eCH-0160 – Archivische Ablieferungsschnittstelle

Name	Archivische Ablieferungsschnittstelle
eCH-Nummer	eCH-0160
Kategorie	Standard
Reifegrad	Implementiert
Version	1.2.0
Status	Abgelöst
Beschluss am	2022-03-01
Ausgabedatum	2022-05-11
Ersetzt Version	1.1 – Minor Change
Voraussetzungen	-
Beilagen	eCH-0160-1.2.0.xsd.zip BEIL1_d_2022-01-14_eCH-0160_V1.2.0_ArchivischeAbliefe-rungsschnittstelle_DataDictionary.xlsx
Sprachen	Deutsch (Original), Französisch (Übersetzung)
Autoren	Fachgruppe Digitale Archivierung
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 Ergänzungen zur besseren Unterteilung und Beschreiung von Archivgut implementiert, Bereinigungen vorgenommen sowie das Data Dictionary mit dem xsd abgeglichen. Bei der Bereinigung wurde insbesondere die Teile im xsd entfernt, welche nicht zur Ablieferungsschnittstelle gehören und bei den SIP nie verwendet werden.

Die Versionen 1.0 oder 1.1. erhalten den Status abgelöst, können jedoch weiterhin verwendet werden. Die Schema behalten ihre Gültigkeit.

Parallel zur Spezifikation werden auf der Homepage der KOST unter https://kost-ceco.ch/cms/ech-0160-archivische-ablieferungsschnittstelle.html diverse eCH-0160-Testpakete publiziert und gepflegt.

Inhaltsverzeichnis

1 Einleitung 6
1.1 Status 6
1.2 Ziel und Zweck 6
1.3 Struktur des Dokuments 6
1.3.1 Aufbau Kapitel 6
1.3.2 ID Anforderungen 7
1.3.3 Unterscheidung zwischen Muss und Kann-Anforderungen 8
1.3.4 Notation Ordner, Dateien und Ordnerstrukturen 8
1.4 Präzisierungsbedarf 8
1.5 Übersicht 9
1.5.1 Abgrenzung 9
1.5.1.1 Definitionen 9
1.5.1.2 Anforderungen pro Ablieferungstyp 10
1.5.2 Anwendungsgebiet 10
1.5.3 Ablieferungsobjekt Paket – SIP 11
1.5.4 Inhalt 12
1.6 Verhältnis Ablieferung – Paket 12
1.7 Sicherheitseinstellungen Dateien im Paket 12
2 Primärdaten 13
2.1 Grundsätzliches 13
2.2 Formate 13
3 Metadaten 14
3.1 Allgemeine Anforderungen 14
3.2 Konzeptuelles Datenmodell 15
3.3 Datenmodell Ablieferungstyp GEVER 17
3.4 Datenmodell Ablieferungstyp FILES 19
3.5 Data Dictionary 21
3.6 XSD 22

3.7 Aufbau Inhaltsverzeichnis im `metadata.xml` ... 23
3.8 Mapping Metadaten des Ursprungssystems auf SIP-Metadaten ... 24
3.9 Metadaten zu den Schutzfristen ... 26
3.10 Metadaten zu den Zeiträumen der Dossiers ... 27
3.11 Metadaten zu den Prüfsummen der Dateien ... 27
3.12 Referenzierung der Zugehörigkeit zu Dossiers / Mappen in den Metadaten ... 28
4 Struktur SIP ... 29
4.1 Rahmenvorgaben und Begrenzungen Grösse Paket ... 29
4.2 Rahmenvorgaben und Begrenzungen Anzahl Dateien im Paket ... 30
4.3 Erlaubte Zeichen bei der Benennung der Ordner und Dateien ... 31
4.4 Aufbau eines SIP ... 32
4.5 Pfadlängen und Hierarchien im SIP ... 36
4.6 Spezifische Vorgaben zu einem SIP für GEVER ... 36
4.7 Spezifische Vorgaben zu einem SIP für FILES ... 40
4.8 Aufbau eines FILES SIP mit integrierter Dokumentation ... 41
5 Transfer ... 43
5.1 Format Package Container ... 43
6 Version und Gültigkeit der Spezifikation ... 43
7 Change-Request-Prozess ... 43
8 Sicherheitsüberlegungen ... 44
9 Haftungsausschluss/Hinweise auf Rechte Dritter ... 45
10 Urheberrechte ... 45
Anhang A – Referenzen & Bibliographie ... 46
Anhang B – Mitarbeit & Überprüfung ... 47
Anhang C – Abkürzungen und Glossar ... 47
Anhang D – Änderungen gegenüber Vorversion ... 50
Anhang E – Abbildungsverzeichnis ... 51
Anhang F – Tabellenverzeichnis ... 52
Anhang H – Ausschnitt Mapping SIP-Metadaten – GEVER-System ... 53
Anhang I – Zeichensätze ... 54

