

eCH-0160 – Archivische Ablieferungsschnittstelle

Name	Archivische Ablieferungsschnittstelle (SIP)
eCH-Nummer	eCH-0160
Kategorie	Standard
Reifegrad	Implementiert
Version	1.1
Status	Genehmigt
Genehmigt am	2015-06-03
Ausgabedatum	2015-06-24
Ersetzt Version	Minor Change
Abhängigkeiten	eCH-0150
Beilagen	XML-Schema eCH-0160-1.1.xsd.zip
Sprachen	Deutsch (Original), Französisch (Übersetzung)
Autoren	Fachgruppe Digitale Archivierung Georg Büchler, KOST georg.buechler@kost.admin.ch 079 463 98 90
Herausgeber / Vertrieb	Verein eCH, Mainaustrasse 30, Postfach, 8034 Zürich T 044 388 74 64, F 044 388 71 80 www.ech.ch / info@ech.ch

Zusammenfassung

Dieses Dokument spezifiziert eine Ablieferungsschnittstelle für Dossiers und Dokumente aus Geschäftsverwaltungssystemen, Fachanwendungen/Datenbanken und Dateisammlungen an ein Archiv. Es handelt sich dabei um eine Präzisierung des Submission Information Packages SIP, einem der Kernkonzepte des Open Archival Information Systems OAIS (ISO 14721:2003), des zentralen Referenzmodells für die Archivierung digitaler Unterlagen.

Die Grundlage für das vorliegende Dokument bildet die Spezifikation Submission Information Package (SIP), Version 4.0, des Schweizerischen Bundesarchivs BAR. In diese Spezifikation sind Ergänzungen und Korrekturen aus den in der Koordinationsstelle für die dauerhafte Archivierung digitaler Unterlagen KOST vertretenen Archiven eingeflossen. Die Schnittstelle wird vom BAR und von verschiedenen Schweizer Staatsarchiven bereits verwendet.

Im Minor Release wurden weitere Lockerungen vorgenommen, archivneutrale Formulierungen in den XSD-Dateien gewählt und das Addendum zur Version 1.0 implementiert.

Die vorliegende Version 1.1 ersetzt nicht die Version 1.0. Das heisst, dass die Version 1.0 nach wie vor gültig bleibt.

Inhaltsverzeichnis

1	Status des Dokuments	5
2	Einleitung	5
2.1	Ziel und Zweck	5
2.2	Struktur des Dokuments	6
2.2.1	Aufbau Kapitel	6
2.2.2	ID Anforderungen	6
2.2.3	Unterscheidung zwischen Muss und Kann-Anforderungen	7
2.2.4	Notation Ordner, Dateien und Ordnerstrukturen	7
2.3	Präzisierungsbedarf	7
2.4	Übersicht	8
2.4.1	Abgrenzung	8
2.4.1.1	Definitionen	8
2.4.1.2	Anforderungen pro Abliefertyp	9
2.4.2	Anwendungsgebiet	9
2.4.2.1	Prozess	9
2.4.3	Ablieferungsobjekt Paket – SIP	10
2.4.4	Inhalt	11
2.5	Verhältnis Ablieferung – Paket	11
2.6	Sicherheitseinstellungen Dateien im Paket	11
3	Primärdaten	12
3.1	Grundsätzliches	12
3.2	Formate	12
4	Metadaten	13
4.1	Allgemeine Anforderungen	13
4.2	Konzeptuelles Datenmodell	13
4.3	Datenmodell Abliefertyp GEVER	16
4.4	Datenmodell Abliefertyp FILES	17
4.5	Data Dictionary	19
4.6	XSD	19
4.7	Aufbau Inhaltsverzeichnis im metadata.xml	20
4.8	Mapping Metadaten des Ursprungssystems auf SIP-Metadaten	21

4.9	Metadaten zu den Schutzfristen.....	23
4.10	Metadaten zu den Zeiträumen der Dossiers	24
4.11	Metadaten zu den Prüfsummen der Dateien.....	24
4.12	Referenzierung der Zugehörigkeit zu Dossiers in den Metadaten	25
5	Struktur SIP.....	26
5.1	Rahmenvorgaben und Begrenzungen Grösse Paket.....	26
5.2	Rahmenvorgaben und Begrenzungen Anzahl Dateien im Paket.....	26
5.3	Erlaubte Zeichen bei der Benennung der Ordner und Dateien.....	27
5.4	Aufbau eines SIP	29
5.5	Pfadlängen und Hierarchien im SIP	33
5.6	Spezifische Vorgaben zu einem SIP für GEVER.....	33
5.7	Spezifische Vorgaben zu einem SIP für FILES	37
5.8	Aufbau eines FILES SIP mit integrierter Dokumentation	38
6	Transfer.....	40
6.1	Format Package Container	40
7	Version und Gültigkeit der Spezifikation	40
8	Change-Request-Prozess	41
9	Sicherheitsüberlegungen.....	41
10	Haftungsausschluss/Hinweise auf Rechte Dritter	42
11	Urheberrechte.....	42
	Anhang A – Referenzen & Bibliographie	43
	Anhang B – Mitarbeit & Überprüfung.....	44
	Anhang C – Abkürzungen.....	45
	Anhang D – Glossar	46
	Anhang E – Änderungen gegenüber Version 1.0.....	48
	Anhang F – Gültigkeit Kapitel.....	50
	Anhang G – Ausschnitt Mapping SIP-Metadaten – GEVER-System	52
	Anhang H – Zeichensätze	53
	Anhang I – Beziehung zu weiteren archivischen Standards	61

1 Status des Dokuments

Genehmigt: Das Dokument wurde vom Expertenausschuss genehmigt. Es hat für das definierte Einsatzgebiet im festgelegten Gültigkeitsbereich normative Kraft.

2 Einleitung

2.1 Ziel und Zweck

Die Spezifikation Archivische Ablieferungsschnittstelle (SIP) hat die folgenden **Ziele**:

- Die Spezifikation enthält die Anforderungen für die Realisierung eines Submission Information Package (SIP), das digitale Informationspaket für die Ablieferung von digitalem Archivgut an ein Archiv. Die Spezifikation legt fest, wie eine digitale Ablieferung an das Archiv aussehen muss und dient damit den abliefernden Stellen einerseits als intern anzuwendende Spezifikation wie auch zur Kommunikation mit ihren Leistungserbringern und Softwareherstellern.
- Die Spezifikation gibt Auskunft über die Anforderungen, die bei der Implementierung von digitalen Schnittstellen in GEVER-Systemen für die Ablieferung und bei der Erstellung von digitalen Ablieferungen aus Datenbanken und aus Dateisystemen eingehalten werden müssen.

Die Spezifikation Archivische Ablieferungsschnittstelle richtet sich an die folgenden

Zielgruppen:

- Abliefernde Stellen
- Leistungserbringer und Softwarehersteller von Applikationen, welche archivische Ablieferungsschnittstellen implementieren.
- Mitarbeitende des Archivs

2.2 Struktur des Dokuments

2.2.1 Aufbau Kapitel

Jedes Kapitel in dieser Spezifikation ist nach demselben Muster aufgebaut. Nach einer kurzen Einleitung werden die Anforderungen in einer Tabelle aufgeführt.

ID	Beschreibung Anforderung	M/K
enthält die ID der Anforderung	enthält den Anforderungstext	definiert ob Muss- oder Kann-Anforderung

Eine Anforderung wird häufig durch Empfehlungen und Beispiele weiter erklärt. Sowohl Empfehlungen wie auch Beispiele sind speziell gekennzeichnet.

ID	Beschreibung Anforderung	M/K
A_4.1-1	<p>Anforderungstext</p> <p>Beispiel Beispieltext</p> <p>Empfehlung <i>Empfehlungstext ist immer kursiv.</i></p>	M

2.2.2 ID Anforderungen

Die Anforderungen sind über eine ID eindeutig identifizierbar.

ID
A_4.1-1

Diese ID ist nach dem folgenden Muster aufgebaut:

A_	Buchstabe + _ identifiziert Hauptkapitel
A_	= Allgemeine Anforderungen
P_	= Anforderungen an die Primärdaten
M_	= Anforderungen an die Metadaten
S_	= Anforderungen an die Paketstruktur
T_	= Anforderungen an den Pakettransfer
4.1-1	Die Nummer beginnt mit der Angabe des Kapitels (Gruppierung der Anforderungen zum gleichen Thema), die Zahl hinter dem Bindestrich wird durchnummeriert und kennzeichnet so alle Anforderungen des Kapitels.

2.2.3 Unterscheidung zwischen Muss und Kann-Anforderungen

Jede Anforderung ist entweder eine Muss- oder eine Kann-Anforderung. Dies wird mit einem Buchstaben kenntlich gemacht, der auf die Verbindlichkeit verweist:

Abkürzung	Bedeutung
M	Muss-Anforderung Diese Anforderung muss erfüllt sein, um eine gültige Ablieferung zu erhalten.
K	Kann-Anforderung Diese Anforderung sollte erfüllt sein. Sie vereinfacht das Handling einer Ablieferung sowohl für die abliefernde Stelle wie auch für das Archiv im Sinne von <i>Best Practice</i> . Das bedeutet, dass es dem der archivierenden Stelle frei steht, die Kann-Anforderungen von den abliefernden Stellen (Paketerstellung) einzufordern oder nicht.

2.2.4 Notation Ordner, Dateien und Ordnerstrukturen

Für die Notation von Ordnern, Dateien etc. werden die folgenden Symbole und Parameter verwendet.

Symbol	Bedeutung
/	Ordner
header/	Ein Ordner mit dem Namen «header»
xy.txt	Datei (mit Datei-Endung «txt»)
dir1/	Beispiel-Ordner (in roter Farbe)
abc.pdf	Beispiel-Dateien (in roter Farbe)
...	Platzhalter für Dateien oder Ordner, die für die Erklärung nicht relevant sind.
[]	Platzhalter für einen Ausdruck oder einen Basistyp wie «string», «integer» etc.

2.3 Präzisierungsbedarf

Institutionen, die die vorliegende Spezifikation implementieren, müssen unter Umständen die folgenden Punkte präzisieren:

- zugelassene archivtaugliche Formate für die Primärdaten (siehe 3.2)
- Maximalgrösse eines SIP (siehe 5.1)
- Maximale Anzahl Dateien in einem Ordner innerhalb des SIP (siehe 5.2)
- Name des Toplevel-Ordners / SIP-Identifikators (siehe 5.4).
- Übernahmevorgaben für das Ordnungssystem (siehe 4.8-2)

2.4 Übersicht

2.4.1 Abgrenzung

Das vorliegende Dokument spezifiziert das Submission Information Package (SIP), das Ablieferungsobjekt für digitale Unterlagen an ein Archiv. Digitale Unterlagen müssen aufbereitet als SIP an das Archiv abgeliefert werden. (Vgl. Abb. 1).

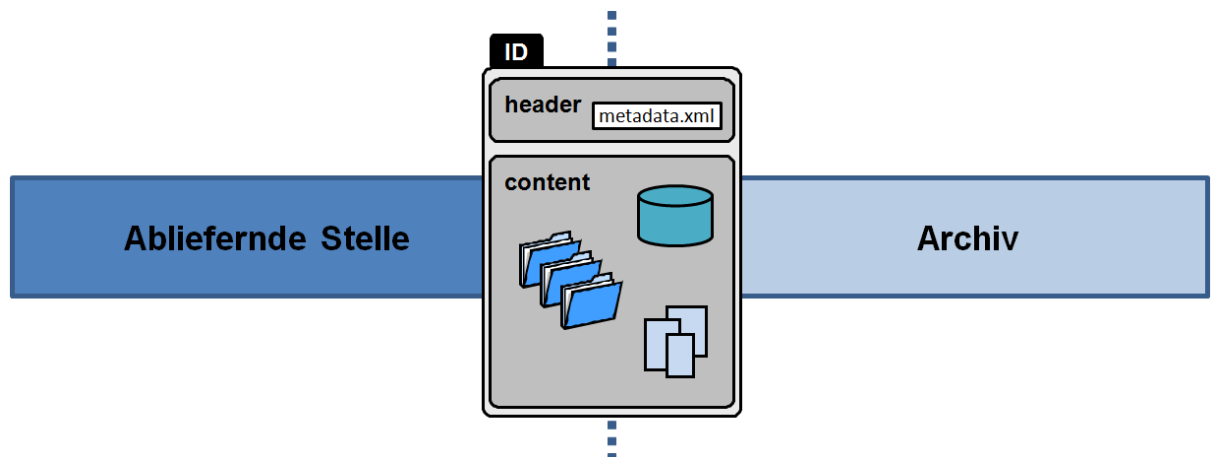


Abb. 1: Abgrenzung Ablieferungsobjekt SIP

Die vorliegende Spezifikation beschreibt die Anforderungen an alle technischen Implementierungen, die zum Ziel haben, ein SIP für das Archiv aus einem bestehenden System zu extrahieren resp. aus einer Datei-Ablage herzustellen. (z. B. aus einem GEVER-System oder mit Daten aus einer Datenbank oder aus einer Sammlung von zusammengehörenden digitalen Unterlagen aus einer Datei-Ablage).

2.4.1.1 Definitionen

Im Rahmen dieser Spezifikation wird auf oberster Ebene zwischen zwei Ablieferungstypen unterschieden: GEVER-Ablieferungen und FILES-Ablieferungen. Der Typ FILES kann zusätzlich noch als FILES mit integrierter Dokumentation vorkommen. Dabei handelt es sich nicht um einen eigenen Ablieferungstyp sondern um einen Untertyp zur FILES-Ablieferung. Diese Typisierung leitet sich aus der vorarchivischen Herkunft des digitalen Archivgutes ab.

Ablieferungstyp	Beschreibung
GEVER	Das digitale Archivgut in der Ablieferung stammt aus einem Geschäftsverwaltungssystem (GEVER-System).
FILES	Das digitale Archivgut in der Ablieferung stammt aus einer Dateiablage oder aus einer relationalen Datenbank oder aus einem anderen System.

Ablieferungstyp		Beschreibung
	FILES-SIP mit integrierter Dokumentation	Falls die FILES-Ablieferung digitale Unterlagen enthält, die aus einer Datenbank oder aus einem anderen System stammen, für das dem Archiv zusätzlich noch eine System-Dokumentation abgeliefert wird, gibt es zusätzlich zu den allgemeinen Vorgaben für FILES-SIP noch weitergehende Anforderungen an die Struktur. Bei diesen Ablieferungen handelt es sich um ein FILES-SIP mit integrierter Dokumentation zu den Daten.

2.4.1.2 Anforderungen pro Ablieferungstyp

Nicht alle Kapitel der vorliegenden Spezifikation müssen bei der Erstellung eines SIP angewendet werden. Die Anforderungen an ein SIP können sich aufgrund der vorarchivischen Herkunft der Unterlagen unterscheiden. Die Abbildung 2 hilft dabei, zu eruieren, welche Kapitel der vorliegenden Spezifikation für welche Ablieferung relevant sind. Die genaue Auflistung der relevanten Kapitel ist anhand der Übersicht im Anhang C ersichtlich.

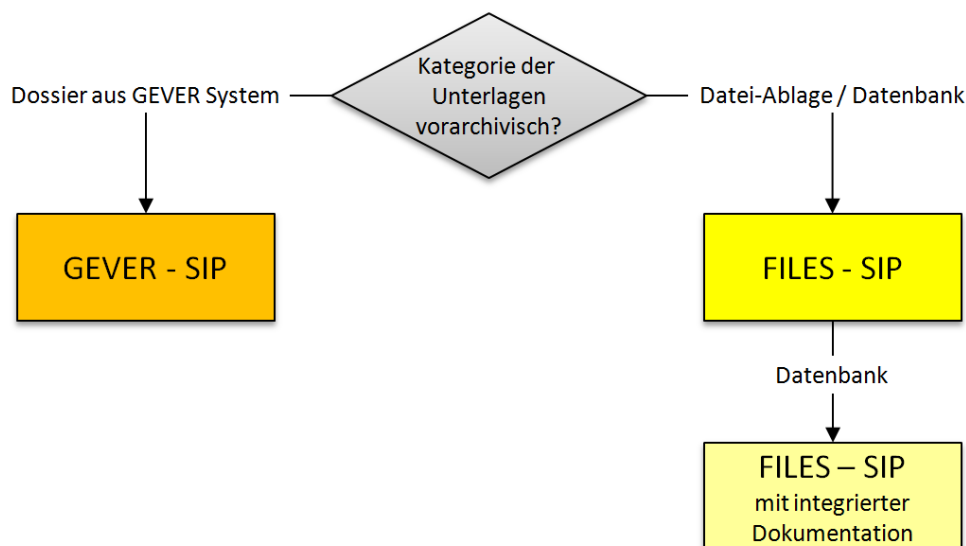


Abb. 2: Ablieferungskategorien SIP

2.4.2 Anwendungsgebiet

2.4.2.1 Prozess

Diese Spezifikation wurde auf der Grundlage des Ablieferungsprozesses des Schweizerischen Bundesarchivs für digitale Unterlagen entwickelt. Die als archivwürdig bewerteten digitalen Unterlagen müssen technisch und inhaltlich gemäss den Anforderungen in dieser Spezifikation aufbereitet und in einem korrekt erstellten SIP geliefert werden. Ist dies nicht der Fall werden die Unterlagen an die abliefernde Stelle zurückgesendet, damit diese die Korrekturen durchführen kann, resp. ein neues SIP gemäss den Anforderungen erstellt.

2.4.3 Ablieferungsobjekt Paket – SIP

Das Ablieferungsobjekt ist ein Informationspaket, das auf dem Modell des OAIS-Standard (Open Archival Information System) basiert (vgl. Abb. 3). Dieser Standard wurde durch das Consultative Committee for Space Data System (CCSDS) entwickelt. Eines der Hauptelemente des OAIS ist, dass alle zusammengehörenden Informationen in einem sogenannten Paket zusammengefasst werden.

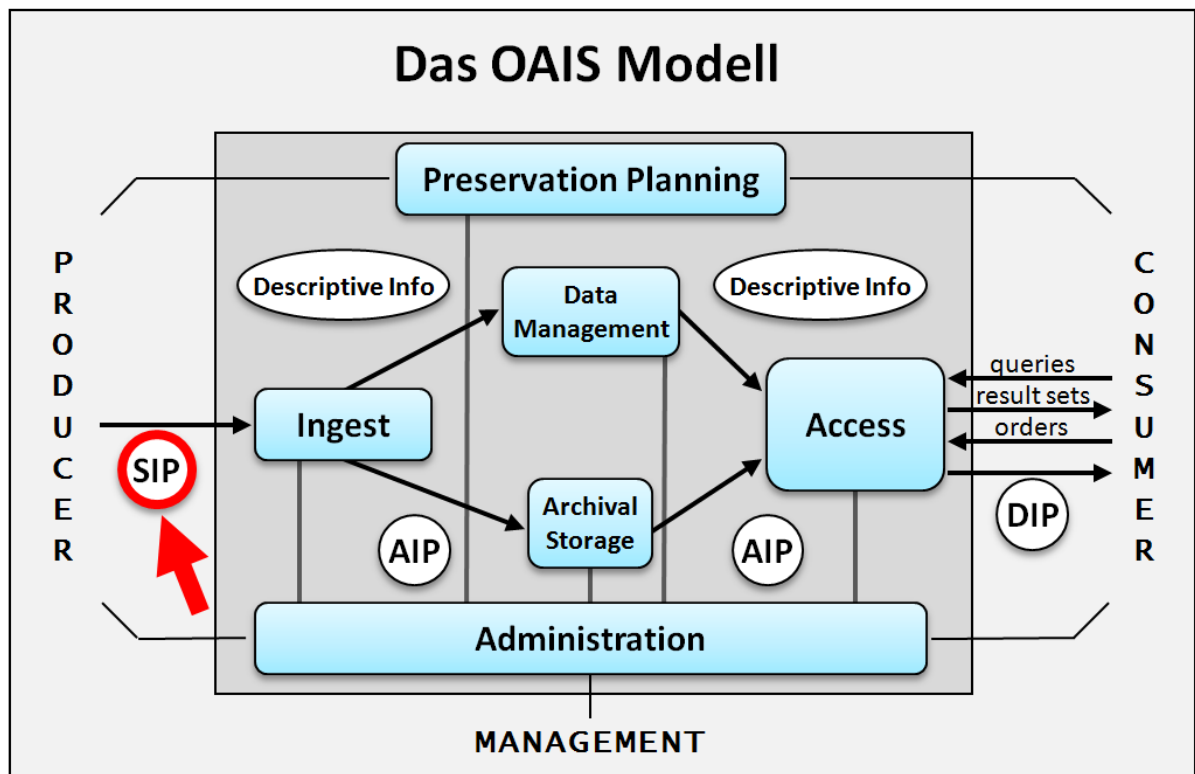


Abb. 3: OAIS Modell mit Kennzeichnung SIP

Es gibt drei Pakettypen: SIP, AIP und DIP.