Anhang J – Beziehung zu weiteren archivischen Standards..... 63

Hinweis

Aus Gründen der besseren Lesbarkeit und Verständlichkeit wird im vorliegenden Dokument bei der Bezeichnung von Personen ausschliesslich die maskuline Form verwendet. Diese Formulierung schliesst Frauen in ihrer jeweiligen Funktion ausdrücklich mit ein.

1 Einleitung

1.1 Status

Abgelöst: Das Dokument wurde durch eine neue, aktuellere Version ersetzt. Die Benutzung ist zwar noch möglich, es wird aber empfohlen, die neuere Version einzusetzen.

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

1.2 Struktur des Dokuments

1.2.1 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	Anforderungstext Beispiel Beispieltext *Empfehlung* *Empfehlungstext ist immer kursiv.*	M

1.2.2 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.		

1.2.3 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 *Best Practice*. Das bedeutet, dass es dem der archivierenden Stelle frei steht, die Kann-Anforderungen von den abliefernden Stellen (Paketerstellung) einzufordern oder nicht.

1.2.4 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.

1.3 Präzisierungsbedarf

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

- zugelassene archivtaugliche Formate für die Primärdaten (siehe 2.2)
- Maximalgrösse eines SIP (siehe 4.1)
- Maximale Anzahl Dateien in einem Ordner innerhalb des SIP (siehe 4.2)
- Name des Toplevel-Ordners / SIP-Identifikators (siehe 5.4).
- Übernahmevorgaben für das Ordnungssystem (siehe 4.8-2)

1.4 Übersicht

1.4.1 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. Abbildung 1).

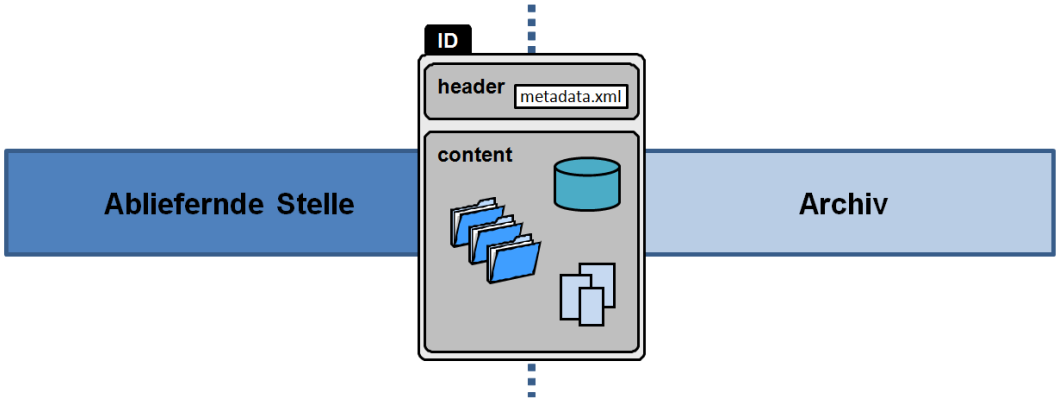


Abbildung 1: Abgrenzung Ablieferungsobjekt SIP

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).