Pakettyp		
SIP	Submission Information Package	Paket zur Ablieferung digitaler Unterlagen
AIP	Archival Information Package	Paket zur Archivierung digitaler Unterlagen
DIP	Dissemination Information Package	Paket zur Vermittlung digitaler Unterlagen

In der vorliegenden Spezifikation werden die Anforderungen an ein SIP definiert (archivisch und technisch). Bei Einhaltung dieser Spezifikation ist sichergestellt, dass alle Pakete mit einer entsprechenden Archivierungsinfrastruktur und Archivierungsprozessen für digitales Archivgut kompatibel sind.

2.4.4 Inhalt

Basierend auf dem OAIS Modell enthält das SIP gemäss der vorliegenden Spezifikation zwei separate Teile, die immer gemeinsam abgeliefert werden müssen. Der erste Teil ist der *Header*. Dieser Teil enthält alle Metadaten (beschreibende Informationen betreffend das Paket und die Primärdaten) im XML-Format. Der zweite Teil ist der *Content*. Dieser Teil enthält alle Primärdaten. (Vgl. Abb. 4.)

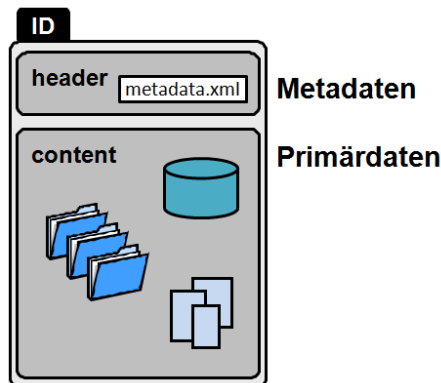


Abb. 4: Aufbau SIP – Metadaten und Primärdaten

2.5 Verhältnis Ablieferung – Paket

Im Fall einer Ablieferung von digitalen Unterlagen muss diese in der Form eines SIP abgeliefert werden.

ID	Beschreibung Anforderung	M/K
A_2.5-1	Eine Ablieferung an das Archiv besteht aus einem oder mehreren Paketen (SIP). Dies verringert die Komplexität und vereinfacht das Handling sowohl für die abliefernde Stelle wie auch für das Archiv.	M

2.6 Sicherheitseinstellungen Dateien im Paket

Die Dateien, die in einem SIP enthalten sind, müssen ohne Einschränkung zugänglich sein. Für den Transfer von digitalen Unterlagen mit erhöhtem Schutzbedarf von der abliefernden Stelle ins Archiv (und nur dafür) kann es nötig sein, das SIP als Ganzes z. B. mit einem Passwort zu schützen. Die Anforderungen für den Transfer sind zwischen der abliefernden Stelle und dem Archiv vorgängig auszuhandeln.

ID	Beschreibung Anforderung	M/K
A_2.6-1	Die Dateien in einem SIP dürfen nicht verschlüsselt sein.	M
A_2.6-2	Die Dateien in einem SIP dürfen nicht passwortgeschützt sein.	M



3 Primärdaten

3.1 Grundsätzliches

Primärdaten sind alle digitalen Unterlagen, die in den abliefernden Stellen (im Rahmen von Geschäften, in Dossiers, in Geschäftsverwaltungssystemen, in Datenbanken) produziert wurden. Es gibt viele verschiedene Typen (wie z. B. Textdateien, Bilder, Datenbanken, Präsentationen, etc.). Primärdaten können «born-digital» Unterlagen sein wie auch «digitalisierte» Unterlagen. Als Primärdaten werden alle in diesem Sinne entstandenen Unterlagen bezeichnet.

3.2 Formate

Die Primärdaten, die in einem SIP, das dem Archiv abgeliefert wird, enthalten sind, müssen zwingend in einem archivtauglichen Format vorliegen. Das zuständige Archiv definiert die als archivtauglich akzeptierten Formate. Diese sind nicht Bestandteil der vorliegenden Spezifikation.

ID	Beschreibung Anforderung	M/K
P_3.2-1	<p>Die Primärdaten müssen in einem für die Unterlagen geeigneten archivtauglichen Format im SIP integriert werden.</p> <p>Empfehlung <i>Es ist von Vorteil, die Unterlagen bereits in einem archivtauglichen Format zu erstellen oder sie so früh wie möglich in ein archivtaugliches Format zu konvertieren. In einem GEVER-System ist die Konversion spätestens bei Dossierabschluss vorzunehmen. Auf jeden Fall muss die Konversion der Primärdaten vor der Integration der Unterlagen in das SIP erfolgen.</i></p>	M
P_3.2-2	<p>Die Dateiendungen der Dateien müssen den üblichen Konventionen in Bezug auf die Formate der jeweiligen Dateien entsprechen.</p> <p>Beispiel Eine Bilddatei im TIFF-Format:  p00010.tif Eine Tondatei im WAVE-Format:  das_lied_1.wav</p>	M

4 Metadaten

Metadaten sind «Informationen über Daten». Sie beschreiben die Primärdaten genauer. Metadaten können in eigenen Dateien abgespeichert (sog. «separate Metadaten» z.B. in einer XML-Datei), oder als so genannte «eingebettete Metadaten» in einer Primärdatei bereits enthalten sein. Ein MS-Word Dokument beinhaltet beispielsweise eingebettete Metadaten in Form von Titel, Thema, Autor etc. Der Erhalt von eingebetteten Metadaten kann aufgrund von möglichen Formatkonvertierungen im Archiv nicht sichergestellt werden.

In diesem Kapitel werden die Metadaten für die SIP der folgenden Ablieferungstypen spezifiziert:

- GEVER: Ablieferung aus einem GEVER-System, sogenannte GEVER-Ablieferung.
- FILES: Ablieferung einer Sammlung von Dateien (sog. FILES-Ablieferung). Unter den FILES-Ablieferungen sind auch die Ablieferungen aus relationalen Datenbanken mit integrierter Dokumentation subsumiert.

4.1 Allgemeine Anforderungen

ID	Beschreibung Anforderung	M/K
M_4.1-1	Alle Metadaten, welche über die gesamten Archivierungsprozesse hinweg erhalten bleiben müssen, müssen als separate Metadaten in einer separaten Datei im XML-Format namens <code>metadata.xml</code> (getrennt von den Primärdaten) vorliegen. Die Spezifikation gibt sowohl die Struktur der Metadatenfile wie auch Inhalt der Metadaten vor (vgl. dazu die Kapitel zur XSD und zum Data Dictionary). Diese Vorgaben müssen exakt eingehalten werden.	M

4.2 Konzeptuelles Datenmodell

Das konzeptuelle Datenmodell vermittelt eine Übersicht über alle grundlegenden Komponenten (Entitäten) im `metadata.xml` sowie über deren Repräsentation im XSD, das dem `metadata.xml` zugrunde liegt. Im konzeptuellen Datenmodell werden die Entitäten, die zum digitalen Informationspaket gehören, wie auch die Beziehungen zwischen diesen Entitäten grafisch dargestellt. Jede Entität umfasst einen Teil der Metadaten im `metadata.xml`. Das konzeptuelle Datenmodell liegt sowohl für das `metadata.xml` für Pakete des Ablieferungstyps GEVER wie auch für Pakete des Ablieferungstyps FILES vor.

ID	Beschreibung Anforderung	M/K
M_4.2-1	Das Einhalten bzw. Umsetzen des konzeptuellen Datenmodells ist für alle Ablieferungstypen eine verbindliche Vorgabe. Es dient der Orientierung und gibt einen Überblick über alle existierenden Entitäten.	M
M_4.2-2	Das konzeptuelle Datenmodell unterscheidet zwischen zwei Ablieferungstypen, nämlich GEVER und FILES. Je nach Herkunft des Archivgutes muss der entsprechende Ablieferungstyp gewählt werden.	M

Notation des konzeptuellen Datenmodells

Das konzeptuelle Datenmodell wird in der vorliegenden Spezifikation gemäss dem UML-Standard (Unified Modeling Language) notiert.

Beschreibung der Entitäten

Entität	Beschreibung
Paket	Die Entität <i>Paket</i> beschreibt die «Verpackung» der Ablieferung in einen technischen Behälter (Ordner). Das Paket trägt je nach Phase im Paket-Lifecycle einen anderen Namen (SIP/AIP/DIP). Für die Ablieferung von digitalem Archivgut handelt es sich immer um ein SIP.
Ablieferung	Die Entität <i>Ablieferung</i> beinhaltet alle Angaben zur abliefernden Stelle und zur Ablieferung ans Archiv.
Provenienz	Die Entität <i>Provenienz</i> repräsentiert den Aktenbildner und bei -Ablieferungen vom Ablieferungstyp FILES zusätzlich das Informationssystem, aus dem die Ablieferung stammt.
Ordnungssystem	Die Entität <i>Ordnungssystem</i> beschreibt den amtsseitig verwendeten Registraturplan, Aktenplan oder Strukturplan resp. die Ordnung des Archivgutes.
Ordnungssystemposition	Die Entität <i>Ordnungssystemposition</i> beschreibt die Eigenschaften der einzelnen Ordnungssystempositionen im Ordnungssystem.
Dossier	Die Entität <i>Dossier</i> beschreibt die Eigenschaften des zusammengehörenden Archivguts aus einem Geschäft oder aus einer anderen Ordnungsstruktur (Datenbank, Datei-Ablage). Die Metadaten aus der Entität sind sowohl für Dossiers wie auch für Subdossiers anzuwenden.
Vorgang	Die Entität <i>Vorgang</i> repräsentieren einzelne, abgrenzbare Geschäftsfälle in Zusammenhang mit einem Dossier. Der Vorgang wird durchgeführt über eine Folge von einzelnen Aktivitäten. Aus Gründen der Rückwärtskompatibilität mit Version 1.0 kann <i>Vorgang</i> auch weiterhin als Attribut der Entität <i>Dossier</i> verwendet werden.
Aktivität	Die Entität <i>Aktivität</i> beschreibt die Tätigkeit, die ausgeführt werden soll.
Dokument	Die Entität <i>Dokument</i> repräsentiert die kleinste Ebene auf der das Archivgut inhaltlich dargestellt wird. Diese Entität enthält die Metadaten, die zum Verständnis der eigentlichen Unterlagen resp. ihrem Inhalt notwendig sind.
Inhaltsverzeichnis	Die Entität <i>Inhaltsverzeichnis</i> repräsentiert das im <code>metadata.xml</code> enthaltene Inhaltsverzeichnis, das alle Ordner und Dateien umfasst.
Ordner	Die Entität <i>Ordner</i> repräsentiert die Ordner im Paket, in denen das Paket strukturiert ist.
Datei	Die Entität <i>Datei</i> repräsentiert die Primärdateien im Dateisystem und beinhaltet auch Angaben zur Identifikation und Beschreibung der einzelnen Dateien.

Entität	Beschreibung
Unstrukturierter Anhang	Die Entität <i>Unstrukturierter Anhang</i> repräsentiert die Beschreibung von Dateien, bei denen es sich nicht um eigentliches Archivgut handelt, die dem Paket als zusätzliche Information in einem speziell dafür gekennzeichneten Ordner beigegeben werden. Diese Entität ist <i>deprecated</i> . Von ihrer Verwendung wird abgeraten. In zukünftigen Versionen des Standards wird sie unter Umständen nicht mehr enthalten sein.
Archivischer Vorgang	Die Entität <i>archivischer Vorgang</i> wird erst im Archiv mit Metadaten gefüllt und dokumentiert Tätigkeiten wie die Eingangskontrolle in einem SIP oder auch Formatkonvertierungen bei einem AIP. Für die Erstellung eines SIP ist diese Entität nicht von Bedeutung. Sie darf bei der Erstellung nicht verwendet werden.
Archivische Notiz	Die Entität <i>archivische Notiz</i> wird vom Archiv verwendet, um Notizen zu Objekten aus den einzelnen Entitäten zu erfassen. Für die Erstellung eines SIP ist diese Entität nicht von Bedeutung. Sie darf bei der Erstellung nicht verwendet werden.

Beziehungen zwischen den Entitäten

Die Beziehungen zwischen den Entitäten sind im konzeptuellen Datenmodell aus der Paket-internen Sicht zu lesen. Das konzeptuelle Datenmodell bildet damit die Welt eines einzelnen Paketes (und damit einer einzelnen digitalen Ablieferung) ab, wie sie im `metadata.xml` dargestellt wird. Dabei können Entitäten referenziell miteinander verbunden sein oder auch sich selbst referenzieren, wie dies bei der Ordnungssystemposition, dem Dossier und dem Ordner der Fall ist. Bei der Ordnungssystemposition ist dies notwendig, weil in einem Ordnungssystem verschiedene Hierarchien bis zur untersten Ordnungssystemposition (der Rubrik) abgebildet werden müssen. Ein Dossier kann verschiedene Subdossiers enthalten, welche aber in der Welt des Datenmodells alle unter der Entität *Dossier* subsumiert sind. Auch die Entität *Ordner* kann sich selber referenzieren und bildet die verschiedenen Hierarchien in der Ordnerstruktur eines Paketes ab.

Bei den Beziehungen von der Entität *Provenienz* zur Entität *Ablieferung* und von der Entität *Ordnungssystem* zur Entität *Ablieferung* handelt es sich im Paket immer um 1:1-Beziehungen. Es kann nur eine Provenienz und ein Ordnungssystem für das digitale Archivgut in einem Paket geben. Allerdings können von einer Provenienz mehrere Ablieferungen gemacht werden und auch aus einem Ordnungssystem werden über die Zeit hinweg mehrere Ablieferungen getätigt. Diese Tatsache wird aber im konzeptuellen Datenmodell nicht abgebildet, da dieses nur die Sicht auf ein einziges Paket enthält.

Die Entitäten *Archivischer Vorgang* und *Archivische Notiz* sind Entitäten, die Metadaten enthalten, die erst im Archiv (nach dem Transfer des Paketes von der abliefernden Stelle ins Archiv) von Bedeutung sind. Für die Erstellung eines Paketes in der abliefernden Stelle sind diese Entitäten nicht von Bedeutung. Sie werden deshalb hier – genauso wie die Entität *Unstrukturierter Anhang*, die für das SIP nicht von Bedeutung ist – nur der Vollständigkeit halber für die Dokumentation aller Entitäten im XSD aufgeführt.

4.3 Datenmodell Ablieferungstyp GEVER

ID	Beschreibung Anforderung	M/K
M_4.3-1	<p>Der Ablieferungstyp GEVER gilt für alle Ablieferungen, die digitale Unterlagen aus einem Geschäftsverwaltungssystem (GEVER) enthalten. Das Datenmodell für GEVER-Ablieferungen ist in der Abb. 5 ersichtlich.</p> <p>Die folgenden Entitäten kommen im GEVER SIP (resp. in den Metadaten) vor:</p> <ul style="list-style-type: none"> • Provenienz • Paket • Ablieferung • Ordnungssystem • Ordnungssystemposition • Inhaltsverzeichnis • Dossier • Vorgang • Aktivität • Dokument • Datei • Ordner <p>Die Entität <i>Archivischer Vorgang</i> wird erst nach dem Transfer des Paketes ins Archiv relevant. Die Anzahl der archivischen Vorgänge in einem SIP einer abliefernden Stelle vor dem Transfer beträgt daher immer 0.</p> <p>Die Entität <i>Archivische Notiz</i> wird erst nach dem Transfer des Paketes ins Archiv relevant. Die Anzahl der archivischen Notizen in einem SIP einer abliefernden Stelle vor dem Transfer beträgt daher immer 0.</p> <p>Die Entität <i>Unstrukturierter Anhang</i> soll in einem GEVER SIP nicht vorkommen.</p>	M

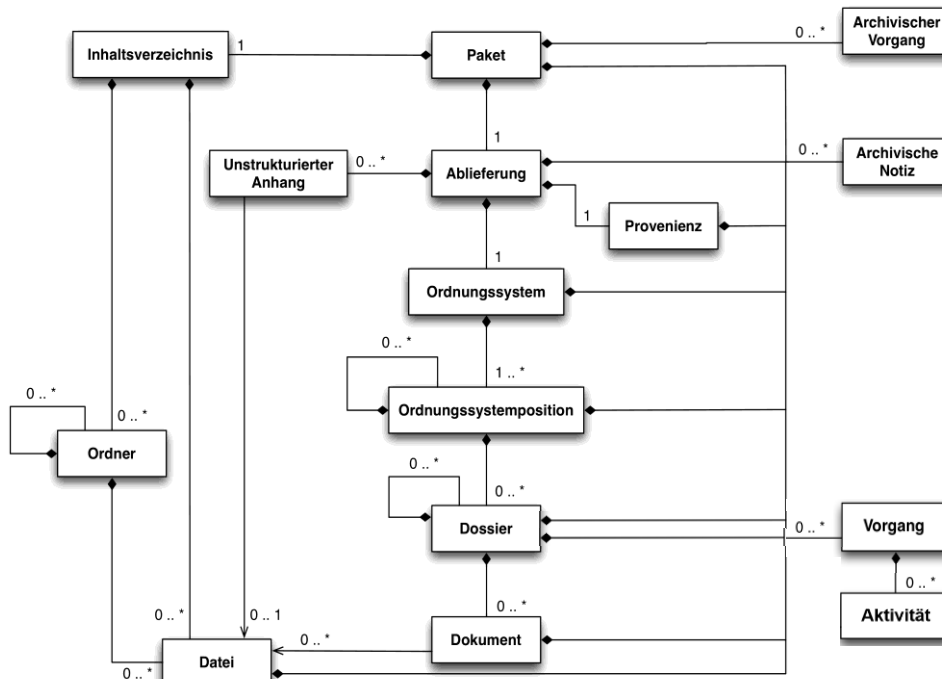


Abb. 5: Konzeptuelles Datenmodell– Anwendungsbereich Ablieferungstyp GEVER

4.4 Datenmodell Ablieferungstyp FILES

ID	Beschreibung Anforderung	M/K
M_4.4-1	<p>Die Ausprägung FILES gilt für alle Ablieferungen, die digitale Unterlagen aus einer relationalen Datenbank oder aus einer Datei-Ablage enthalten. Das Datenmodell für FILES-Ablieferungen ist in der Abb. 6 ersichtlich.</p> <p>Die folgenden kommen im FILES SIP (resp. in den Metadaten) vor:</p> <ul style="list-style-type: none"> • Provenienz • Paket • Ablieferung • Ordnungssystem • Ordnungssystemposition • Inhaltsverzeichnis • Dossier • Dokument • Datei • Ordner <p>Die Entitäten <i>Dokument</i> muss in einem FILES SIP nicht zwingend vorkommen. Sie kann bei Bedarf verwendet werden.</p> <p>Die Entität <i>Archivischer Vorgang</i> wird erst nach dem Transfer des Paketes ins Archiv relevant. Die Anzahl der archivischen Vorgänge in einem SIP einer abliefernden Stelle vor dem Transfer beträgt daher immer 0.</p> <p>Die Entität <i>Archivische Notiz</i> wird erst nach dem Transfer des Paketes ins Archiv relevant. Die Anzahl der archivischen Notizen in einem SIP einer abliefernden Stelle vor dem Transfer beträgt daher immer 0.</p> <p>Die Entitäten <i>Vorgang</i>, <i>Aktivität</i> und <i>Unstrukturierter Anhang</i> sollen in einem FILES SIP nicht vorkommen.</p>	M

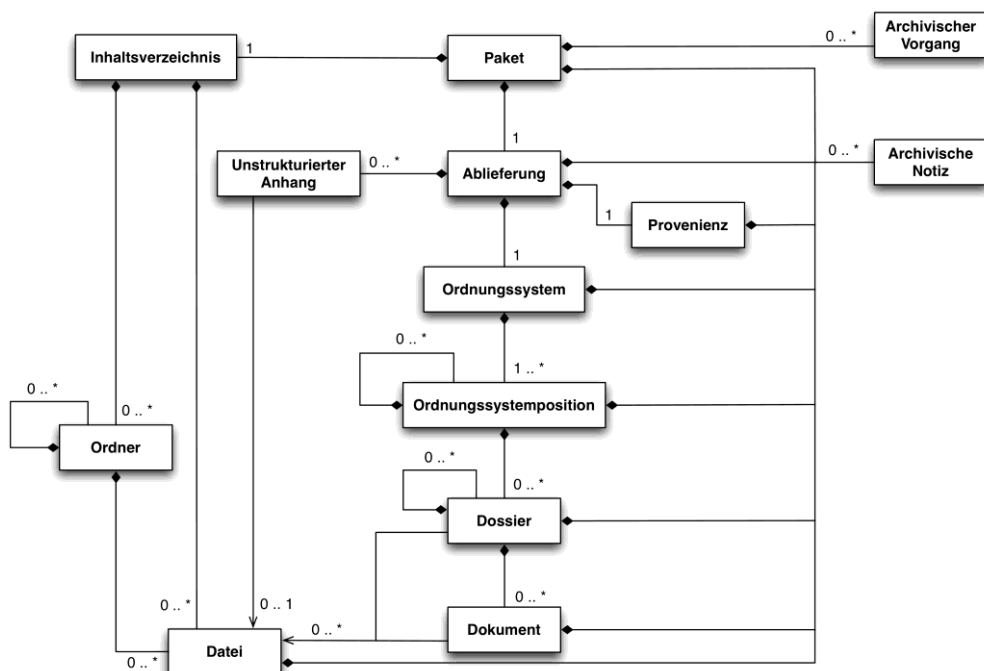


Abb. 6: Konzeptuelles Datenmodell– Anwendungsbereich Ablieferungstyp FILES

ID	Beschreibung Anforderung	M/K
M_4.4-2	<p>Für die Beziehungen zwischen den Entitäten Dossier, Dokument und Datei gibt es im FILES SIP zwei Fälle (vgl. Abb. 7).</p> <p>Fall A (blau markiert): Eine Datei ist direkt einem Dossier zugeordnet. Ein Dossier besitzt 0..* Dateien. Es gibt keine Dokumente.</p> <p>Fall B (grün markiert): Eine Datei ist einem Dokument zugeordnet. Das Dokument ist dem Dossier zugeordnet. Ein Dossier besitzt 0..* Dateien. 1 Dokument besteht aus 0..* Dateien. (wie bei der Ausprägung GEVER)</p> <p>Fall A und B können gleichzeitig in einem Paket vorkommen. Fall A und B können gleichzeitig in einem Dossier vorkommen.</p>	M

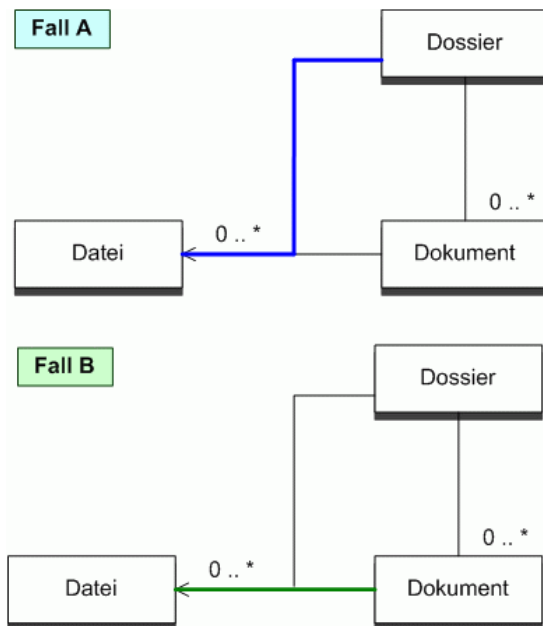


Abb. 7: Unterschiedliche Möglichkeiten der Beziehungen zwischen Dossier, Dokument und Datei

4.5 Data Dictionary

Im Data Dictionary sind die einzelnen Metadatenelemente nach Entitäten aufgeteilt und spezifiziert. Das Data Dictionary basiert auf dem bereits vorgestellten Datenmodell. Jedes Kapitel des Data Dictionary behandelt eine Entität.