1.4.1.1 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.
	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.

1.4.1.2 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.

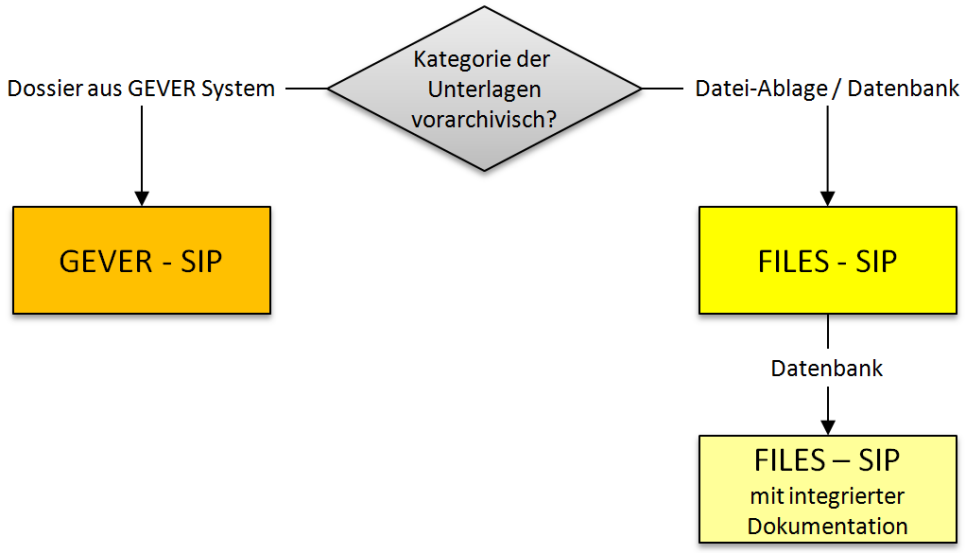


Abbildung 2: Ablieferungskategorien SIP

1.4.2 Anwendungsgebiet

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 erstellen SIP geliefert werden. Ist dies nicht der Fall werden die Unterlagen an die ablieferende Stelle zurückgesendet, damit diese die Korrekturen durchführen kann, resp. ein neues SIP gemäss den Anforderungen erstellt.

1.4.3 Ablieferungsobjekt Paket – SIP

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

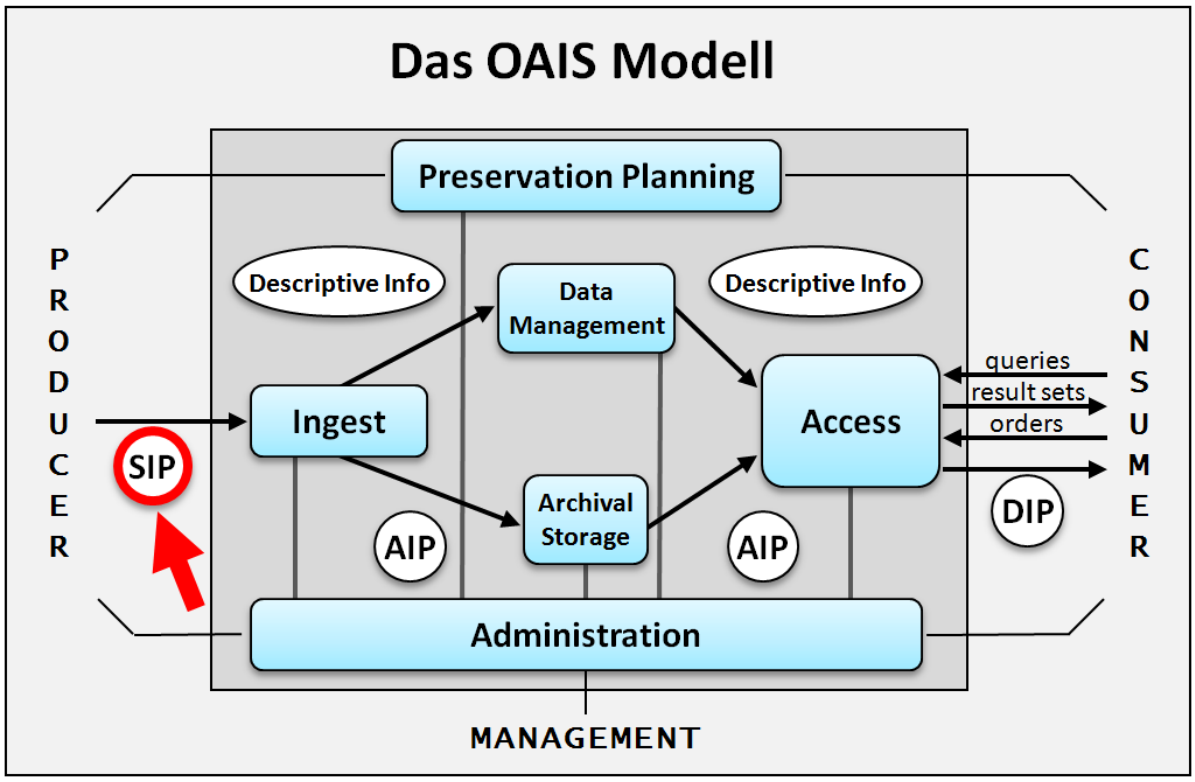


Abbildung 3: OAIS Modell mit Kennzeichnung SIP

Es gibt im OAIS Modell 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

Tabelle 1: Pakettypen

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.

1.4.4 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. Abbildung 4)

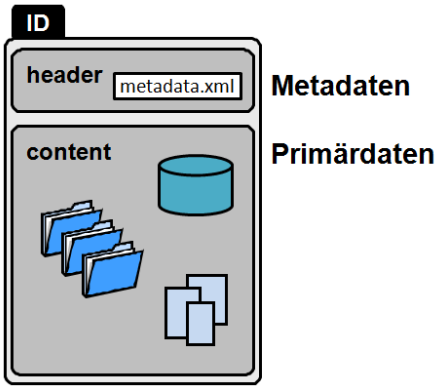


Abbildung 4: Aufbau SIP – Metadaten und Primärdaten

1.5 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

1.6 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

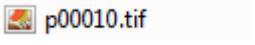
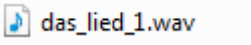
2 Primärdaten

2.1 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.

2.2 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	Die Primärdaten müssen in einem für die Unterlagen geeigneten archivtauglichen Format im SIP integriert werden. *Empfehlung* *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.*	M
P_3.2-2	Die Dateiendungen der Dateien müssen den üblichen Konventionen in Bezug auf die Formate der jeweiligen Dateien entsprechen. Beispiel Eine Bilddatei im TIFF-Format:  Eine Tondatei im WAVE-Format:  	M

3 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.

3.1 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 `metadata.xml` (getrennt von den Primärdaten) vorliegen. Die Spezifikation gibt sowohl die Struktur der Metadatendatei wie auch Inhalt der Metadaten vor (vgl. dazu die Kapitel zur XSD und zum Data Dictionary). Diese Vorgaben müssen exakt eingehalten werden.	M
M_4.1-2	Die archivische Ablieferungsschnittstelle hat die Version 1.2.0. Das Schema hat die Version 5.0. Damit die älteren SIP gegen die neuen XSD validiert werden können, muss der ursprüngliche Namespace behalten werden: `http://bar.admin.ch/arelda/v4`	M

M_4.1-3	Da mehrere Tools mit Präfixe im XML-Namespace nicht korrekt umgehen können, wird dringend empfohlen in `metadata.xml` auf Präfixe im XML-Namespace zu verzichten.
	