ID	Beschreibung Anforderung	M/K
M_4.5-1	Alle Metadaten, die im Data Dictionary als Muss bezeichnet sind, müssen ins SIP übertragen werden.	M
M_4.5-2	<p>Metadaten, die aus dem Ursprungssystem (GEVER, FILES) und über alle anderen Elemente der Ablieferung bekannt sind und für die ein geeignetes Attribut in den Metadaten des SIP vorhanden ist, müssen zwingend ins SIP übertragen werden.</p> <p>Empfehlung</p> <p><i>Je genauer die digitalen Unterlagen mittels Metadaten beschrieben sind, desto einfacher gestaltet sich auch der Umgang mit dem Archivgut, sobald es wieder bereitgestellt und vermittelt wird.</i></p>	M

Das Data Dictionary mit allen Metadatenelementen des SIP sowie den für die Spezifikation der Metadaten verwendeten Attribute liegt als Beilage zur Spezifikation vor.

4.6 XSD

Die XSD (XML Schema Definition) inkl. Dokumentation beschreibt die Struktur (oder Grammatik) der XML Datei, welche die Metadaten einer Ablieferung enthält (`metadata.xml`). Es ist die technische Umsetzung des Data Dictionary und des Datenmodells und zugleich die Grundlage für die Validierung des `metadata.xml` im SIP.

ID	Beschreibung Anforderung	M/K
M_4.6-1	<p>Die XSD ist für das <code>metadata.xml</code> verbindlich einzuhalten. Das heisst, jedes <code>metadata.xml</code> in einem SIP muss gegenüber der XSD positiv validiert werden können.</p> <p>Als XSD ist entweder die Schemadatei <code>arelda.xsd</code> aus der Beilage zur eCH-0160 Spezifikation Version 1.1 oder die Schemadatei aus der eCH Schemasammlung zu verwenden: http://www.ech.ch/xmlns/eCH-0160/Schema/1.1/arelda.xsd</p> <p>Die XSD <code>arelda.xsd</code> von eCH-0160 Version 1.1 hat die interne Version 4.1</p>	M
M_4.6-2	<p>Als Schema Location kann entweder auf das lokale Schema im SIP Ordner <code>/xsd</code>, oder auf die eCH Schemasammlung verwiesen werden. Oder es wird auf eine Schema Location gänzlich verzichtet.</p> <p>Beispiel</p> <pre>xsi:schemaLocation="http://bar.admin.ch/arelda/v4 xsd/arelda.xsd" xsi:schemaLocation="http://bar.admin.ch/arelda/v4 http://www.ech.ch/xmlns/eCH-0160/Schema/1.1/arelda.xsd"</pre>	K

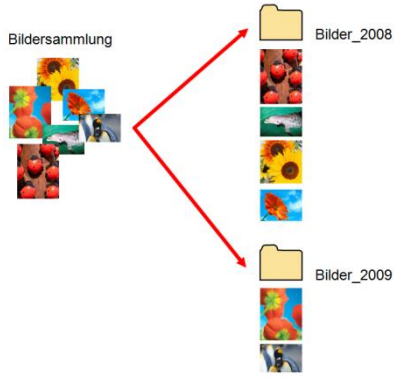




4.7 Aufbau Inhaltsverzeichnis im metadata.xml









ID	Beschreibung Anforderung	M/K
M_4.7-1	<p>Alle Ordner und Dateien im Paket (Inhalte des Ordners <code>content/</code> und Ordners <code>header/</code>) müssen im Inhaltsverzeichnis des <code>metadata.xml</code> gemäss den Vorgaben mit allen Metadaten aufgeführt sein.</p> <p>Die einzige Ausnahme dieser Regel gilt für die Datei <code>metadata.xml</code> im Ordner <code>header/</code>. Diese Datei darf nicht im <code>metadata.xml</code> aufgeführt werden, da sie sonst auf sich selbst verweisen würde.</p> <p>Beispiel Aufbau des SIP (schematisch)</p> <pre> SIP_20090824_Bm/ header/ metadata.xml xsd/ ablieferung.xsd ...xsd content/ d0001/ d0002/ p0001.pdf </pre> <p>Entsprechendes Inhaltsverzeichnis eines <code>metadata.xml</code></p> <pre> <?xml version="1.0" encoding="UTF-8"?> <paket xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance" xmlns="http://bar.admin.ch/arelda/v4" xsi:type="paketSIP" schemaVersion="4.0"> <paketTyp>SIP</paketTyp> <inhaltsverzeichnis> <ordner> <name>header</name> <originalName>header</originalName> <ordner> <name>xsd</name> <originalName>xsd</originalName> <datei id="dat00001"> <name>ablieferung.xsd</name> <originalName>ablieferung.xsd</originalName> <pruefalgorithmus>MD5</pruefalgorithmus> <pruefsumme>aaf17ad13064e1643ac85478296165af</pruefsumme> </datei> </ordner> </ordner> ... <ordner> <name>content</name> <originalName>content</originalName> <ordner> <name>d000001</name> <originalName>Veranstaltungen</originalName> <ordner> <name>d000002</name> <originalName>Ausstellung</originalName> <datei id="dat000015"> <name>p000015.pdf</name> <originalName>Flyer_ausstellung.pdf</originalName> <pruefalgorithmus>MD5</pruefalgorithmus> <pruefsumme>f20168fd9b04d7393b8a68e59f092c69</pruefsumme> </datei> </ordner> </ordner> </ordner> </inhaltsverzeichnis> ... </pre>	M

4.8 Mapping Metadaten des Ursprungssystems auf SIP-Metadaten

Um die Metadaten des SIP korrekt ausfüllen und aus dem System extrahieren zu können, braucht es ein Mapping zwischen den Metadatenfeldern im System des Aktenbildners und denjenigen im SIP.

ID	Beschreibung Anforderung	M/K
M_4.8-1	Die Metadaten, die bei der abliefernden Stelle bereits vorhanden sind, entsprechen in ihren Inhalten, ihrer Aufbereitung und bei bereits vorhandenen Datenmodellen häufig nicht den Strukturen, die für Erstellung eines SIP benötigt werden. Deshalb muss ein Mapping zwischen den Metadaten, die der Aktenbildner resp. die abliefernde Stelle erstellt hat und den Metadaten im SIP zur Ablieferung des Archivgutes ans Archiv vorgenommen werden. Dies muss von der abliefernden Stelle sowohl für GEVER-SIP wie auch für FILES-SIP sichergestellt werden.	M
M_4.8-2	Bei einem GEVER-SIP ist das Ordnungssystem ins SIP zu übernehmen. Empfehlung <i>Es wird empfohlen, zunächst ein Mapping zwischen den Metadaten des ursprünglichen GEVER-Systems und den Metadaten des SIPs herzustellen. Ein Beispiel eines Teils eines solchen Mappings für ein GEVER-System findet sich im Anhang D.</i>	M
M_4.8-3	Bei einem FILES-SIP muss nicht nur die Ablagestruktur (physische Ordnung) auf Ordner- und Dateiebene, sondern auch die entsprechende Systematik auf Ordnungssystem- und Unterlagenebene (logische Ordnung) in den Metadaten verzeichnet werden. Die Unterlagen werden dadurch gemäss ihren Inhalten in sinnvolle Einheiten zusammengefasst und zu einem vermittelbaren Dossier zusammengefügt. Empfehlung <i>Falls keine solche Ordnung bereits vorhanden ist, sollte sie aufgrund der Inhalte der Unterlagen erstellt werden.</i> <i>Es wird empfohlen, bereits vorhandene oder sinnvolle Systematiken und Ordnungen von Datei-Strukturen oberhalb der Dossiers im Ordnungssystem abzubilden und zusammengehörende Unterlagen in Dossiers zusammenzufassen.</i>	M

ID	Beschreibung Anforderung	M/K
	<p>Beispiel</p> <p>Eine Dateiablage beinhaltet eine Bildersammlung aus unzähligen ungeordneten Bildern. Die Bilder könnten nun z. B. nach Entstehungsjahr geordnet und auf Dateiablage-Ebene in Ordnern zusammengefasst werden. Der Inhalt dieser Ordner wird dann jeweils als Dossier in den Metadaten beschrieben.</p>  <p>Der Order content im SIP sieht dann folgendermassen aus:</p> <pre> content/ Bilder_2008/ Kaefer.tif Delfin.tif Sonnenblume.tif Orange_Blume.tif Bilder_2009/ Mehrere_Blumen.tif Pinguine.tif </pre> <p>Modellhaft könnte diese Ordnung mit der Zuordnung zu den Dossiers folgendermassen dargestellt werden:</p>  <p>  = Ordnungssystem  = Ordnungssystemposition  = Dossier </p>	

ID	Beschreibung Anforderung	M/K
M_4.8-4	<p>Für FILES-SIP mit integrierter Dokumentation (z. B. Datenbankarchivierungen mit SIARD) ist ein einfaches Ordnungssystem zu erstellen, um die Dossiers mit der Dokumentation und den Daten zu integrieren.</p> <p>Beispiel</p> <p>Modellhaft könnte eine solche Ordnung folgendermassen dargestellt werden:</p> <ul style="list-style-type: none">  Name der Datenbank <ul style="list-style-type: none">  1 Dokumentation <ul style="list-style-type: none">  Dossier mit der Dokumentation zur Datenbank  2 Daten <ul style="list-style-type: none">  Dossier mit den Daten  = Ordnungssystem  = Ordnungssystemposition  = Dossier 	M

4.9 Metadaten zu den Schutzfristen

Die Angaben zu den Schutzfristen der abgelieferten Unterlagen müssen in den Metadaten des SIP festgehalten werden.

ID	Beschreibung Anforderung	M/K
M_4.9-1	<p>Die Angaben zu den Schutzfristen der Unterlagen im Paket sollten in die dazu zur Verfügung stehenden Metadaten im <code>metadata.xml</code> eingetragen werden.</p> <p>Bei den Metadaten handelt es sich um die folgenden:</p> <p><code><schutzfristenkategorie></code>: Angabe des relevanten Artikels in den gesetzlichen Grundlagen (falls anwendbar)</p> <p><code><schutzfrist></code>: Dauer der Schutzfrist in Jahren</p> <p>Das Metadatenpaar, das die Schutzfristen festhält, steht in drei Entitäten zur Verfügung:</p> <p>Ablieferung</p> <p>Ordnungssystemposition</p> <p>Dossier</p> <p><i>Empfehlung</i></p> <p><i>Es wird empfohlen, die Metadaten zu Datenschutz, Klassifizierung und Öffentlichkeitsstatus für Dossiers und Dokumente bereits im GEVER-System zu führen. Auf der Basis dieser Angaben können die Schutzfristen bei der Erstellung eines SIP hergeleitet und direkt in den Metadaten verzeichnet werden.</i></p>	K

ID	Beschreibung Anforderung	M/K
M_4.9-2	Wenn Schutzfristen erfasst werden, müssen diese entweder global für die gesamte Ablieferung (gleiche Schutzfrist für alle Unterlagen) oder pro Ordnungssystemposition oder pro Dossier in den Metadaten festgehalten werden.	K

4.10 Metadaten zu den Zeiträumen der Dossiers

ID	Beschreibung Anforderung	M/K
M_4.10-1	Sobald eine Angabe zu einem Entstehungszeitraum eines Dossiers (Metadatum <code><entstehungszeitraum></code> in der Entität <i>Dossier</i>) geschätzt wird, ist eine Begründung für diese Schätzung in das Metadatum <code><entstehungszeitraumAnmerkung></code> beim betroffenen Dossier einzutragen.	M

4.11 Metadaten zu den Prüfsummen der Dateien

ID	Beschreibung Anforderung	M/K
M_4.11-1	<p>Zu jeder Datei im Paket (ausser zur Datei <code>metadata.xml</code>) muss eine Prüfsumme berechnet und in den Metadaten zu dieser Datei im <code>metadata.xml</code> eingetragen werden. Dazu stehen die folgenden Metadaten in der Entität <i>Datei</i> zur Verfügung:</p> <p><code><pruefalgorithmus></code> : Bezeichnung des bei der Erstellung der Prüfsumme verwendeten Algorithmus.</p> <p><code><pruefsumme></code> : Prüfsummenwert, abhängig vom verwendeten Prüfalgorithmus</p> <p>Als Prüfalgorithmus können folgende Hash-Funktionen verwendet werden:</p> <p>MD5 SHA-1 SHA-256 SHA-512</p>	M

4.12 Referenzierung der Zugehörigkeit zu Dossiers in den Metadaten

ID	Beschreibung Anforderung	M/K
M_4.12-1	<p>Das Metadatum <dateiRef> dient dazu, den Zusammenhang zwischen den Dateien im Paket und deren Zugehörigkeit zu den Dokumenten in den Dossiers oder direkt zu einem Dossier herzustellen. Dieser Zusammenhang wird durch die entsprechende Referenzierung der <dateiRef> zu der entsprechenden Datei im Inhaltsverzeichnis des metadata.xml festgehalten.</p> <p>Diese Referenzierung ist die eindeutige Zuweisung der Inhalte der Dateien zu den Dossiers.</p> <pre> <paket> <inhaltsverzeichnis> <ordner> <name>header</name> ... </ordner> <ordner> <name>content</name> <ordner> <name>FLZ_1</name> <datei id="dat15"> <name>Modell_A.tif</name> <originalName>Modell_A.tif</originalName> <pruefalgorithmus>MD5</pruefalgorithmus> <pruefsumme>3b0b980e1a1bb76ee49abd9db66b0a47<pruefsumme> </datei> ... </ordner> </inhaltsverzeichnis> <ablieferung xsi:type="ablieferungFilesSIP"> <ablieferungstyp>Files</ablieferungstyp> ... <dossier id="dos1"> <titel>Flugzeuge 2009</titel> <erscheinungsform>digital</erscheinungsform> <entstehungszeitraum> <von> <datum>2009</datum> </von> <bis> <datum>2009</datum> </bis> </entstehungszeitraum> <bemerkung></bemerkung> <dateiRef>dat15</dateiRef> </dossier> ... </ablieferung> </paket> </pre> <p>Hinweis: In einem FILES-SIP kann ab der XSD-Version 4.0 eine Datei auch mehreren Dokumenten oder mehreren Dossiers gleichzeitig zugewiesen werden. Diese Möglichkeit sollte aber nur aufgrund eines fachlich zwingenden Grunds eingesetzt werden. Beispiel: Mehrere Dokumente beschreiben den Inhalt der gleichen Videodatei in aufeinanderfolgenden Abschnitten.</p>	M

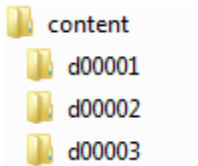
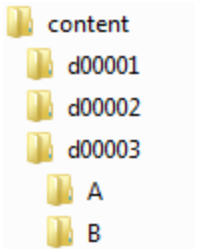
5 Struktur SIP

5.1 Rahmenvorgaben und Begrenzungen Grösse Paket

ID	Beschreibung Anforderung	M/K
S_5.1-1	<p>Ein SIP sollte maximal 8 GB gross sein.</p> <p>Empfehlung <i>Es wird empfohlen, aus Gründen der schnelleren Übertragung und Vermittlung, die Grösse eines einzelnen SIP für die Ablieferung unter 2 GB zu halten.</i> <i>Dies kann durch eine regelmässige Ablieferung von Unterlagen an das Archiv und durch eine gute Aussonderungsplanung erreicht werden.</i></p>	K
S_5.1-2	Die abliefernde Stelle muss vor der Erstellung der Ablieferung und dem Transfer des Paketes mit dem Archiv Kontakt aufnehmen, wenn das SIP grösser als 8 GB ist.	M
S_5.1-3	Die abliefernde Stelle muss vor der Erstellung der SIP mit dem Archiv Kontakt aufnehmen, wenn die abzuliefernden Unterlagen aus Gründen der Grösse auf mehrere SIP (und demzufolge auch auf mehrere Ablieferungen) aufgeteilt werden müssen.	M

5.2 Rahmenvorgaben und Begrenzungen Anzahl Dateien im Paket
















ID	Beschreibung Anforderung	M/K
S_5.2-1	Ein SIP darf nicht mehr als 1'000'000 Dateien enthalten.	M

ID	Beschreibung Anforderung	M/K
S_5.2-2	<p>Ein einzelner Ordner innerhalb des SIP sollte nicht mehr als 5'000 Dateien enthalten.</p> <p>Empfehlung <i>Die Anforderung sollte eingehalten werden, weil sonst Probleme mit der Darstellung des SIP auf dem Explorer auftreten können. Falls ein einzelner Ordner mehr als 5000 Dateien enthält, wird empfohlen, Ordner zur weiteren Unterteilung einzufügen.</i></p> <p>Beispiel</p>  <p>Der Ordner d00003/ enthält ursprünglich 9'000 Dateien.</p>  <p>Im Ordner d00003/ werden zwei Ordner eingefügt, und die 9'000 Dateien so darauf verteilt, dass kein Ordner mehr als 5'000 Dateien enthält. Die Benennung dieser Ordner ist grundsätzlich frei (hier A/ und B/), sollte aber innerhalb des Ordners d00003/ eindeutig und möglichst kurz gehalten werden, um die Pfadlänge nicht übermässig zu verlängern.</p>	K

5.3 Erlaubte Zeichen bei der Benennung der Ordner und Dateien

In den Namen der Dateien und Ordner darf nur ein begrenzter Satz von Zeichen verwendet werden. Dadurch werden falsche Konvertierungen innerhalb der Namen verhindert, die sonst auftreten könnten, wenn Pakete zwischen unterschiedlichen Betriebssystemen transferiert und verarbeitet werden.