	Empfehlung
	Präfixe im XML-Namespace sollen nicht verwendet werden. Folgender Aufbau des `metadata.xml` *wird dringend empfohlen:*
	`<?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:schemaLocation="http://bar.admin.ch/arelda/v4 xsd/arelda.xsd"`
	`xsi:type="paketSIP" schemaVersion="5.0">`
	`<paketTyp>SIP</paketTyp>`
	`...`
	`</paket>`
		K

3.2 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 *Paket* 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 *Ablieferung* beinhaltet alle Angaben zur abliefernden Stelle und zur Ablieferung ans Archiv.
Provenienz	Die Entität *Provenienz* repräsentiert den Aktenbildner und bei -Ablieferungen vom Ablieferungstyp FILES zusätzlich das Informationssystem, aus dem die Ablieferung stammt.
Ordnungssystem	Die Entität *Ordnungssystem* beschreibt den amtsseitig verwendeten Registraturplan, Aktenplan oder Strukturplan resp. die Ordnung des Archivgutes.
Ordnungssystemposition	Die Entität *Ordnungssystemposition* beschreibt die Eigenschaften der einzelnen Ordnungssystempositionen im Ordnungssystem.
Dossier	Die Entität *Dossier* 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.

Entität	Beschreibung
Vorgang	Die Entität *Vorgang* 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 *Vorgang* auch weiterhin als Attribut der Entität *Dossier* verwendet werden.
Aktivität	Die Entität *Aktivität* beschreibt die Tätigkeit, die ausgeführt werden soll.
Mappe	Die Entität *Mappe* beschreibt die Eigenschaften des zusammengehörenden Archivguts aus einem Geschäft oder aus einer anderen Ordnungsstruktur. Sie dient der einfachen Gliederung.
Dokument	Die Entität *Dokument* 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 *Inhaltsverzeichnis* repräsentiert das im `metadata.xml` enthaltene Inhaltsverzeichnis, das alle Ordner und Dateien umfasst.
Ordner	Die Entität *Ordner* repräsentiert die Ordner im Paket, in denen das Paket strukturiert ist.
Datei	Die Entität *Datei* repräsentiert die Primärdateien im Dateisystem und beinhaltet auch Angaben zur Identifikation und Beschreibung der einzelnen Dateien.
Archivischer Vorgang	Die Entität *archivischer Vorgang* wird erst im Archiv mit Metadaten gefüllt und dokumentiert Tätigkeiten wie die Eingangskontrolle in einem SIP. Für die Erstellung eines SIP ist diese Entität nicht von Bedeutung. Sie darf bei der Erstellung nicht verwendet werden.
Archivische Notiz	Die Entität *archivische Notiz* 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, der Mappe 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 nur der Vollständigkeit halber für die Dokumentation aller Entitäten im XSD aufgeführt.

3.3 Datenmodell Ablieferungstyp GEVER

ID	Beschreibung Anforderung	M/K
M_4.3-1	Der Ablieferungstyp GEVER gilt für alle Ablieferungen, die digitale Unterlagen aus einem Geschäftsverwaltungssystem (GEVER) enthalten. Das Datenmodell für GEVER-Ablieferungen ist in der Abbildung 5 ersichtlich. Die folgenden Entitäten kommen im GEVER SIP (resp. in den Metadaten) vor: • Provenienz • Paket • Ablieferung • Ordnungssystem • Ordnungssystemposition • Inhaltsverzeichnis • Dossier • Vorgang • Aktivität • Mappe • Dokument • Datei • Ordner Die Entität *Archivischer Vorgang* 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. Die Entität *Archivische Notiz* 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. Die Entität *Mappe* dient dazu Ordnungssystempositionen und Dossiers weiter zu strukturieren. Mappe ist jedoch nicht rekursiv. Die Entität *Unstrukturierter Anhang* ist veraltet und soll in einem GEVER SIP nicht vorkommen.	M

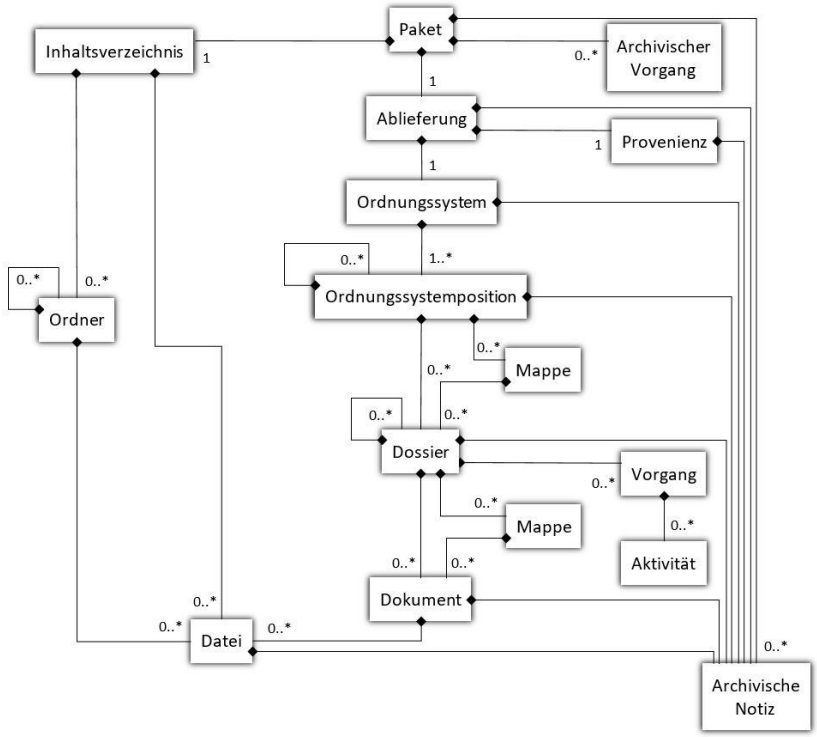


Abbildung 5: Konzeptuelles Datenmodell– Anwendungsbereich Ablieferungstyp GEVER

ID	Beschreibung Anforderung	M/K
M_4.3-2	Für die Beziehungen zwischen den Entitäten Dossier und Mappe zur Entität Dokument gibt es im GEVER SIP zwei Fälle (vgl. Abbildung 6). Fall A (blau markiert): Ein Dokument ist direkt einem Dossier zugeordnet. Ein Dossier besitzt 0..* Dokumente. Fall B (violett markiert): Ein Dokument ist direkt einer Dossier-Mappe zugeordnet. Eine Dossier-Mappe besitzt 0..* Dokumente. Fall A und B können gleichzeitig in einem Paket und in einem Dossier vorkommen.	M

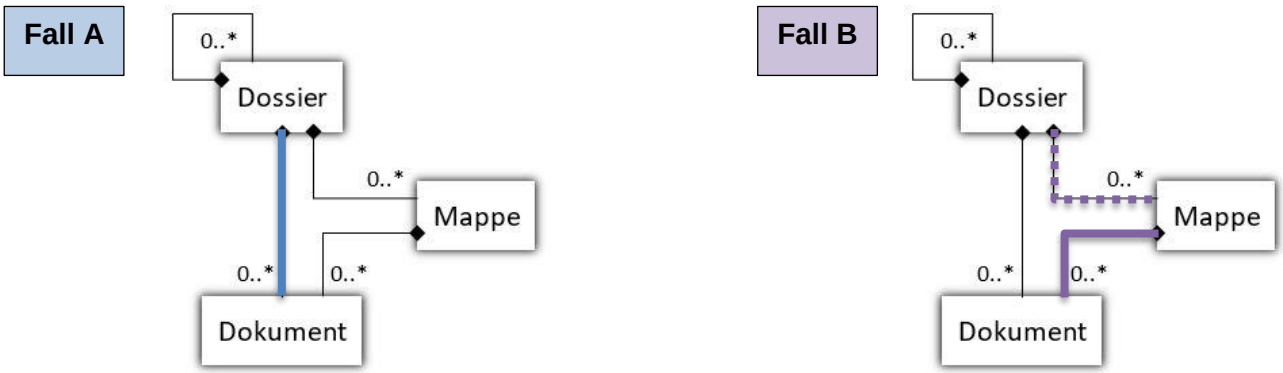


Abbildung 6: Möglichkeiten der Beziehungen zwischen Dossier und Mappe zu Dokument (vereinfachte GEVER-Darstellung)