ID	Beschreibung Anforderung	M/K
S_5.3-1	Die Vorgaben in Bezug auf den zu verwendenden Zeichensatz zur Benennung von Ordnern und Dateien müssen vollumfänglich eingehalten werden.	M

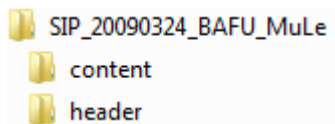
ID	Beschreibung Anforderung	M/K
S_5.3-2	<p>Der erlaubte Zeichensatz zur Benennung von Ordnern und Dateien ist ein Subset von US-ASCII.</p> <p>Die folgenden Zeichen sind erlaubt:</p> <p style="margin-left: 40px;">Buchstaben: A-Z a-z Ziffern: 0-9 Andere: ! # \$ % () + , - . = @ [] { } ~ _ SPACE</p> <p>Die Zeichencodes der erlaubten Zeichen sind in den Zeichensätzen US-ASCII, ISO-8859 und UTF-8 identisch.</p>	M
S_5.3-3	<p>Wenn die Namen von Ordnern und Dateien nicht erlaubte Zeichen enthalten, müssen diese vor der Ablieferung an das Archiv normalisiert werden.</p> <p>Empfehlung</p> <p><i>Im Anhang E finden sich Empfehlungen im Sinne einer Best Practice, die bei der Normalisierung von Namen befolgt werden können.</i></p>	M
S_5.3-4	<p>Wenn zwei Dateien oder Ordner, die im selben Ordner abgelegt sind, auf denselben Namen normalisiert werden (was eher selten der Fall ist), treten Kollisionen auf. Um diese Kollisionen zu vermeiden, müssen die betroffenen Dateien während oder vor der Normalisierung geeignet ergänzt werden, z. B. durch anhängen eines Suffixes an den Basisnamen.</p> <p>Der ursprüngliche Dateiname muss entsprechend zuvor in den Metadaten festgehalten werden und geht dadurch nicht verloren.</p> <p>Beispiel</p> <p>Kollision bei Normalisierung:</p> <div style="margin-left: 40px;">  Berufe_J  Jaeger.pdf  Jäger.pdf  Jäger.pdf →  Jaeger.pdf  Jaeger.pdf →  Jaeger.pdf </div> <p>Geeignete Ergänzung:</p> <div style="margin-left: 40px;">  Jäger.pdf →  Jäger_1.pdf →  Jaeger_1.pdf  Jaeger.pdf →  Jaeger_2.pdf →  Jaeger_2.pdf </div>	M
S_5.3-5	<p>Werden existierende Dateien im Paket bei der Paketerstellung (z. B. aus technischen Gründen) umbenannt, müssen die originalen Dateinamen im Metadaten-Element originalName in der Entität Datei eingetragen werden.</p> <p>Beispiel</p> <p>Umbenennung Dateiname bei Dateien im Paket</p> <div style="margin-left: 40px;">  Jäger.pdf →  Jaeger.pdf </div> <p>Einträge in Metadaten-Elementen im metadata.xml</p> <pre><dateiName>Jaeger.pdf</dateiName> <originalName>Jäger.pdf</originalName></pre>	M






5.4 Aufbau eines SIP

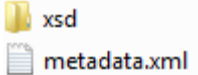
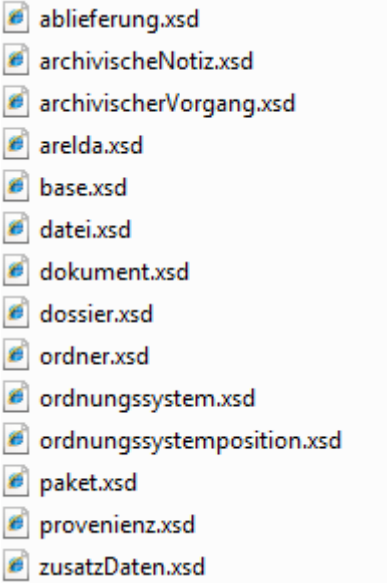
Die Basis des SIP ist eine Datei-Ordner-Struktur. Für diese gibt es sowohl Vorgaben über den Aufbau wie auch für die Benennungskonventionen (Inhalt und Aufbau der Namen) der einzelnen Ordner und Dateien. Als erste Übersicht folgen untenstehend die obersten Ordner eines SIP (Abb. 8).

```
SIP_[Identifikatoren]/
  header/
    metadata.xml
    xsd/
      arelda.xsd
      ...xsd
      ...xsd
  content/
```

Abb. 8: Grundstruktur eines Informationspaketes

ID	Beschreibung Anforderung	M/K
S_5.4-1	<p>Jedes Paket hat genau einen Toplevel-Ordner. Der Toplevel-Ordner ist der oberste Ordner, der den gesamten Inhalt eines einzelnen Paketes umfasst.</p> <p>Beispiele</p> <p>Der Toplevel-Ordner des Paketes heisst im untenstehenden Beispiel SIP_20090324_BAFU_MuLe.</p> 	M

ID	Beschreibung Anforderung	M/K						
S_5.4-2	<p>Der Name des Toplevel-Ordners beginnt bei einem SIP immer mit der Zeichenfolge <code>SIP_</code> gefolgt von bestimmten identifizierenden Angaben [Identifikatoren].</p> <p><i>Empfehlung</i> Der Name des Toplevel-Ordners soll gemäss dem folgenden Muster aufgebaut sein:</p> <table><tr><td><code>SIP_</code></td><td><code>[Ablieferungsdatum]</code></td><td><code>_</code></td><td><code>[Name der abliefernden Stelle]</code></td><td><code>_</code></td><td><code>[Referenz]</code></td></tr></table> <p><i>[Ablieferungsdatum]: Aufbau des Ablieferungsdatums: JJJJMMTT</i> <i>[Name der abliefernden Stelle]: Der Name der abliefernden Stelle wird mit der offiziellen Abkürzung angegeben (z. B. EPA, DEZA, BFS). Wenn keine offizielle Abkürzung vorhanden ist, ist eine kurze, aber sprechende Bezeichnung der abliefernden Stelle zu wählen.</i> <i>[Referenz]: Die Referenz kann durch die abliefernde Stelle optional verwendet werden. Hier kann z. B. das Kürzel des Namens der Kontaktperson bzw. der Datenbank oder die Ablieferungsnummer angegeben werden.</i></p> <p><i>Beispiel</i> Dargestellt sind 3 SIP, bei denen je der Toplevel-Ordner sichtbar ist.</p> <div> <code>SIP_20090324_BAFU_MuLe</code></div> <div> <code>SIP_20100425_DEZA_AmAn</code></div> <div> <code>SIP_20090612_BFS_DBWohnungen</code></div>	<code>SIP_</code>	<code>[Ablieferungsdatum]</code>	<code>_</code>	<code>[Name der abliefernden Stelle]</code>	<code>_</code>	<code>[Referenz]</code>	M
<code>SIP_</code>	<code>[Ablieferungsdatum]</code>	<code>_</code>	<code>[Name der abliefernden Stelle]</code>	<code>_</code>	<code>[Referenz]</code>			
S_5.4-3	<p>Der Toplevel-Ordner enthält immer genau zwei Ordner. Diese Ordner haben die Namen <code>header/</code> und <code>content/</code>.</p> <p>Es ist nicht erlaubt, andere Ordner oder Dateien direkt im Toplevel-Ordner zu platzieren.</p> <p>Beispiel Der Inhalt des Toplevel-Ordners sieht bei allen SIP folgendermassen aus:</p> <div> <code>content</code></div> <div> <code>header</code></div>	M						

ID	Beschreibung Anforderung	M/K
S_5.4-4	<p>Der Ordner <code>header/</code> enthält die Metadaten und deren XSD-Dateien. Der Ordner <code>header/</code> enthält niemals Primärdaten.</p> <p>Der vordefinierte Ordner im Ordner <code>header/</code> eines SIP für die Ablieferung heisst: <code>xsd/</code></p> <p>Es sind keine weiteren Ordner im Ordner <code>header/</code> erlaubt.</p> <p>Die vordefinierte Datei im Ordner <code>header/</code> heisst immer <code>metadata.xml</code>. Diese Datei ist die einzige Datei, die im Ordner <code>header/</code> erlaubt ist. In dieser Datei werden die Metadaten des Paketes gespeichert.</p> <p>Beispiel Der Inhalt des Ordners <code>header/</code> eines SIP, das abgeliefert wird.</p> 	M
S_5.4-5	<p>Alle XSD-Dateien die zur benutzten XSD-Version gehören, werden im Ordner <code>header/</code> in einem separaten Ordner mit dem Namen <code>xsd/</code> abgelegt. Die in <code>metadata.xml</code> referenzierte Schemadatei heisst zwingend <code>arelda.xsd</code> und wird als <code>schemaVersion="4.1"</code> referenziert</p> <p>Beispiel Alle XSD-Dateien für die Metadatenversion ARELDA_v4.1 im Ordner <code>xsd/</code>.</p> 	M
S_5.4-6	<p>Der Ordner <code>content/</code> enthält die Primärdaten eines SIP. Das heisst, der Ordner enthält alle Dateien, aus denen sich die Inhalte der Dossiers im SIP zusammensetzen.</p>	M

ID	Beschreibung Anforderung	M/K
S_5.4-7	<p>Es wird empfohlen, alle Ordner, die einem Dossier entsprechen, auf demselben Ordnerlevel innerhalb des Ordners <code>/content</code> zu platzieren.</p> <p>Beispiel Beispiel GEVER Die blau eingefärbten Ordner enthalten je Dateien die zum selben Dossier gehören. Ein Ordner (z. B. <code>d00001</code>) entspricht dabei einem Dossier.</p> <pre>content/ d00001/ d00002/ d00003/ d00004/ d00005/</pre> <p>Beispiel FILES Die blau eingefärbten Ordner enthalten je Dateien die zum selben Dossier gehören. Ein Ordner (z. B. <code>Baum_1</code>) entspricht dabei einem Dossier.</p> <pre>content/ teil_1/ Baum_1/ Baum_2/ teil_2/ Fisch_1/ Fisch_2/ Fisch_3/</pre>	K

5.5 Pfadlängen und Hierarchien im SIP









ID	Beschreibung Anforderung	M/K
S_5.5-1	<p>Die Pfadlänge zu jeder Datei und zu jedem Ordner innerhalb des Informationspaketes sollte weniger als 180 Zeichen betragen. Der Pfad beinhaltet dabei immer auch den Toplevel-Ordner. Auch die / müssen mitgezählt werden.</p> <p>Die Länge des Namens eines Ordners oder einer Datei sollte verkürzt werden, wenn der Pfad zu diesem Ordner oder zu dieser Datei über 180 Zeichen lang ist. Die Namen in einem Pfad müssen solange gekürzt werden, bis die Länge des gesamten Pfades weniger als 180 Zeichen beträgt.</p> <p>Beispiel</p> <p>SIP_20091220_EPA_hp/header/metadata.xml Dieser Pfad besteht aus 39 Zeichen.</p> <p>SIP_20091220_SBF_hp/content/d00001245/p00123453.pdf Dieser Pfad besteht aus 51 Zeichen.</p> <p>SIP_20091220_BFS_hp/content/orange_zettel/zettel_1.tif Dieser Pfad besteht aus 54 Zeichen.</p> <p>Empfehlung</p> <p><i>Diese Anforderung sollte eingehalten werden, weil sonst Probleme beim Export des SIPs in ein Dateisystems auftreten können</i></p>	K









5.6 Spezifische Vorgaben zu einem SIP für GEVER

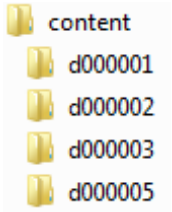
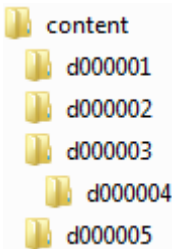
Zu den allgemeinen Vorgaben an die Struktur eines SIP gibt es für GEVER SIP noch spezifische Vorgaben, die nur für SIP aus GEVER Systemen gelten. Die Struktur eines SIP aus einem GEVER System ist in Abb. 9 ersichtlich.

<pre> SIP_20090130_BFS_hpm/ header/ metadata.xml xsd/ arelda.xsd ... content/ d00001/ p00001.pdf p00002.txt p00003.tif d00002/ p00004.pdf d00003/ p00005.pdf d00004/ p00006.pdf p00007.tif </pre>

Abb. 9: Struktur eines GEVER SIP

ID	Beschreibung Anforderung	M/K																							
S_5.6-1	Es wird im Sinne eines Best-Practice-Lösungsansatzes empfohlen, dass jedes Dossier und jedes Subdossier in der Ablieferung einem Ordner im Ordner <code>content/</code> oder tiefer entspricht.	K																							
S_5.6-2	<p>Es wird im Sinne eines Best-Practice-Lösungsansatzes empfohlen, dass bei jedem Ordner, der einem Dossier oder einem Subdossier entspricht, die Ordnernamen mit dem Buchstaben <code>d</code> beginnen. Alle Ordner werden mittels einer Laufnummer durchnummeriert.</p> <p>Muster:</p> <table><tr><td>d</td><td>n</td><td>n</td><td>n</td></tr></table> <table><tr><td>d</td><td>n</td><td>n</td><td>n</td><td>n</td><td>n</td><td>n</td></tr></table> <p>d: ist die Abkürzung für (Sub-)Dossier n: ist ein Integer mit einer innerhalb des Paketes fixen Anzahl Ziffern für die Laufnummer. Nicht benutzte Stellen werden mit einer 0 ausgefüllt. Die maximale Anzahl der Ziffern für die Laufnummer ist 6.</p> <p>Die Laufnummern (aus <code>d</code> und dem Integer) sind als Identifikatoren eindeutig und werden innerhalb des SIP nur einmal vergeben.</p> <p>Beispiel Ein Ordner, der einem Dossier oder einem Subdossier entspricht, wird wie folgt benannt:</p> <table><tr><td>d</td><td>n</td><td>n</td><td>n</td><td>n</td></tr></table> <p> <code>d0001</code>  <code>d0002</code>  <code>d0003</code>  <code>d0004</code></p> <p>oder</p> <table><tr><td>d</td><td>n</td><td>n</td><td>n</td><td>n</td><td>n</td><td>n</td></tr></table> <p> <code>d000001</code>  <code>d000002</code>  <code>d000003</code>  <code>d000004</code></p>	d	n	n	n	d	n	n	n	n	n	n	d	n	n	n	n	d	n	n	n	n	n	n	K
d	n	n	n																						
d	n	n	n	n	n	n																			
d	n	n	n	n																					
d	n	n	n	n	n	n																			

ID	Beschreibung Anforderung	M/K																																	
S_5.6-3	<p>Benennung von Dateien innerhalb des Ordners <code>content/</code>:</p> <p>Es wird im Sinne eines Best-Practice-Lösungsansatzes empfohlen, dass bei jeder Datei der Dateiname mit dem Buchstaben <code>p</code> beginnt. Alle Dateien werden mittels einer Laufnummer durchnummeriert. Die originalen Dateinamen sind dabei wie in Anforderung S_5.3-5 beschrieben in den Metadaten einzutragen und können auf diese Weise erhalten werden.</p> <p>Muster:</p> <table><tr><td>p</td><td>n</td><td>n</td><td>n</td><td>n</td><td>.</td><td>[Dateiendung]</td></tr></table> <table><tr><td>p</td><td>n</td><td>n</td><td>n</td><td>n</td><td>n</td><td>n</td><td>.</td><td>[Dateiendung]</td></tr></table> <p>p: ist die Abkürzung für (Primär-)Datei.</p> <p>n: ist ein Integer mit einer innerhalb des Paketes fixen Anzahl Ziffern für die Laufnummer. Nicht benutzte Stellen werden mit einer 0 ausgefüllt. Die maximale Anzahl der Ziffern für die Laufnummer ist 6.</p> <p>[Dateiendung]: Die Dateiendung gehört zum Dateinamen. Sie muss nach dem Punkt ebenfalls angegeben werden.</p> <p>Die Laufnummern (aus <code>p</code> und dem Integer) sind als Identifikatoren eindeutig und werden innerhalb des SIP nur einmal vergeben.</p> <p>Beispiel</p> <table><tr><td>p</td><td>n</td><td>n</td><td>n</td><td>n</td><td>n</td><td>.</td><td>[Dateiendung]</td></tr></table> <div> p00007.wav</div> <div> p00008.pdf</div> <div> p00009.txt</div> <div> p00010.tif</div> <p>oder</p> <table><tr><td>p</td><td>n</td><td>n</td><td>n</td><td>n</td><td>n</td><td>n</td><td>.</td><td>[Dateiendung]</td></tr></table> <div> p000011.wav</div> <div> p000008.pdf</div> <div> p000009.txt</div> <div> p000010.tif</div>	p	n	n	n	n	.	[Dateiendung]	p	n	n	n	n	n	n	.	[Dateiendung]	p	n	n	n	n	n	.	[Dateiendung]	p	n	n	n	n	n	n	.	[Dateiendung]	K
p	n	n	n	n	.	[Dateiendung]																													
p	n	n	n	n	n	n	.	[Dateiendung]																											
p	n	n	n	n	n	.	[Dateiendung]																												
p	n	n	n	n	n	n	.	[Dateiendung]																											

ID	Beschreibung Anforderung	M/K
S_5.6-4	<p>Alle Ordner, die Dossiers aus dem GEVER-System entsprechen, sind mit Vorteil auf der obersten Ebene im Ordner <code>content/</code> verzeichnet. Die Ordnungssystemstruktur sollte nicht in Ordnern abgebildet werden.</p> <p>Beispiel</p> <p>Oberste Ebene → Ordner für Dossiers im Ordner <code>content/</code>.</p>  <p>Weitere Ebenen → Ordner für Subdossiers</p>  <p>Der Ordner <code>d000004</code> entspricht einem Subdossier.</p>	K

5.7 Spezifische Vorgaben zu einem SIP für FILES

Zu den allgemeinen Vorgaben an die Struktur eines SIP gibt es für FILES SIP noch spezifische Vorgaben, die nur für SIP aus Datei-Ablagen oder für SIP aus relationalen Datenbanken gelten.

Die Struktur eines FILES SIP ist in Abb. 10 ersichtlich.

<pre> SIP_20080130_BAK_hanspeter.meier/ header/ metadata.xml xsd/ arelda.xsd content/ Kunstsammlung_Meier/ Einfuehrung/ Dokumentation.txt Fruehwerk/ Loewe.tif rote_Phase/ Kamel_gross.tif Kamel_klein.tif blaue_Phase/ Das_meer.tif dunkelblaue_Phase_Dezember1867_April1868/ der_ozean.tif Mittleres_Schaffen/ katze.tif Spaetwerk/ ein_Rentier.tif Notizbuecher/ Notizen_2000_2002.tif </pre>
--

Abb. 10: Struktur eines FILES SIP

ID	Beschreibung Anforderung	M/K
S_5.7-1	<p>Die ganze Ordnerstruktur der Primärdaten der Ablieferung ist im Ordner <code>content/</code> enthalten und abgebildet. Deshalb entsprechen die Ordner im <code>content/</code> nicht zwingend Dossiers und Subdossiers.</p> <p><i>Empfehlung</i> <i>Es wird im Interesse der Übersichtlichkeit empfohlen, in einem FILES SIP die Ordner, die Dossiers entsprechen, auf einer Hierarchiestufe darzustellen.</i></p>	M
S_5.7-2	<p>Für die Benennung von Ordnern und Dateien innerhalb des Ordners <code>content/</code> gibt es keine inhaltlichen Restriktionen oder Vorgaben. Die Namen der Ordner und Dateien können frei gewählt werden, solange sie den Anforderungen hinsichtlich Zeichensatz und Pfadlänge entsprechen.</p>	M

ID	Beschreibung Anforderung	M/K
S_5.7-3	Es ist die Aufgabe der abliefernden Stelle, die Zuordnung zwischen den Dateien im Paket und deren Gruppierung in den Dossiers im metadata.xml herzustellen. Dies geschieht technisch über die <dateiRef> im Teil <ablieferung>. Diese referenziert ein <datei>-Element im <inhaltsverzeichnis> über das id-Attribut. Im <datei>-Element sind Angaben zur Datei wie Name und Prüfsumme enthalten. Das <inhaltsverzeichnis> listet alle Ordner und Dateien der Ablieferung auf. Der Pfad einer Datei ergibt sich aus der Position in der Ordnerstruktur im Inhaltsverzeichnis. (Vgl. dazu Abb. 11).	M

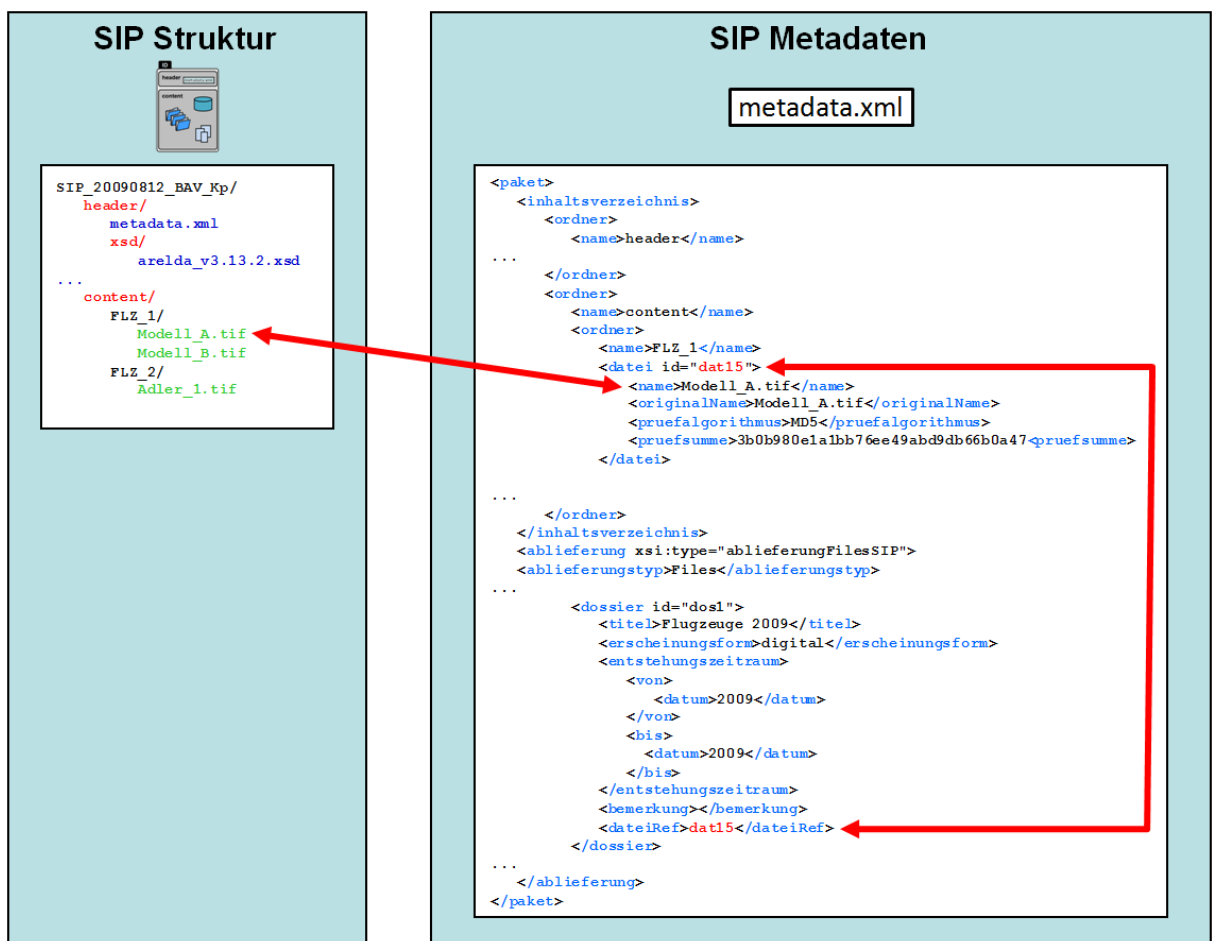


Abb. 11: Struktur eines FILES SIP – Referenzierung der Datei aus dem Dossier zur Dateibeschreibung im Inhaltsverzeichnis


5.8 Aufbau eines FILES SIP mit integrierter Dokumentation

Ein SIP einer Ablieferung mit Archivgut aus einer relationalen Datenbank oder aus einer Fachapplikation ist grundsätzlich ein SIP mit einer SIP Struktur für FILES Ablieferungen. Es gelten für ein solches SIP allerdings noch weitere spezifische Vorgaben. Vor allem ist es wichtig, dass das System, aus dem die Daten stammen (Datenbank oder Fachapplikation), in einer Dokumentation beschrieben wird.

Die Struktur eines solchen SIP (in diesem Fall ein SIP mit den Daten einer relationalen Datenbank, die im Format SIARD archiviert wurde) ist in Abb. 12 ersichtlich.

<pre> SIP_20070130_BFS_DB_Statistik_Verkehr header/ metadata.xml xsd/ arelda.xsd ... content/ 1_DOK/ Beschreibung_DB_Verkehr.pdf Datenmodell.pdf Daten_ListeA.pdf Daten_Liste1.pdf Daten_Liste2.pdf Daten_Liste3.pdf Schulungsunterlagen_2000.tif 2_DATEN/ Datenbank_Statistik_Verkehr.siard </pre>

Abb. 12: Struktur eines FILES SIP mit integrierter Dokumentation

ID	Beschreibung Anforderung	M/K
S_5.8-1	Die Dokumentation in einem FILES-SIP mit integrierter Dokumentation, muss im Ordner 1_DOK/ abgelegt werden.	M
S_5.8-2	Die Daten in einem FILES-SIP mit integrierter Dokumentation (z. B. die Daten einer archivierten relationalen Datenbank) müssen in einem Ordner mit dem Namen 2_DATEN/ abgelegt werden.	M
S_5.8-3	<p>In einem FILES SIP mit Archivgut aus einer relationalen Datenbank, die im SIARD-Format archiviert wurde, ist deshalb mindestens 1 Dossier enthalten, das dem Ordner 2_DATEN/ entspricht.</p> <p>Beispiel</p> <p>Bei FILES-SIP mit Daten aus Datenbanken, die mit SIARD archiviert wurden, enthält der Ordner 1_DOK/ die Dokumentation, der Ordner 2_DATEN/ das SIARD-Extrakt (Datei mit der Endung .siard).</p> 	M

6 Transfer

6.1 Format Package Container

ID	Beschreibung Anforderung	M/K
T_6.1-1	<p>Um einen schnelleren und einfacheren Umgang mit Informationspaketen zu ermöglichen, kann das SIP in einen Container (z. B. ZIP) verpackt werden.</p> <p>Das Containerformat muss vor dem Transfer zwischen abliefernder Stelle und Archiv vereinbart werden.</p>	K

7 Version und Gültigkeit der Spezifikation

Die vorliegende Version 1.1 ersetzt nicht die Version 1.0. Das heisst, dass die Version 1.0 nach wie vor gültig bleibt.

Dieser Minor Release Version 1.1 ist vollständig rückwärtskompatibel mit der Version 1.0, so dass keine Softwareänderungen an bestehenden Lösungen notwendig sind. In diesem Release wurden die eCH-inhärente Forderungen, die rückwärtskompatiblen RFCs, das Addendum zu eCH-0160 sowie das White Paper „Extraktion deskriptiver Metadaten aus der archivischen Ablieferungsschnittstelle eCH-0160“ aufgenommen. Details sind im Anhang E ersichtlich.

8 Change-Request-Prozess

Der Change-Request-Prozess wird nach eCH-0150 durchgeführt.

Die Rollen im Change Management mit dessen Aufgaben und Kompetenzen für den vorliegenden Standard eCH-0160 sind folgendermassen definiert:

Strategische Ebene: eCH-Fachgruppe Digitale Archivierung

Diese Ebene umfasst u.a. die langfristige Planung von neuen und die Weiterentwicklung bestehender eCH-Standards sowie die Definition der dafür notwendigen strategischen Ausrichtung der eCH-Fachgruppe und deren Tätigkeiten (Leitbild).

Taktische Ebene: Change Board entspricht der Themengruppe eCH-0160

Diese Ebene umfasst die mittelfristige Planung der Weiterentwicklung bestehender eCH-Standards (u.a. Versionenplanung) und die dafür notwendigen Entscheidungen über die Umsetzung oder Ablehnung von RFC unter Berücksichtigung der strategischen Vorgaben der eCH-Fachgruppe.

Operative Ebene: Change Manager ist die Geschäftsstelle der KOST

Diese Ebene umfasst die kurzfristige Planung der Weiterentwicklung (u.a. Qualifizierung von RFC, Sitzungsplanung Change Board) bestehender eCH-Standards und die operative Umsetzung von RFC aufgrund der Entscheidungen.

Gegenüber anderen eCH-Fachgruppen, der eCH-Geschäftsstelle oder anderen Dritten ist stets die eCH-Fachgruppe für das Change und Release Management – und somit für die Weiterentwicklung und die Governance von eCH-Standards verantwortlich.

9 Sicherheitsüberlegungen

Keine

10 Haftungsausschluss/Hinweise auf Rechte Dritter

eCH-Standards, welche der Verein eCH dem Benutzer zur unentgeltlichen Nutzung zur Verfügung stellt, oder welche eCH referenziert, haben nur den Status von Empfehlungen. Der Verein eCH haftet in keinem Fall für Entscheidungen oder Massnahmen, welche der Benutzer auf Grund dieser Dokumente trifft und / oder ergreift. Der Benutzer ist verpflichtet, die Dokumente vor deren Nutzung selbst zu überprüfen und sich gegebenenfalls beraten zu lassen. eCH-Standards können und sollen die technische, organisatorische oder juristische Beratung im konkreten Einzelfall nicht ersetzen.

In eCH-Standards referenzierte Dokumente, Verfahren, Methoden, Produkte und Standards sind unter Umständen markenrechtlich, urheberrechtlich oder patentrechtlich geschützt. Es liegt in der ausschliesslichen Verantwortlichkeit des Benutzers, sich die allenfalls erforderlichen Rechte bei den jeweils berechtigten Personen und/oder Organisationen zu beschaffen.

Obwohl der Verein eCH all seine Sorgfalt darauf verwendet, die eCH-Standards sorgfältig auszuarbeiten, kann keine Zusicherung oder Garantie auf Aktualität, Vollständigkeit, Richtigkeit bzw. Fehlerfreiheit der zur Verfügung gestellten Informationen und Dokumente gegeben werden. Der Inhalt von eCH-Standards kann jederzeit und ohne Ankündigung geändert werden.

Jede Haftung für Schäden, welche dem Benutzer aus dem Gebrauch der eCH-Standards entstehen ist, soweit gesetzlich zulässig, wegbedungen.

11 Urheberrechte

Wer eCH-Standards erarbeitet, behält das geistige Eigentum an diesen. Allerdings verpflichtet sich der Erarbeitende sein betreffendes geistiges Eigentum oder seine Rechte an geistigem Eigentum anderer, sofern möglich, den jeweiligen Fachgruppen und dem Verein eCH kostenlos zur uneingeschränkten Nutzung und Weiterentwicklung im Rahmen des Vereinszweckes zur Verfügung zu stellen.

Die von den Fachgruppen erarbeiteten Standards können unter Nennung der jeweiligen Urheber von eCH unentgeltlich und uneingeschränkt genutzt, weiterverbreitet und weiterentwickelt werden.

eCH-Standards sind vollständig dokumentiert und frei von lizenz- und/oder patentrechtlichen Einschränkungen. Die dazugehörige Dokumentation kann unentgeltlich bezogen werden.

Diese Bestimmungen gelten ausschliesslich für die von eCH erarbeiteten Standards, nicht jedoch für Standards oder Produkte Dritter, auf welche in den eCH-Standards Bezug genommen wird. Die Standards enthalten die entsprechenden Hinweise auf die Rechte Dritter.

Anhang A – Referenzen & Bibliographie

Begriff	Beschreibung
eCH-0150 [EAD]	Change und Release Management von eCH-Standards Encoded Archival Description, Version 2002 http://www.loc.gov/ead/
[eCH-0160]	Archivische Ablieferungsschnittstelle (SIP) http://www.ech.ch/vechweb/page?p=dossier&documentNumber=eCH-0160&documentVersion=1.0
[eCH-0160 DataDictionary]	Data Dictionary Archivische Ablieferungsschnittstelle (SIP) http://www.ech.ch/vechweb/page?p=dossier&documentNumber=eCH-0160&documentVersion=1.0
[I017]	I017 – GEVER Metadaten. Version 2.0 – ISB (Informatiksteuerungsorgan des Bundes) vom 28. Juni 2004 http://www.isb.admin.ch/themen/standards/alle/03228/
[ISAD(G)]	ISAD(G) – Internationale Grundsätze für die archivische Verzeichnung. Zweite, überarbeitete Ausgabe. Veröffentlichungen der Archivschule Marburg, Institut für Archivwissenschaft, Nr. 23, 2002 http://www.ica.org/download.php?id=1684 (Deutsche Übersetzung von ISAD(G): General International Standard Archival Description. Second Edition. 2000. http://www.ica.org/download.php?id=1687)
[Richtlinie ISAD(G)]	Verband Schweizerischer Archivarinnen und Archivare VSA (Hg.), Schweizerische Richtlinie für die Umsetzung von ISAD(G) – International Standard Archival Description (General). 2009. http://www.vsa-aas.org/fileadmin/user_upload/texte/ag_n_und_s/Richtlinien_ISAD_G_VSA_d.pdf
[xlsadg DataDictionary]	Rahmenspezifikation für die digitale Archivierung (Projekt bentō): Descriptive Information: Data Dictionary & XML Schema Beschreibung (xlsadg_DataDictionary_v1.6.pdf) http://kost-ceco.ch/cms/index.php?bento_de
[METS]	Metadata Encoding and Transmission Standard http://www.loc.gov/standards/mets/
[Matterhorn METS Profile]	Matterhorn METS Profile http://www.loc.gov/standards/mets/profiles/00000041.xml
[OAIS]	nestor (Hg.), Referenzmodell für ein Offenes Archiv-Informations-System – Deutsche Übersetzung 2.0. nestor-Materialien 16, 2013. http://nbn-resolving.de/urn/resolver.pl?urn=urn:nbn:de:0008-2013082706 (Deutsche Übersetzung von CCSDS 650.0-M-2, Reference Model for an Open Archival Information System (OAIS), Magenta Book, 2012. http://public.ccsds.org/publications/archive/650x0m2.pdf , inhaltsgleich mit ISO 14721:2012: Space data and information transfer systems -- Open archival information system (OAIS) -- Reference model. http://www.iso.org/iso/home/store/catalogue_tc/catalogue_detail.htm?csnumber=57284)

Anhang B – Mitarbeit & Überprüfung

- Marguérite Bos, Schweizerisches Bundesarchiv
- Georg Büchler, KOST
- Alain Dubois, Staatsarchiv Wallis
- Markus Lischer, Staatsarchiv Luzern
- Martin Kaiser, KOST
- Lambert Kansy, Staatsarchiv Basel-Stadt
- Ursina Rodenkirch-Brändli, Staatsarchiv Graubünden
- Claire Röthlisberger-Jourdan, KOST

Anhang C – Abkürzungen

Begriff	Beschreibung
AIP	Archival Information Package: AIP entstehen aus SIP im Laufe des Archivierungsprozesses der digitalen Unterlagen. AIP stellen diejenige Form der Informationspakete dar, in welcher die digitalen Unterlagen im digitalen Magazin gespeichert werden.
BAR	Schweizerisches Bundesarchiv
DIP	Dissemination Information Package: Ein DIP ist der Behälter für diejenigen Dossiers, welche von einem Benutzer in einem Bestellvorgang bestellt werden.
Dossier	Gesamtheit (Kollektiv) der Unterlagen zu einem Geschäft. Grundsätzlich entspricht ein Dossier einem Geschäft. Durch Zusammenfassen artverwandter Geschäfte bzw. durch Aufteilung von Dossiers in Subdossiers kann diese Grundstruktur aber den jeweiligen Bedürfnissen angepasst werden. Die Dossierbildung erfolgt auf der Grundlage des Ordnungssystems.
EAD	Encoded Archival Description. Ein dokumentarischer XML-Standard, der von der Library of Congress herausgegeben wird. Siehe Referenzen [EAD]
GEVER	Geschäftsverwaltung: Gesamtheit der Aktivitäten und Regeln für die Planung, Steuerung und Kontrolle sowie den Nachweis von Geschäften. Die GEVER umfasst die dokumenten- bzw. dossierbezogene Aktenführung (Führen von Geschäftsunterlagen) und das geschäftsbezogene Prozessmanagement. Die GEVER dient der wirksamen und effizienten Geschäftstätigkeit der Verwaltungseinheiten.
OAIS	Open Archival Information System: Als ISO 14721 verabschiedetes Referenzmodell beschreibt OAIS ein Archiv als Organisation, in dem Menschen und Systeme mit der Aufgabenstellung zusammenwirken, Informationen zu erhalten und einer definierten Nutzerschaft verfügbar zu machen. Siehe auch Referenzen [OAIS]
SIP	Submission Information Package: SIP sind Informationspakete, die von den aktenbildenden Stellen an das Archiv übermittelt werden. Sie enthalten die digitalen Unterlagen (Primärdaten und Metadaten).

Anhang D – Glossar

Begriff	Beschreibung
Ablieferung	Die Ablieferung bezeichnet sowohl den Prozess, bei dem eine Stelle Unterlagen dem Archiv übergibt, wie auch die Gesamtmenge der von einem Amt während einer Ablieferung übergebenen Unterlagen.
Abliefernde Stelle	Bezeichnung der Stelle bzw. Organisationseinheit, welche die Unterlagen an das Archiv abliefern. Die abliefernde Stelle ist häufig identisch mit dem Aktenbildner, dies ist jedoch nicht zwingend der Fall.
Attribut	Eine atomare Einheit innerhalb einer grösseren Einheit (Entität), hier von Metadaten. Ein solches einzelnes Metadatum wird auch Element (bei EAD) oder Feld (bei Datenbanken) genannt.
Aktenbildner	Bezeichnung der Stelle bzw. Organisationseinheit, welche die Unterlagen gebildet und geführt hat.
Archiv	1. Institution/Stelle, die Archivgut erfasst, aufbewahrt, konserviert und zugänglich macht. 2. Archivierte Unterlagen einer Organisation. 3. Gebäude oder Institution, das/die für die Archivierung von Unterlagen gebaut oder hergerichtet wurde.
Archivischer Vorgang	Ein archivischer Vorgang findet im Archiv statt. Er umfasst und dokumentiert Tätigkeiten im Bereich der Übernahme oder der Erhaltung.
archivtauglich	Archivtaugliche Dateiformate genügen den Anforderungen des Archivs für die Erhaltung der digitalen Unterlagen.
archivwürdig	Archivwürdige Unterlagen sind Unterlagen, die von juristischer oder administrativer Bedeutung sind oder einen grossen Informationswert haben (d.h. aus historischer, sozialer oder kultureller Sicht wertvoll sind).
Archivgut	Als Archivgut gelten Unterlagen, die vom Archiv zur Aufbewahrung übernommen worden sind oder von anderen Stellen nach den gleichen Grundsätzen selbständig archiviert werden.
Bestand	Gesamtheit des übernommenen Schriftguts einer Behörde oder Amtsstelle im Archiv. Ein Bestand besteht aus einer oder mehreren Ablieferungen der gleichen Behörde (Provenienz).
Datei-Ablage	Eine Datei-Ablage ist primär eine Menge von Dateien. Im Rahmen der vorliegenden Spezifikation wird der Begriff Datei-Ablage für Ablieferungen verwendet, die Dateien enthalten, die ohne ein Ordnungssystem im Sinne der Aktenführung mit einem GEVER-System abgeliefert werden. Die Dateien können aber durchaus gemäss einer anderen Systematik geordnet sein.
Digitales Archiv	Bezeichnung für die Institution, die digitales Archivgut bewertet, sichert, erschliesst und zugänglich macht.

Begriff	Beschreibung
Digitales Magazin	Aufbewahrungsort für digitales Archivgut.
Dossier	Als Dossier gilt die Gesamtheit (Kollektiv) der Unterlagen zu einem Geschäft. Grundsätzlich entspricht ein Dossier einem Geschäft. Durch Zusammenfassen artverwandter Geschäfte bzw. durch Aufteilung von Dossiers in Subdossiers kann diese Grundstruktur aber den jeweiligen Bedürfnissen angepasst werden. Die Dossierbildung erfolgt auf der Grundlage des Ordnungssystems.
Element	Bezeichnung für ein atomares Metadatum in EAD und andern Metadatenstandards, hier Attribut genannt.
Entität	Eine Einheit von Attributen, die eine Sache oder einen Sachverhalt beschreiben, bzw. ein Objekt, das durch Attribute bestimmt wird. In eCH-0160 ist das Objekt eine Verzeichnungsebene.
Metadaten	Metadaten können als «Informationen über die Primärdaten» (Daten über Daten) bezeichnet werden, da sie einen beschreibenden Charakter haben.
Metadatum	Atomare Einheit in einem Metadatenmodell.
Metadatenstandard	Normalisierung der Art und Weise, auf welche Art Metadaten eine gegebene Entität, Sache oder Sachverhalt beschreiben.
Primärdaten	Primärdaten sind Daten, die ausschliesslich direkt bei den einzelnen Aktenbildnern entstehen.
Serie	Gleichförmige Menge von Dossiers innerhalb eines Archivbestandes, synonym für Serie.
Schutzfrist	Über die Schutzfristen wird der Zugang zum Archivgut geregelt. Unterlagen, welche noch einer Schutzfrist unterstehen, dürfen (ausser von der jeweiligen Akten abliefernden Stelle selbst) nur nach einem entsprechenden Bewilligungsverfahren mit positivem Entscheid eingesehen werden.
Unterlagen	Unterlagen sind alle aufgezeichneten Informationen, unabhängig vom Informationsträger, welche bei der Erfüllung öffentlicher Aufgaben empfangen oder erstellt worden sind, sowie alle Hilfsmittel und ergänzenden Daten, die für das Verständnis dieser Informationen und deren Nutzung notwendig sind
Verzeichnungseinheit	Ein Set von Metadatenelementen, welches einen Verzeichnungsknoten auf einer beliebigen Stufe einer archivischen Tektonik beschreibt. Innerhalb einer Tektonik sind Verzeichnungseinheiten hierarchisch angeordnet.

Anhang E – Änderungen gegenüber Version 1.0

Folgende Änderungen wurden von der Version 1.0 zur Version 1.1 vorgenommen.

Kapitel / ID / Dokument	Anpassung	RFC
Kapitel 8	Change-Request-Prozess neu gemäss eCH-0150. Die Rollen wurden definiert.	2013-12
S_5.5-1	Die Einschränkung der Pfadlänge für Dateien im SIP auf 180 Zeichen war mit "Muss" zu restriktiv und wurde zu "Kann" gelockert.	2013-8
S_5.1-1ff	Die Beschränkung der Paketgrösse wurde mit "Kann" gelockert. Im Gegenzug muss bei Paketgrösse über 8GB oder bei gesplitteten SIPs vorgängig das Archiv informiert werden.	2013-9
M_4.9-1ff ablieferung.xsd dossier.xsd ordnungssystemposition.xsd	Die Metadaten zu den Schutzfristen sind neu optional. Im Gegenzug wurde ein separater Punkt M_4.9-2 erstellt, welcher die Muss-Anforderungen enthält. In den XSD-Dateien wurden nebst dieser Anpassung die Passagen rund um die Schutzfristen archivneutral formuliert.	2014-96
ablieferung.xsd	Die Dokumentation der Ablieferungsnummer wurde archivneutral formuliert und mit dem Hinweis ergänzt, dass Buchstaben enthalten sein dürfen.	2014-97
archivischerVorgang.xsd dokument.xsd	Archivneutrale Formulierungen in der Beschreibung.	2014-97
arelda.xsd	Implementierung Addendum_2014-03-25_eCH-0160_V1.0_ArchivischeAblieferungsschnittstelle	2014-35
Kapitel 4.2 und 4.3 dossier.xsd arelda.xsd DataDictionary	Die zusätzlichen, optionalen Entitäten <i>Vorgang</i> und <i>Aktivität</i> und die damit verbundenen Detailmetadaten wurden hinzugefügt, um Workflowmetadaten nach [I017] strukturiert abzubilden. Aus Gründen der Rückwärtskompatibilität mit Version 1.0 kann <i>Vorgang</i> auch weiterhin als Attribut der Entität <i>Dossier</i> verwendet werden.	2013-98
M_4.6-1 & M_4.6-2	XSD Schema Name und Schema Location genauer spezifiziert S_5.4-5	Vernehmlassung
S_5.4-5	Die Schemadatei heisst zwingend arelda.xsd	
Anhang A	Umfangreiche Ergänzungen.	White-Paper
Anhang C	Abkürzungen erweitert um: EAD.	

Kapitel / ID / Dokument	Anpassung	RFC
Anhang D	Glossar erweitert um: Ablieferung, Attribut, Bestand, Element, Entität, Metadatenstandard, Serie und Verzeichnungseinheit.	White-Paper
Anhang I	WhitePaper „Extraktion deskriptiver Metadaten aus der archivischen Ablieferungsschnittstelle eCH-0160“.	White-Paper

Anhang F – Gültigkeit Kapitel

Legende Farben

	ist für Ablieferungskategorie GEVER anzuwenden
	ist für Ablieferungskategorie FILES anzuwenden
	ist für Ablieferungskategorie FILES mit integrierter Dokumentation anzuwenden

X	Alle Anforderungen dieses Kapitels sind anzuwenden
P	Nur die Anforderungen, die explizit auf GEVER oder FILES verweisen, sind anzuwenden

Kapitel		Gültig für Ablieferungskategorien		
		GEVER	FILES	FILES mit integrierter Dokumentation
1	Einleitung			
1.1	Status des Dokuments	X	X	X
1.2	Ziel und Zweck	X	X	X
1.3	Struktur des Dokuments	X	X	X
1.3.1	Aufbau Kapitel	X	X	X
1.3.2	ID Anforderungen	X	X	X
1.3.3	Unterscheidung zwischen Muss und Kann-Anforderungen	X	X	X
1.3.4	Notation Ordner, Dateien und Ordnerstrukturen	X	X	X
1.4	Präzisierungsbedarf	X	X	X
2	Übersicht			
2.1	Abgrenzung	X	X	X
2.1.1	Definitionen	X	X	X
2.1.2	Anforderungen pro Ablieferungstyp	X	X	X
2.2	Anwendungsgebiet	X	X	X
2.2.1	Prozess	X	X	X
2.3	Ablieferungsobjekt Paket – SIP	X	X	X
2.4	Inhalt	X	X	X
2.5	Verhältnis Ablieferung – Paket	X	X	X
2.6	Sicherheitseinstellungen Dateien im Paket	X	X	X
3	Primärdaten			
3.1	Grundsätzliches	X	X	X
3.2	Formate	X	X	X
4	Metadaten			
4.1	Allgemeine Anforderungen	X	X	X
4.2	Konzeptuelles Datenmodell	X	X	X
4.3	Datenmodell Ablieferungstyp GEVER	X		
4.4	Datenmodell Ablieferungstyp FILES		X	X

Kapitel		Gültig für Ablieferungskategorien		
		GEVER	FILES	FILES mit integrierter Dokumentation
4.5	Data Dictionary	X	X	X
4.6	XSD	X	X	X
4.7	Aufbau Inhaltsverzeichnis im metadata.xml	X	X	X
4.8	Mapping Metadaten des Ursprungssystems auf SIP-Metadaten	P	P	P
4.9	Metadaten zu den Schutzfristen	X	X	X
4.10	Metadaten zu den Zeiträumen der Dossiers	X	X	X
4.11	Metadaten zu den Prüfsummen der Dateien	X	X	X
4.12	Referenzierung der Zugehörigkeit zu Dossiers in den Metadaten	X	X	X
5	STRUKTUR SIP			
5.1	Rahmenvorgaben und Begrenzungen Grösse Paket	X	X	X
5.2	Rahmenvorgaben und Begrenzungen Anzahl Dateien im Paket	X	X	X
5.3	Erlaubte Zeichen bei der Benennung der Ordner und Dateien	X	X	X
5.4	Aufbau eines SIP	X	X	X
5.5	Pfadlängen und Hierarchien im SIP	X	X	X
5.6	Spezifische Vorgaben zu einem SIP für GEVER	X		
5.7	Spezifische Vorgaben zu einem SIP für FILES		X	X
5.8	Aufbau eines FILES SIP mit integrierter Dokumentation			X
6	TRANSFER			
6.1	Format Package Container	X	X	X

Anhang G – Ausschnitt Mapping SIP-Metadaten – GEVER-System

Die unterstehende Tabelle zeigt ein **Beispiel**, wie ein Mapping zwischen den Metadaten für das SIP und den Metadaten des Herkunftssystems (z. B. eines GEVER-Systems) der zu archivierenden Unterlagen resp. deren Metadaten aussehen kann. Die darin enthaltenen Daten sind als Beispiele zu verstehen. Im unterstehenden Beispiel sind nur die ersten vier Metadaten der Entität Dokument aufgeführt. Das Mapping muss für alle relevanten Entitäten einer Ablieferung erstellt werden.

Entität – DOKUMENT – GEVER SIP							
XSD						GEVER-System	
Element-Name	Definition	muss (m) kann (k)	Daten- typ	Werte- bereich	mehrfach vorkommend	Mapping	auf Feld(er)
id	paketweit eindeutige ID (Primärschlüssel)	m	text		-	ja	COO-Id
titel	Kurze Beschreibung des im Dokument behandelten Gegenstandes	m	text	länge 3	-	ja	Eigenschaft (Feld XXXX)
autor	Verfasser eines Dokuments	k	text	länge 2	0..n	ja	zuständiger Sachbearbeiter (Feld XXXX)
erscheinungsform	Angabe, ob es sich beim Dokument zum Zeitpunkt der Ablieferung um ein digitales Dokument, oder um ein nicht-digitales Dokument handelt (Papier, audiovisuell). Ein Dokument kann nur einer der beiden Erscheinungsformen (entweder digital oder nicht-digital) zugewiesen werden. Dokumente, die vor der Ablieferung ins Archiv aus einem digitalen und einem nicht-digitalen Teil bestanden, müssen als zwei getrennte Dokumente abgeliefert werden.	m	enumeration	digital, nicht digital		fix	fixer Eintrag: digital Begründung: alle Unterlagen sind digital (keine analogen Teile)
dokumenttyp	Aussehen des Dokuments, das zu erhalten ist.	k	text	länge 3	-	nein	Begründung: kann für vorliegende Dokumente nicht spezifisch angegeben werden.
...

Anhang H – Zeichensätze

Dieser Anhang definiert die erlaubten Zeichen in Datei- bzw. Ordnernamen in SIP und gibt Normalisierungstabellen für einige häufig anzutreffende Zeichensätze an. Damit werden Inkompatibilitäten verhindert, die durch unterschiedliche Zeichensätze bzw. Sonderzeichen in verschiedenen Anwendungsumgebungen entstehen können. Die ursprünglichen Dateinamen müssen in den Metadaten festgehalten werden und gehen nicht verloren.

Der Einfachheit halber wird im Folgenden nur noch von Dateinamen gesprochen. Gemeint sind damit immer Datei- und Ordnernamen.

H.1: Codepage-Problematik

Eine *Codepage* ist eine Tabelle mit einer Zeichenkodierung verschiedener Zeichen. Auf Computern dargestellte Zeichen werden bei der Programmierung entsprechend einer Codepage codiert. Für die Codierung gibt es normalerweise 2^8 Zeichen, das sind insgesamt 256. In der Tabelle werden diese Zeichen von 0 bis 255 dargestellt. Die erste Hälfte (Zeichen 1 bis 126) ist dabei in fast allen Codepages identisch mit US-ASCII.

Auf jedem Rechner kann eine unterschiedliche Codepage installiert sein. Regionen bzw. Länder haben auf ihre Sprachen angepasste Codepages. Werden Dateien über sprachliche Grenzen hinaus ausgetauscht, die unterschiedliche Codepages verwenden, hat dies zur Folge, dass Texte mit Umlauten oder Akzentzeichen einen «Datensalat» verursachen, weil sie nicht richtig dargestellt werden. Dies kann auch beim Austausch zwischen Computern mit unterschiedlichen Betriebssystemen geschehen. Was für den Inhalt einer Datei gilt, trifft ebenso auf die Dateinamen zu.

Windows beispielsweise benutzt in westeuropäischen Ländern die Codepages 1252 und Unicode (UTF-16 encoding), Linux und Unix Systeme eher den ISO-8859-1 Standard.

Folgende Beispiele zeigen auf, wozu es führt, wenn Dateinamen unterschiedlich kodiert und in veränderten Umgebungen dargestellt werden:

Dateiname	Kreiert in Umgebung mit	Dargestellt in Umgebung mit	Ergebnis	Kommentar
René Zürcher	UTF-8	ISO-8859-1	RenÃ© ZÃ¼rcher	
René Zürcher	ISO-8859-1	ISO-8859-5	Renщ Zърcher	
René Zürcher	ISO-8859-1	UTF-8	Ren Zrcher Ren? Z?rcher	Illegale UTF-8 Sequenz. Die Zeichen werden dann oft mit ? oder gar nicht dargestellt

H.2: Dateinamen-Normalisierungen

Zur Vermeidung der genannten Problematik beim Arbeiten mit SIPs in Umgebungen mit unterschiedlichen Codepages müssen Dateinamen normalisiert werden. Als Basis wird der US-ASCII-Zeichensatz genommen (E.2.1). Mit einigen zusätzlichen Restriktionen für Dateina-

men (E.2.2), die durch die Betriebssysteme gegeben sind, werden nachfolgend Normalisierungstabellen zur Umwandlung der Dateinamen vorgeschlagen (E.2.3ff).

H.2.1 US-ASCII als Basiszeichensatz

UTF-8 wie auch alle ISO-8859 sind Supersets von US-ASCII, d.h. jede US-ASCII-Zeichenfolge ist ebenfalls eine gültige UTF-8-Zeichenfolge bzw. ISO-8859-Zeichenfolge. US-ASCII-Zeichenfolgen werden demnach immer gleich interpretiert und dargestellt, unabhängig davon, ob die Software für US-ASCII, UTF-8 oder ISO-8859-Verarbeitung konzipiert oder konfiguriert wurde (Mit Software sind hier nicht nur Applikationen wie Editor, Web-Browser etc. gemeint, sondern auch Betriebssysteme, Dateisystem-Treiber).

H.2.2 Erlaubte und reservierte (nicht-erlaubte) Zeichen in Dateinamen

Übersicht über Restriktionen bzw. spezielle Bedeutungen von Zeichen in Betriebssystemen und Sprachdefinitionen (Für eine ausführlichere Übersicht siehe Wikipedia-Eintrag «Filename» (<http://en.wikipedia.org/wiki/Filename>) mit detaillierten Beschreibungen und weiteren Referenzen.)

System	Gross-/ Kleinschreibung	reservierte Zeichen	Kommentare
WinXP NTFS	optional	< > : " / \ ? *	
WinXP HPFS	nicht relevant	< > : " / \ ? *	
Unix / Linux	relevant	/	
Mac OS HFS+	nicht relevant	: / (im Unix layer Mac OS X)	
XML	relevant	< > & ' "	
Uniform Resource Identifiers (URI)	relevant	: / ? # [] @ (generic component delimiters) ! \$ & ' () * + , ; = (sub-component delimiters)	wird nicht berücksichtigt

Für die Normalisierung von Dateinamen werden URIs zur Einschränkung der erlaubten Zeichen aus dem US-ASCII-Zeichensatz nicht berücksichtigt. Sollte ein Dateiname reservierte Zeichen eines URI enthalten und dies für die weitere Verarbeitung innerhalb eines URI wesentlich sein, wird der Dateiname durch percent-encoding (RFC 3986 Uniform Resource Identifier (URI): Generic Syntax) entsprechend angepasst.

H.2.2.1 Erlaubte Zeichen

Die erlaubten Zeichen befinden sich ausschliesslich im Bereich des US-ASCII-Zeichensatzes, das heisst, innerhalb 0x20 bis 0x7E. Jedoch ist nicht jedes US-ASCII-Zeichen erlaubt. Die folgende Tabelle listet die erlaubten Zeichen auf.

US-ASCII / ISO-8859-1	Unicode	Unicode Description	Symbol
0x20	U+0020	Space	SP
0x21	U+0021	Exclamation mark	!
0x23	U+0023	Number sign	#
0x24	U+0024	Dollar sign	\$
0x25	U+0025	Percent sign	%
0x28	U+0028	Left parenthesis	(
0x29	U+0029	Right parenthesis)
0x2B	U+002B	Plus sign	+
0x2C	U+002C	Comma	,
0x2D	U+002D	Hyphen/Minus sign	-
0x2E	U+002E	Full stop	.
0x30	U+0030	Digit Zero	0
...
0x39	U+0039	Digit Nine	9
0x3D	U+003D	Equal sign	=
0x40	U+0040	At sign	@
0x41	U+0041	Latin Capital letter A	A
...
0x5A	U+005A	Latin Capital letter Z	Z
0x5B	U+005B	Left Square Bracket	[
0x5D	U+005D	Right Square Bracket]
0x5F	U+005F	Low line	_
0x61	U+0061	Latin Small Letter A	a
...
0x7A	U+007A	Latin Small Letter Z	z
0x7B	U+007B	Left Curly Bracket	{
0x7D	U+007D	Right Curly Bracket	}
0x7E	U+007E	Tilde	~

H.2.2.2 Nicht erlaubte Zeichen

Nicht erlaubt sind diejenigen Zeichen, die per Definition in Betriebssystemen bzw. Dateisystemen für Dateinamen nicht erlaubt sind, sowie Zeichen, die bei Datenübertragungen zwischen unterschiedlichen Systemen Probleme verursachen können (alle Zeichen ausserhalb des US-ASCII-Bereichs).

US-ASCII / ISO-8859-1	Unicode	Unicode Description	Symbol
0x00	U+0000	Control Characters	
...	...		
0x1F	U+001F	Control Characters	
0x22	U+0022	Quotation mark	"
0x26	U+0026	Ampersand	&
0x27	U+0027	Apostrophe	'
0x2A	U+002A	Asterisk	*
0x2F	U+002F	Slash	/
0x3A	U+003A	Colon	:
0x3B	U+003B	Semicolon	;
0x3C	U+003C	Less-than sign	<
0x3E	U+003E	Greater-than sign	>
0x3F	U+003F	Question mark	?
0x5C	U+005C	Backslash	\

US-ASCII / ISO-8859-1	Unicode	Unicode Description	Symbol
0x5E	U+005E	Circumflex accent	^
0x60	U+0060	Grave accent	`
0x7C	U+007C	Vertical bar	
	U+007F	Delete	
...	

Zeichencodes, die über dem Wert 0x7F bzw. U+007F liegen, sind nicht erlaubt.

H.2.3 Normalisierung US-ASCII im Bereich 0x20–0x7E

Die Steuerzeichen (0x00..0x1F, 0x7F) sind in Dateinamen illegal und werden für die Normalisierung nicht weiter berücksichtigt. Sollten sie trotzdem auftreten, sind sie zu eliminieren und es ist eine Fehlermeldung zu generieren. In Dateinamen nicht erlaubte Zeichen (siehe Kapitel E.2.2.2) werden bei der Normalisierung auf «_» abgebildet, erlaubte Zeichen bleiben gleich:

Von...				Nach...		
US-ASCII / ISO-8859-1	Unicode	Unicode Description	Symbol	Unicode	Unicode Description	Symbols
0x20	U+0020	Space	SP	U+005F	Space	SP
0x21	U+0021	Exclamation mark	!	U+0021	Exclamation mark	!
0x22	U+0022	Quotation mark	"	U+005F	Low Line	_
0x23	U+0023	Number sign	#	U+0023	Number sign	#
0x24	U+0024	Dollar sign	\$	U+0024	Dollar sign	\$
0x25	U+0025	Percent sign	%	U+0025	Percent sign	%
0x26	U+0026	Ampersand	&	U+005F	Low Line	_
0x27	U+0027	Apostrophe	'	U+005F	Low Line	_
0x28	U+0028	Left parenthesis	(U+0028	Left parenthesis	(
0x29	U+0029	Right parenthesis)	U+0029	Right parenthesis)
0x2A	U+002A	Asterisk	*	U+005F	Low Line	_
0x2B	U+002B	Plus sign	+	U+002B	Plus sign	+
0x2C	U+002C	Comma	,	U+002C	Comma	,
0x2D	U+002D	Hyphen/Minus sign	-	U+002D	Hyphen/Minus sign	-
0x2E	U+002E	Full stop	.	U+002E	Full stop	.
0x2F	U+002F	Slash	/	U+005F	Low Line	_
0x30	U+0030	Digit Zero	0	U+0030	Digit Zero	0

0x39	U+0039	Digit Nine	9	U+0039	Digit Nine	9
0x3A	U+003A	Colon	:	U+005F	Low Line	_
0x3B	U+003B	Semicolon	;	U+005F	Low Line	_
0x3C	U+003C	Less-than sign	<	U+005F	Low Line	_
0x3D	U+003D	Equal sign	=	U+003D	Equal sign	=
0x3E	U+003E	Greater-than sign	>	U+005F	Low Line	_
0x3F	U+003F	Question mark	?	U+005F	Low Line	_
0x40	U+0040	At sign	@	U+0040	At sign	@
0x41	U+0041	Latin Capital letter A	A	U+0041	Latin Capital letter A	A

0x5A	U+005A	Latin Capital letter Z	Z	U+005A	Latin Capital letter Z	Z
0x5B	U+005B	Left Square Bracket	[U+005B	Left Square Bracket	[
0x5C	U+005C	Backslash	\	U+005F	Low Line	_
0x5D	U+005D	Right Square Bracket]	U+005D	Right Square Bracket]
0x5E	U+005E	Circumflex accent	^	U+005F	Low Line	_
0x5F	U+005F	Low line	_	U+005F	Low Line	_
0x60	U+0060	Grave accent	`	U+005F	Low Line	_
	U+0061	Latin Small Letter A	a	U+0061	Latin Small Letter A	a

	U+007A	Latin Small Letter Z	z	U+007A	Latin Small Letter Z	z

Von...				Nach...		
US-ASCII / ISO-8859-1	Unicode	Unicode Description	Symbol	Unicode	Unicode Description	Symbols
0x7B	U+007B	Left Curly Bracket	{	U+007B	Left Curly Bracket	{
0x7C	U+007C	Vertical bar		U+005F	Low Line	_
0x7D	U+007D	Right Curly Bracket	}	U+007D	Right Curly Bracket	}
0x7E	U+007E	Tilde	~	U+007E	Tilde	~

H.2.4 Normalisierung Codepage-1252, ISO-8859 und Unicode im Bereich 0x80–0x9F (U+0080 – U+009F)

Codepage-1252 wird als *Westeuropäisch (Western European)* bezeichnet. Sie weicht von ISO-8859-1 im Bereich 0x80–9F ab, dessen 32 Positionen hier 27 darstellbare Zeichen beinhalten, u. a. die in ISO 8859-15 hinzugekommenen und einige für bessere Typographie notwendige. Bei den nicht darstellbaren Zeichen bleibt in der unterstehenden Tabelle das Feld *Symbol* leer. Die Unterschiede zwischen all diesen Kodierungen sowie die generell mangelnde Konsequenz bei der Unterstützung verschiedener Zeichensätze sind ein häufiges Interoperabilitätsproblem.

Die Zeichen (0x80..0x9F bzw. U+0080..U+009F) sind in ISO-8859 und in Unicode Steuerzeichen. Sie sind in Dateinamen illegal und werden für die Normalisierung nicht weiter berücksichtigt. Sollten sie trotzdem auftreten, sind sie zu eliminieren und es ist eine Fehlermeldung zu generieren.

Die folgende Tabelle enthält die Umsetzung der betroffenen Zeichen in *visuell ähnliche* US-ASCII Zeichen:

Von...		Nach...	
CP-1252	SYMBOL	US-ASCII	SYMBOLS
0x80	€	0x45 0x3D	E=
0x81			
0x82	,	0x27	'
0x83	f	0x66	f
0x84	„	0x27	'
0x85	...	0x2E 0x2E 0x2E	...
0x86	†	0x5F	_
0x87	‡	0x5F	_
0x88	^	0x5F	_
0x89	‰	0x25 0x30	%0
0x8A	Š	0x53	S
0x8B	‹	0x27	'
0x8C	Œ	0x4F 0x45	OE
0x8D		0x5F	_
0x8E	Ž	0x5A	Z
0x8F		0x5F	_
0x90		0x5F	_
0x91	‘	0x27	'
0x92	’	0x27	'
0x93	“	0x27	'
0x94	”	0x27	'
0x95	•	0x5F	_
0x96	–	0x2D 0x2D	--
0x97	—	0x2D 0x2D 0x2D	---
0x98	~	0x7E	~
0x99	™	0x54 0x4D	TM
0x9A	š	0x73	s

Von...		NACH...	
CP-1252	SYMBOL	US-ASCII	SYMBOLS
0x9B	›	0x27	'
0x9C	œ	0x6F 0x65	oe
0x9D		0x5F	–
0x9E	ž	0x7A	z
0x9F	ÿ	0x59	Y

H.2.5 Normalisierung Codepage-1252, ISO-8859 und Unicode im Bereich 0xA0–0xFF (U+00A0 – U+00FF)

Diese und die folgenden Mappings bilden Zeichen aus UTF-8 (Teilmenge) und ISO-8859, die nicht im US-ASCII-Bereich liegen, auf *visuell ähnliche* US-ASCII-Zeichen ab.

Die relevanten und erlaubten Code Pages des ISO-8859 Standards sind:

Code Page 1 Latin-1, Westeuropäisch

Code Page 15 Latin-9, Westeuropäisch

Die folgende Tabelle enthält das Mapping von ISO-8859-1-Zeichen und Unicode-Zeichen ausserhalb des US-ASCII-Bereiches (0xA0 – 0xFF) in visuell ähnliche US-ASCII Zeichen bzw. Zeichenfolgen unter Berücksichtigung der Einschränkungen des obigen Abschnittes E.2.2.1.

Von...				Nach...	
ISO-8859-1	Unicode	Unicode Description	Symbol	US-ASCII	Symbols
0xA0	U+00A0	Non-breaking space		0x20	SP
0xA1	U+00A1	Inverted Exclamation Mark	¡	0x5F	_
0xA2	U+00A2	Cent sign	¢	0x63	c
0xA3	U+00A3	Pound sign	£	0x4C 0x3D	L=
0xA4	U+00A4	Currency sign	¤	0x49 0x3D	I=
0xA5	U+00A5	Yen sign	¥	0x59 0x3D	Y=
0xA6	U+00A6	Broken bar	¦	0x5F	_
0xA7	U+00A7	Section sign	§	0x53 0x53	SS
0xA8	U+00A8	Diaeresis	¨	0x5F	_
0xA9	U+00A9	Copyright sign	©	0x28 0x63 0x29	(c)
0xAA	U+00AA	Feminine Ordinal Indicator	ª	0x61	a
0xAB	U+00AB	Left-pointing double angle quotation mark	«	0x5F	_
0xAC	U+00AC	Not sign	¬	0x5F	_
0xAD	U+00AD	Soft hyphen	­	0x5F	_
0xAE	U+00AE	Registered sign	®	0x28 0x72 0x29	(r)
0xAF	U+00AF	Macron	¯	0x5F	_
0xB0	U+00B0	Degree symbol	°	0x64 0x65 0x67	deg
0xB1	U+00B1	Plus-minus sign	±	0x2B 0x2D	+–
0xB2	U+00B2	Superscript two	²	0x32	2
0xB3	U+00B3	Superscript three	³	0x33	3
0xB4	U+00B4	Acute accent	´	0x5F	_
0xB5	U+00B5	Micro sign	µ	0x75	u
0xB6	U+00B6	Pilcrow sign	¶	0x50	P
0xB7	U+00B7	Middle dot	·	0x2E	.
0xB8	U+00B8	Cedilla	¸	0x2C	,
0xB9	U+00B9	Superscript one	¹	0x31	1
0xBA	U+00BA	Masculine ordinal indicator	º	0x6F	o
0xBB	U+00BB	Right-pointing double-angle quotation mark	»	0x5F	_
0xBC	U+00BC	Vulgar fraction one quarter	¼	0x5F	_

Von...				Nach...	
ISO-8859-1	Unicode	Unicode Description	Symbol	US-ASCII	Symbols
0xBD	U+00BD	Vulgar fraction one half	½	0x5F	_
0xBE	U+00BE	Vulgar fraction three quarters	¾	0x5F	_
0xBF	U+00BF	Inverted Question Mark	¿	0x5F	_
0xC0	U+00C0	Latin Capital Letter A with grave	À	0x41	A
0xC1	U+00C1	Latin Capital letter A with acute	Á	0x41	A
0xC2	U+00C2	Latin Capital letter A with circumflex	Â	0x41	A
0xC3	U+00C3	Latin Capital letter A with tilde	Ã	0x41	A
0xC4	U+00C4	Latin Capital letter A with diaeresis	Ä	0x41 0x65	Ae
0xC5	U+00C5	Latin Capital letter A with ring above	Å	0x41	A
0xC6	U+00C6	Latin Capital letter AE	Æ	0x41 0x65	Ae
0xC7	U+00C7	Latin Capital letter C with cedilla	Ç	0x43	C
0xC8	U+00C8	Latin Capital letter E with grave	È	0x45	E
0xC9	U+00C9	Latin Capital letter E with acute	É	0x45	E
0xCA	U+00CA	Latin Capital letter E with circumflex	Ê	0x45	E
0xCB	U+00CB	Latin Capital letter E with diaeresis	Ë	0x45	E
0xCC	U+00CC	Latin Capital letter I with grave	Ì	0x49	I
0xCD	U+00CD	Latin Capital letter I with acute	Í	0x49	I
0xCE	U+00CE	Latin Capital letter I with circumflex	Î	0x49	I
0xCF	U+00CF	Latin Capital letter I with diaeresis	Ï	0x49	I
0xD0	U+00D0	Latin Capital letter Eth	Ð	0x44	D
0xD1	U+00D1	Latin Capital letter N with tilde	Ñ	0x4E	N
0xD2	U+00D2	Latin Capital letter O with grave	Ò	0x4F	O
0xD3	U+00D3	Latin Capital letter O with acute	Ó	0x4F	O
0xD4	U+00D4	Latin Capital letter O with circumflex	Ô	0x4F	O
0xD5	U+00D5	Latin Capital letter O with tilde	Õ	0x4F	O
0xD6	U+00D6	Latin Capital letter O with diaeresis	Ö	0x4F 0x65	Oe
0xD7	U+00D7	Multiplication sign	×	0x78	x
0xD8	U+00D8	Latin Capital letter O with stroke	Ø	0x4F	O
0xD9	U+00D9	Latin Capital letter U with grave	Ù	0x55	U
0xDA	U+00DA	Latin Capital letter U with acute	Ú	0x55	U
0xDB	U+00DB	Latin Capital Letter U with circumflex	Û	0x55	U
0xDC	U+00DC	Latin Capital Letter U with diaeresis	Ü	0x55 0x65	Ue
0xDD	U+00DD	Latin Capital Letter Y with acute	Ý	0x59	Y
0xDE	U+00DE	Latin Capital Letter Thorn	Þ	0x54 0x68	Th
0xDF	U+00DF	Latin Small Letter sharp S	ß	0x73 0x73	ss
0xE0	U+00E0	Latin Small Letter A with grave	à	0x61	a
0xE1	U+00E1	Latin Small Letter A with acute	á	0x61	a
0xE2	U+00E2	Latin Small Letter A with circumflex	â	0x61	a
0xE3	U+00E3	Latin Small Letter A with tilde	ã	0x61	a
0xE4	U+00E4	Latin Small Letter A with diaeresis	ä	0x61 0x65	ae
0xE5	U+00E5	Latin Small Letter A with ring above	å	0x61	a
0xE6	U+00E6	Latin Small Letter AE	æ	0x61 0x65	ae
0xE7	U+00E7	Latin Small Letter C with cedilla	ç	0x63	c
0xE8	U+00E8	Latin Small Letter E with grave	è	0x65	e
0xE9	U+00E9	Latin Small Letter E with acute	é	0x65	e
0xEA	U+00EA	Latin Small Letter E with circumflex	ê	0x65	e
0xEB	U+00EB	Latin Small Letter E with diaeresis	ë	0x65	e
0xEC	U+00EC	Latin Small Letter I with grave	ì	0x69	i
0xED	U+00ED	Latin Small Letter I with acute	í	0x69	i
0xEE	U+00EE	Latin Small Letter I with circumflex	î	0x69	i
0xEF	U+00EF	Latin Small Letter I with diaeresis	ï	0x69	i
0xF0	U+00F0	Latin Small Letter Eth	ð	0x64	d
0xF1	U+00F1	Latin Small Letter N with tilde	ñ	0x6E	n
0xF2	U+00F2	Latin Small Letter O with grave	ò	0x6F	o
0xF3	U+00F3	Latin Small Letter O with acute	ó	0x6F	o
0xF4	U+00F4	Latin Small Letter O with circumflex	ô	0x6F	o
0xF5	U+00F5	Latin Small Letter O with tilde	õ	0x6F	o
0xF6	U+00F6	Latin Small Letter O with diaeresis	ö	0x6F	oe

Von...				Nach...	
ISO-8859-1	Unicode	Unicode Description	Symbol	US-ASCII	Symbols
0xF7	U+00F7	Division sign	÷	0x5F	_
0xF8	U+00F8	Latin Small Letter O with stroke	ø	0x6F	o
0xF9	U+00F9	Latin Small Letter U with grave	ù	0x75	u
0xFA	U+00FA	Latin Small Letter U with acute	ú	0x75	u
0xFB	U+00FB	Latin Small Letter U with circumflex	û	0x75	u
0xFC	U+00FC	Latin Small Letter U with diaeresis	ü	0x75 0x65	ue
0xFD	U+00FD	Latin Small Letter Y with acute	ý	0x79	y
0xFE	U+00FE	Latin Small Letter Thorn	þ	0x74 0x68	th
0xFF	U+00FF	Latin Small Letter Y with diaeresis	ÿ	0x79	y

Für die Zeichen des ISO-8859-15, die sich von ISO-8859-1 unterscheiden, ist nach dem gleichen Rezept wie oben vorzugehen.

H.2.6 Normalisierung von Unicode ab dem Zeichen U+0100

Treten Zeichen in Dateinamen auf, die Unicodes U+0100 oder grösser sind, sind diese alle in erlaubte US-ASCII-Zeichen bzw. Zeichensequenzen umzuwandeln. Dabei ist nach dem gleichen Rezept wie oben vorzugehen.

Anhang I – Beziehung zu weiteren archivischen Standards

I.1 Einleitung

Dieser Anhang beschreibt einen *Crosswalk* beziehungsweise eine Metadaten transformation zwischen der archivischen Ablieferungsschnittstelle [eCH-0160] und dem massgeblichen archivischen Verzeichnungsstandard [ISAD(G)], respektive seiner Kodierung in [EAD]. Das Ziel ist nicht die vollständige Abbildung aller Metadaten der Ablieferungsschnittstelle eCH-0160, sondern das Extrahieren der Erschliessungsinformationen, im OAIS Modell als *Descriptive Information* bezeichnet. Der Anhang basiert auf dem WhitePaper „Extraktion deskriptiver Metadaten aus der archivischen Ablieferungsschnittstelle eCH-0160“, das am 7. 4. 2014 von der eCH-Fachgruppe Digitale Archivierung veröffentlicht wurde.

Jeder Crosswalk oder Metadaten transformation beinhaltet eine gewisse Willkürlichkeit bei der Wahl der Abbildung, d.h. es ist nicht in allen Fällen eine eindeutige Abbildung möglich. Dieser Anhang beschreibt deshalb nicht einen allgemeinen Crosswalk zwischen archivischen Standards, sondern fokussiert auf die Metadaten transformation von real implementierten Interpretationen dieser Standards bei Schweizer Archiven, nämlich ISAD(G) in der Ausprägung xlsadg der *bentō*-Rahmenspezifikation der KOST [xlsadg] mit einer Implementierung im Staatsarchiv Basel-Stadt und EAD im [Matterhorn METS Profile], wie es von der Firma docuteam und dem Staatsarchiv Wallis bei der *Library of Congress* eingereicht und auch implementiert worden ist.

Bei der Transformation ist nicht die vollständige Abbildung aller Metadaten das Ziel, sondern die Abbildung aller Erschliessungsinformationen, die im OAIS Modell als *Descriptive Information* bezeichnet sind. Metadaten der Ablieferungsebene bleiben unberücksichtigt.

Als Beilagen zum Anhang sind die beiden Extraktionen als XSL-Transformationen realisiert; dies um die technische Umsetzung im Detail zu verdeutlichen.

I.1.1 Anwendungsgebiet

Die hier beschriebenen Metadaten transformationen sind für alle Institutionen nützlich, welche digitale Unterlagen in SIPs nach eCH-0160 zur Archivierung übernehmen und die archivischen Metadaten in ein Verzeichnis, Findmittel oder Archivinformationssystem (im OAIS-Modell auch *Data Management* genannt) entweder im EAD- oder im xlsadg-Format übernehmen möchten. Die Transformationen können auch dienlich sein, wenn das Ziel die Transformation von eCH-0160-SIPs in AIPs mit entsprechender Metadatenstruktur ist.

I.1.2 Vorteile

Die Extraktion deskriptiver Metadaten aus SIPs trägt zur Lösung zweier miteinander verbundener Probleme der digitalen Archivierung bei:

- Die Erzeugung von Metadaten ist aufwändig. Naheliegend ist deshalb die Weiterverwendung bereits existierender Metadaten. Die Extraktion von Metadaten aus den Ursprungssystemen und ihre Übergabe an ein Archivinformationssystem über eine Ab-

lieferung als eCH-0160-SIP ermöglicht es, weitgehend von Vorleistungen zu profitieren.

- Die Archivierung grosser Mengen von digitalen Unterlagen, wie sämtliche Archivinstitutionen sie zukünftig realisieren müssen, ist ohne eine möglichst weitgehende Automatisierung nicht möglich. Die Extraktion von Metadaten erlaubt es, eine minimale oder eine darüber hinausgehende archivische Verzeichnung mit geringem individuellem Zusatzaufwand zu realisieren.

Die Normierung der Metadaten transformation vereinfacht die findmittelübergreifende Suche, wenn in jedem ISAD(G)- oder EAD-Feld jeweils die gleichen Metadaten aus der Ablieferung zu finden sind.

Ein Metadatenmapping ist eine aufwändige und konfliktbehaftete Arbeit; eine Vorgabe hilft hier eindeutig, Zeit und Ressourcen zu sparen.

I.1.3 Beilagen

Zum Anhang I „Beziehung zu weiteren archivischen Standards“ gehören zwei weitere integrale Beilagen: eine Excel-Datei mit den Transformationen in tabellarischer Form (eCH-0160_xlsadg&EAD.xlsx) und eine ZIP-Datei mit den in XSL realisierten Transformationen (xsl-Transformation.zip).

I.2 Kontext

I.2.1 OAIS

Das international massgebende Referenzmodell für die digitale Archivierung ist das *Reference Model for an Open Archival Information System* OAIS. Es umfasst insbesondere ein Datenmodell und ein funktionales Modell.

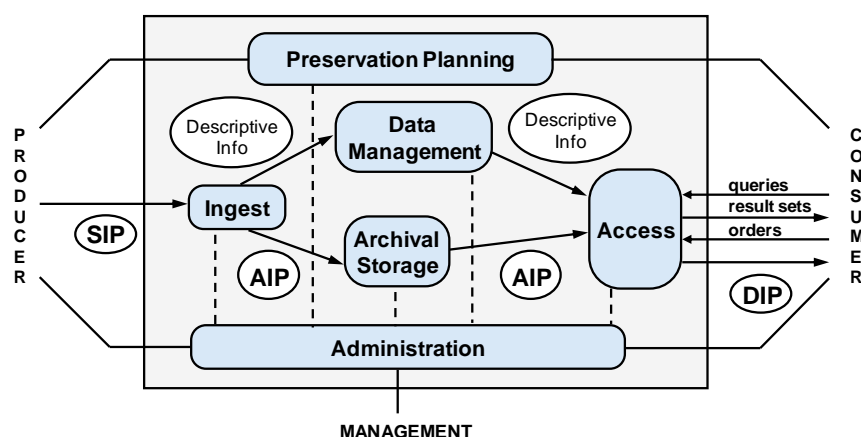


Abb. I.1: Das funktionale Modell von OAIS

Das funktionale Modell des OAIS unterscheidet sechs Funktionsbereiche: *Ingest* (Übernahme), *Data Management* (Datenverwaltung, unter anderem Verwaltung der Erschliessungsinformationen), *Archival Storage* (Archivspeicher), *Access* (Zugang), *Administration* (Administration) und *Preservation Planning* (Erhaltungsplanung).

Die hier beschriebene Extraktion deskriptiver Metadaten gehört in den Funktionsbereich *In-gest*, genauer zur Funktion *Generate Descriptive Information* (Erschliessungsinformationen erzeugen).

I.2.2 Existierende und verwendete Standards

Das OAIS ist auf einem hohen konzeptionellen Niveau angesiedelt. Seine Datenmodelle und Funktionen sind für die Implementierung im einzelnen zu spezifizieren. In diesem Kontext sind die folgenden nationalen und internationalen Standards von Bedeutung.

I.2.2.1 eCH-0160, Archivische Ablieferungsschnittstelle (SIP)

Als Standard für die archivische Ablieferungsschnittstelle (SIP) steht eCH-0160 zur Verfügung. Dieser Standard definiert das Datenmodell eines SIPs und ermöglicht die strukturierte Ablieferung von Primär- und Metadaten aus GEVER-Systemen, Datensammlungen und Datenbanken an ein Archiv. Das SIP gemäss eCH-0160 gestattet insbesondere die Übernahme und Weiterverwendung der im Ursprungssystem erzeugten und abgelegten Metadaten.

I.2.2.2 General International Standard Archival Description ISAD(G)

Der vom *International Council on Archives* ICA entwickelte und betreute Standard ISAD(G) definiert ein Set von beschreibenden Metadaten für die archivische Verzeichnung von Unterlagen. ISAD(G) beruht auf dem Rekursionsprinzip und der stufengerechten Verzeichnung. Der Standard liegt allen massgeblichen aktuellen Archivinformationssystemen (AIS) zugrunde und eignet sich daher ideal für die standardisierte Übernahme von Verzeichnungsdaten. Eine schweizerische Interpretation von ISAD(G) wurde durch den Verband Schweizerischer Archivarinnen und Archivare (VSA) publiziert, siehe Anhang B [Richtlinie ISAD(G)]

I.2.2.3 xlsadg

xlsadg ist eine konzeptionelle XML-Implementation von ISAD(G). Um die automatische Verzeichnung von Archivalien zu fördern, haben verschiedene Archive xlsadg im Projekt *bentō* der KOST entwickelt. Die Metadaten-Elemente von ISAD(G) wurden dabei speziell für die Verwendung bei der Archivierung digitaler Unterlagen um einige zusätzliche Elemente ergänzt. xlsadg ist inzwischen beim Staatsarchiv Basel-Stadt implementiert.

xlsadg ist einer der Endpunkte der hier vorgestellten Extraktionen.

I.2.2.4 Encoded Archival Description EAD

EAD (*Encoded Archival Description*) ist ein XML-Standard, der zur elektronischen Beschreibung von Findmitteln von der *Library of Congress* entwickelt wurde. Der Anwendungsbereich ist weit gefasst und umfasst Findmittel und Findbücher in Bibliotheken, Museen, aber auch Archiven. Die aktuelle Version ist EAD 2002, EAD 3 ist als Beta-Version verfügbar.

ISAD(G) und EAD wurden unabhängig voneinander in verschiedenen Kontexten und für verschiedene Zwecke entwickelt. Der Crosswalk zwischen den zwei Standards wird im Appendix A der EAD Tag-Library definiert, siehe: http://www.loc.gov/ead/tglib/appendix_a.html.

EAD ist einer der Endpunkte der hier vorgestellten Extraktionen.

I.2.2.5 Matterhorn METS Profile

Der *Metadata Encoding and Transfer Standard* METS ist ein sehr offen definiertes Containerformat für die digitale Archivierung. Die Verwendung von METS wird mittels sogenannter Profile für konkrete Anwendungen genau spezifiziert. Das vom Staatsarchiv Wallis und von Docuteam bei der *Library of Congress* registrierte Matterhorn METS Profile bildet das gesamte OAIS-Informationsmodell mittels international etablierter Standards ab. Es zeichnet sich insbesondere durch die Einbettung von EAD für die deskriptiven Informationen aus.

Die hier vorgestellte Metadaten transformation von eCH-0160 zu EAD ist in Zusammenarbeit mit dem Staatsarchiv Wallis und Docuteam entstanden. Die Transformation bildet eine der Grundlagen für die Überführung einer eCH-0160-Ablieferung in ein archivtaugliches Matterhorn-METS-Paket.

I.3 Realisierung

I.3.1 Transformationsrichtung

eCH-0160 definiert den Ausgangspunkt der hier dargestellten Metadaten transformation. Als erster Zwischenschritt wird die hierarchische Ordnungsstruktur von eCH-0160 auf eine ISAD(G)-konforme Verzeichnungsstruktur abgebildet. Als zweiter Zwischenschritt werden die eCH-0160-Metadaten jeder Ordnungsstrukturebene mit den entsprechenden ISAD(G)-Metadaten in Beziehung gesetzt. In der abschliessenden Metadaten transformation werden die ISAD(G)-Metadaten mit der konkreten xlsadg-Metadaten spezifikation bzw. dem EAD-Metadatenstandard für Erschliessungsinformationen gemappt.

I.3.2 Stratifizierung nach Verzeichnungseinheiten

Wir unterscheiden in eCH-0160 vier Entitäten [Spalten in der Exceltabelle] Paket/Ablieferung (1), Ordnungssystemposition (2), Dossier (3) und Dokument (4). Die Entität *Ordnungssystemposition* und *Dossier* kann rekursiv wiederholt werden. Diese vier Entitäten werden in ISAD(G) Bestand¹, Serie, Dossier und Dokument genannt. Es gibt in eCH-0160 Metadatenattribute, die in allen oder mehreren Entitäten gleichermassen vorkommen (grün); Attribute, die nur in einer Entität anzutreffen sind (gelb); und unterschiedliche Attribute mit ähnlichem Inhalt in mehreren Entitäten (rosa).

¹ Ein Bestand setzt sich in der Regel aus mehreren Ablieferungen der gleichen Behörde (Provenienz) zusammen; wir müssen hier also genaugenommen von einem *Ablieferungsbestand* sprechen.

Die einzelnen gleichwertigen Attribute aus verschiedenen Entitäten sind in Bahnen [Zeilen in der Exceltabelle] zusammengefasst und können so mit den entsprechenden Attributen einer ISAD(G)-Verzeichnungseinheit in Beziehung gesetzt werden.

Der Aufbau der Exceltabelle basiert auf dem eCH-0160 Data Dictionary, dem Data Dictionary zur *bentō*-Rahmenspezifikation und der deutschen Übersetzung von ISAD(G).

1

2

3

4

ISAD(G)

Ablieferung	Provenienz	Ordnungssystem	Ordnungssystemposition	Dossier	Dokument	ISAD(G)
ISAD(G): Bestand			ISAD(G): Serie	ISAD(G): Dossier	ISAD(G): Dokument	3.1.4 Verzeichnungsstufe
	aktenbildnerName					3.2.1 Name der Provenienzstelle
ablieferndeStelle						3.2.4 Abgebende Stelle
	geschichteAktenbildner					3.2.2 Verwaltungsgeschichte / Biographische Angaben
	extern: Archivkuerzel und Bestandessignatur + automatisch	generierte Nummer (fortlaufend SIG)	SIG.2 oder hierarchisch SIG.1 SIG.1			3.1.1 Signatur
ablieferungstyp	systemBeschreibung		zusatzmerkmal			3.2.3 Bestandesgeschichte
angebotsnummer	verwandteSysteme					
	extern: Bestand-Titel	titel	titel	titel		3.1.2 Titel
ablieferungsteile		id	inhalt	id		3.3.1 Form und Inhalt
			erscheinungsform	erscheinungsform		3.4.4 Physische Beschaffenheit und technische Anforderungen
			forminhalt	dokumenttyp		3.3.1 Form und Inhalt
			umfang	anwendung		3.1.5 Umfang (Menge und Abmessung)
		federfuehrendeOrganisationseinheit	federfuehrendeOrganisationseinheit	autor		
		eroeffnungsdatum	abschlussdatum	registrierdatum		
entstehungszeitraum	existenzzeitraum	anwendungszeitraum	entstehungszeitraum	entstehungszeitraum		3.1.3 Entstehungszeitraum / Laufzeit
		mitbenutzung	entstehungszeitraumAnmerkung			
referenzSchutzfristenFormular						
schutzfristenkategorie		schutzfristenkategorie	schutzfrist			
schutzfrist		schutzfrist	schutzfristenBegrueundung			
		schutzfristenBegrueundung	klassifizierungskategorie			
		klassifizierungskategorie	datenschutz			
		datenschutz	oeffentlichkeitsstatus			
		oeffentlichkeitsstatus	oeffentlichkeitsstatusBegrueundung			
		oeffentlichkeitsstatusBegrueundung	sonstigeBestimmungen			
		sonstigeBestimmungen	vorgang			
bemerkung	bemerkung	bemerkung	bemerkung	bemerkung		3.6.1 Allgemeine Anmerkungen
zusatzDaten		zusatzDaten	dateRef	dateRef		
	archivierungsmodusLoeschvorschriften	zusatzDaten	zusatzDaten	zusatzDaten		
referenzBewertungsentscheid						3.3.2 Bewertung und Kassation
alleinstehende Attribute						
auf mehreren Ebenen vertretenes Attribut						
auf mehreren Ebenen vertretene unterschiedliche Attribute mit ähnlichem Inhalt						

Abb. I.2: Exceltabelle in eCH-0160_xlsadg&EAD.xlsx

I.3.3 Metadatenabbildung

Wenn wir die Abbildung einzelner Attribute von eCH-0160 auf ISAD(G) betrachten, können wir verschiedene Fälle unterscheiden:

I.3.3.1 Obligatorische ISAD(G)-Attribute

Die obligatorischen ISAD(G)-Attribute sind nicht alle vollumfänglich in eCH-0160 vorhanden². Auf allen Stufen sind das *Signatur*, *Titel* und *Verzeichnungsstufe*. Nur auf bestimmten Stufen obligatorisch sind 3.2.1 *Name der Provenienzstelle* (obligatorisch auf der Stufe Bestand),

² Das rührt daher, dass eCH-0160 auf dem GEVER-Metadatenkonzept [I017] basiert.

3.1.3 *Entstehungszeitraum / Laufzeit* (obligatorisch auf den Stufen Bestand und Dossier) und 3.1.5 *Umfang* (obligatorisch auf der Stufe Bestand). Diese sind im Abschnitt 0

I.3.3.4 Vererbte oder aggregierte Attribute behandelt.0

3.1.1 *Signatur* muss aus externem Archivkürzel, externer Bestandessignatur und automatisch generierter Nummer (fortlaufend SIG.1 SIG.2 oder hierarchisch SIG.1 SIG.1.1) erzeugt werden.

3.1.2 *Titel* ist auf Bestandesebene nicht gegeben und muss manuell ergänzt werden.

3.1.4 *Verzeichnungsstufe* wird nach der oben bereits erläuterten Regel generiert (Paket/Ablieferung -> Bestand, Ordnungssystemposition -> Serie, Dossier -> Dossier und Dokument -> Dokument).

I.3.3.2 1:1-Abbildung

Kein eCH-0160-Attribut kann aus allen Entitäten 1:1 auf ISAD(G) abgebildet werden. In der Regel fehlt das entsprechende Attribut in einer Entität. Am ehesten können wir folgende eCH-0160-Attribute nennen: *titel*, *entstehungszeitraum*, *bemerkung* und die Attribute für die Zugangsbestimmungen. Siehe dazu aber im Detail 0

I.3.3.4 Vererbte oder aggregierte Attribute.

I.3.3.3 Kein entsprechendes ISAD(G)-Attribut

Eine ganze Reihe Attribute aus eCH-0160 finden keine Entsprechung in ISAD(G), da sie nicht Teil der Erschliessungsinformationen sind, sondern entweder

- *Technische Referenzen* (*id* und *dateiRef*);
- Teil der *Ablieferungsinformation* (*verwandteSysteme*, *angebotsnummer*, *mitbenutzung*, *archivierungsmodus* *Löschvorschriften*);
- Spezifische *GEVER-Informationen*: (*federfuehrendeOrganisationseinheit*, *autor*, *eroffnungsdatum*, *registrierdatum*, *abschlussdatum*, *vorgang*);
- *Schutzfristenvorgaben* (*referenzSchutzfristenFormular*, *schutzfristenkategorie*, *schutzfristen*, *schutzfristenbegründung*);
- Unstrukturierte *Zusatzdaten* (*zusatzDaten*).

I.3.3.4 Vererbte oder aggregierte Attribute

Weitere Attribute sind nicht auf jeder ISAD(G)-Verzeichnungsstufe vorhanden. Wir können dabei zwei Fälle unterscheiden: Gewisse Attribute vererben sich implizit oder explizit auf die tieferen Verzeichnungsstufen (*aktenbildnerName*, *ablieferndeStelle*, *geschichteAktenbildner*); andere Attribute werden aggregiert³ (*erscheinungsform*, *formInhalt*, *dokumenttyp*, *umfang*,

³ Aggregiert (lateinisch *aggregatio*: Anhäufung, Vereinigung) bedeutet, dass das Element auf mindestens einer Hierarchiestufe vorhanden sein muss und die Werte auf die jeweils darüber liegenden Stu-

entstehungszeitraum, klassifizierungskategorie, datenschutz, oeffentlichkeitsstatus, oeffentlichkeitsstatusBegrueundung, sonstigeBestimmungen).

Beim Vererben und Aggregieren werden nicht nur in eCH-0160 fehlende Attribute ergänzt, sondern auch leere Attribute implizit oder explizit gefüllt. Eine Tabelle aus dem Data Dictionary zur KOST *bentō* Rahmenspezifikation erläutert diesen Sachverhalt für ISAD(G):

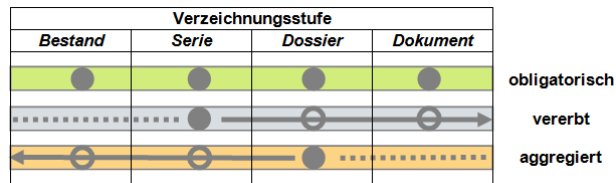


Abb. I.3: Vererbung und Aggregation (aus [xlsadg DataDictionary], S. 12)

Das Vorkommen der Elemente: obligatorisch, vererbt oder aggregiert, graphisch dargestellt:

	Verzeichnungselement ISAD(G)	Bestand	Serie	Dossier	Dokument
3.1	Identifikation				
3.1.1	Signatur				
3.1.2	Titel				
3.1.3	Entstehungszeitraum / Laufzeit				
3.1.4	Verzeichnungsstufe				
3.1.5	Umfang (Menge und Abmessung)				
3.2	Kontext				
3.2.1	Name der Provenienzstelle				
3.2.2	Verwaltungsgeschichte / Biographische Angaben				
3.2.3	Bestandesgeschichte				
3.2.4	Abgebende Stelle				
3.3	Inhalt und innere Ordnung				
3.3.1	Form und Inhalt				
3.3.2	Bewertung und Kassation				
3.3.3	Neuzugänge				
3.3.4	Ordnung und Klassifikation				
3.4	Zugangs- und Benutzungsbedingungen				
3.4.1	Zugangsbestimmungen				
3.4.2	Reproduktionsbestimmungen				
3.4.3	Sprache / Schrift				
3.4.4	Physische Beschaffenheit und technische Anforderungen				
3.4.5	Findhilfsmittel				
3.5	Sachverwandte Unterlagen				
3.5.1	Aufbewahrungsort der Originale				
3.5.2	Kopien bzw. Reproduktionen				
3.5.3	Verwandte Verzeichnungseinheiten				
3.5.4	Veröffentlichungen				
3.6	Anmerkungen				
3.6.1	Allgemeine Anmerkungen				
3.7	Verzeichnungskontrolle				
3.7.1	Information des Bearbeiters				
3.7.2	Verzeichnungsgrundsätze				
3.7.3	Datum oder Zeitraum der Verzeichnung				

Legende: obligatorisch vererbt aggregiert

Abb. I.4: obligatorische, vererbte und aggregierte Attribute in ISAD(G) (aus [xlsadg DataDictionary], S. 11)

fen logisch zusammengeführt werden; auf den darunter liegenden Hierarchiestufen ist das Element fakultativ.

I.3.4 Abbildung auf xlsadg und EAD

Die Abbildung von ISAD(G) auf xlsadg oder EAD folgt wie oben bereits erläutert den konkreten Implementierungen in den Staatsarchiven Basel-Stadt und Wallis.

Die Transformation eCH-0160 → EAD bildet sämtliche eCH-0160-Attribute ab. Es werden auch Nicht-ISAD(G)-Felder gemappt (Details in der entsprechenden Excel Datei *eCH-0160_xlsadg&EAD.xlsx*). Damit ist sichergestellt, dass sämtliche deskriptiven Informationen der eCH-0160-Ablieferung im Archiv genutzt werden können. Die Transformation eCH-0160 → xlsadg hingegen reflektiert nur das Exzerpieren von Erschliessungsinformationen (*Descriptive Information*).

Die Spalte xlsadg folgt einer relativen XPath Notation, während bei der EAD-Spalte eine Punkt-Notation (*current node*) verwendet wurde.

ISAD(G)	xlsadg	EAD (XPath relative to C or ArchDesc element)
3.1.4 Verzeichnungsstufe	//identity/descriptionLevel	./[@level = 'otherlevel']/@otherlevel
3.2.1 Name der Provenienzstelle	//context/creator	./EAD:did/EAD:origination
3.2.4 Abgebende Stelle	//context/acqInfo	./EAD:acqinfo/EAD:p
3.2.2 Verwaltungsgeschichte / Biographische Angaben	//context/adminBioHistory	./EAD:bioghist/EAD:p
3.1.1 Signatur	//identity/referenceCode	./EAD:did/EAD:unitid[@type = 'refCode']
3.2.3 Bestandesgeschichte	//context/archivalHistory	./EAD:custodhist/EAD:p
		./EAD:did/EAD:physdesc/EAD:physfacet
		./EAD:did/EAD:unitid[@type = 'refCodeAdmin']
3.1.2 Titel	//identity/title	./EAD:did/EAD:unitid[@label = 'main']
3.3.1 Form und Inhalt	//scopeContent/content	./EAD:did/EAD:abstract
		./EAD:did/EAD:unitid[@type = 'refCodeAdmin']
3.4.4 Physische Beschaffenheit und technische Anforderungen	//physTech	./EAD:physTech/EAD:p
3.3.1 Form und Inhalt	//scopeContent/scope	./EAD:scopecontent/EAD:p
		./EAD:controlaccess/EAD:genreform
3.1.5 Umfang (Menge und Abmessung)	//extentMedium/medium	./EAD:did/EAD:physdesc/EAD:physfacet
		./EAD:did/EAD:origination
		./EAD:did/EAD:unitdate[@label = 'fromYear']
		./EAD:did/EAD:unitdate[@label = 'toYear']
3.1.3 Entstehungszeitraum / Laufzeit	//identity/dates/fromDate //identity/dates/toDate	./EAD:did/EAD:unitdate[@label = 'creationPeriod']
		./EAD:did/EAD:unitdate[@label = 'creationPeriodNotes']
		./EAD:controlaccess/EAD:persname[@role = 'involved']
3.4.1 Zugangsbestimmungen	//accessConditions/hasPrivacyProtection //accessConditions/openToThePublic //accessConditions/classification //accessConditions/otherAccessConditions //accessConditions/accessConditionsNotes	./EAD:accessrestrict[@type = 'restrictions']/EAD:p
3.6.1 Allgemeine Anmerkungen	//notes/note	./EAD:note/EAD:p
		./EAD:dao[@xlink:role = 'simple']/@xlink:href
		./EAD:odd[@type = 'usage']/EAD:p
3.3.2 Bewertung und Kassation	//contentStructure/appraisalDestruction	./EAD:appraisal/EAD:p

Abb.I.5: Auszug aus den Tabellen in eCH-0160_xlsadg&EAD.xlsx

I.3.5 XSL-Transformation

Damit auch eine Möglichkeit besteht, die Transformationen konkret zu überprüfen, sind beide Transformationen neben der Darstellung in Excel auch als XSL-Transformationen ausgeführt. Damit können mit einem entsprechenden Programm (saxon, xalan u.a.) XML-Metadatendateien nach eCH-0160 in die jeweils anderen Metadaten-Formate überführt werden.

Der Übersichtlichkeit halber sind die Transformationen nach den vier Ebenen in eCH-0160 gegliedert *..fond.xsl* *..serie.xsl* *..file.xsl*⁴ und *..item.xsl*. Zugangsbestimmungen, Entstehungszeitraum und Signatur sind gesondert in *..access.xsl* *..date.xsl* und *..reference.xsl* abgehandelt.

Für den Online-Test steht eine Webapplikation zur Verfügung:

<http://kost-ceco.ch/isadg/app/input.php>

KOST Referenzimplementierung - Descriptive Information

Diese *Descriptive Information* Referenzimplementierung konvertiert eine eCH-0160 Metadaten Datei in eine xIsadg oder EAD Datei. Die importierte Metadaten Datei wird mit einer Schema Validierung überprüft, verschiedene Metadatenschema werden erkannt, konvertiert werden können aber nur eCH-0160 Metadaten. Ziel ist das Extrahieren der ISAD(G) relevanten Metadaten, die *Descriptive Information* nach OAIS für den Import in ein *Datamanagement* oder ein Findmittel und nicht aber die vollständige Abbildung aller eCH-0160 Metadaten in einem anderen Schema. Grundlage der Konvertierung ist der folgender Crosswalk [eCH-0160 zu xIsadg_v3.xlsx](#) und [CH-0160 zu EAD_v1.xlsx](#). Technisch basiert die Konvertierung auf folgenden XSLT Transformation [eCH2xIsadg.zip](#) und [eCH2EAD.zip](#). Eine Windows Commandline Version benötigt Saxon9 und xmllint und ist hier zu finden [eCH2ISADG_cmd.zip](#). Die Basisdokumente für die beiden Metadatenschema sind die folgenden *Data Dictionaries*: [eCH-0160_V1.0_ArchivischeAblieferungsschnittstelle_DataDictionary.pdf](#) und [xIsadg_DataDictionary_v1.6.pdf](#).

Eine eCH-0160 Metadaten Datei für die Konvertierung auswählen:

Folgende Dateien sind bereits übertragen:

- metadata.xml (29645 bytes) eCH-0160 / arelda_v4 SIP Metadata

Folgende Datei steht zur Konvertierung bereit: **metadata.xml**

Manuelle Angaben zur Zielakzession (Archivtektonik):

Bestand-Signatur: Bestandstitel:

Angaben zur Signaturbildung:

Signaturstil: ☒ fortlaufende Nummerierung ☐ dezimal Gliederung

Ausgabeformat: ☒ xIsadg ☐ EAD

Copyright © 2005-2013 by Koordinationsstelle für die dauerhafte Archivierung elektronischer Unterlagen

info@kost-ceco.ch

Abb.I.6: Screenshot der Webapplikation

⁴ NB: Die englischen Termini *file* und *item* entsprechen den deutschen „Dossier“ und „Dokument“.