3.4 Datenmodell Ablieferungstyp FILES

ID	Beschreibung Anforderung	M/K
M_4.4-1	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 Abbildung 7 ersichtlich. Die folgenden kommen im FILES SIP (resp. in den Metadaten) vor: • Provenienz • Paket • Ablieferung • Ordnungssystem • Ordnungssystemposition • Inhaltsverzeichnis • Dossier • Mappe • Dokument • Datei • Ordner Die Entitäten *Dokument* muss in einem FILES SIP nicht zwingend vorkommen. Sie kann bei Bedarf verwendet werden. Die Entität *Archivischer Vorgang* 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. Die Entität *Archivische Notiz* 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. Die Entität *Mappe* dient dazu Ordnungssystem, Ordnungssystempositionen und Dossiers weiter zu strukturieren. Diese Mappen sind jedoch nicht rekursiv. Kann eine Ablieferung keinem Ordnungssystem zugeordnet werden, kann die Ablieferung auch nur aus Mappen und Submappen bestehen (z.B. Geo-SIP). Die Entität *Unstrukturierter Anhang* ist veraltet und soll in einem FILES SIP nicht vorkommen.	M

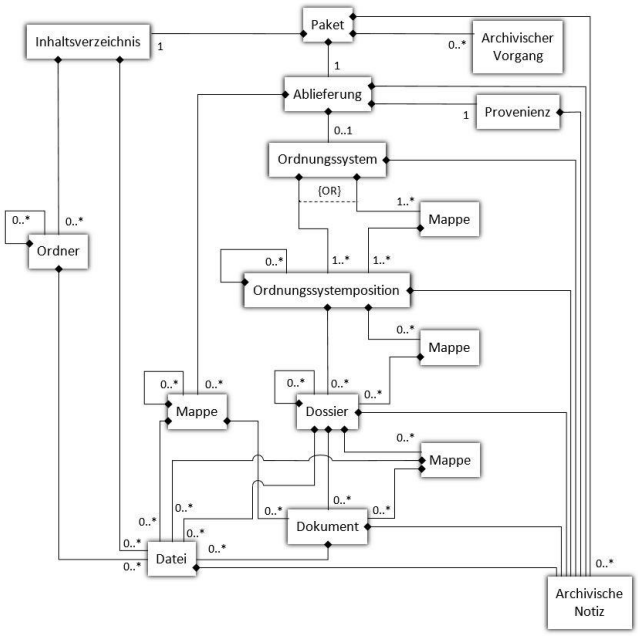


Abbildung 7: Konzeptuelles Datenmodell– Anwendungsbereich Ablieferungstyp FILES

ID	Beschreibung Anforderung	M/K
M_4.4-2	Für die Beziehungen zwischen den Entitäten Dossier, Mappe und Dokument zur Entität Datei gibt es im FILES SIP vier Fälle (vgl. Abbildung 8). Fall A (blau markiert): Eine Datei ist direkt einem Dossier zugeordnet. Ein Dossier besitzt 0..* Dateien. Es gibt keine Dokumente. Fall B (violett markiert): Eine Datei ist direkt einer Dossier-Mappe zugeordnet. Eine Dossier-Mappe besitzt 0..* Dateien. Es gibt keine Dokumente. Fall C (grün markiert): Eine Datei ist einem Dokument zugeordnet. Ein Dokument besteht aus 0..* Dateien. (wie bei der Ausprägung GEVER) Fall D (rot markiert): Eine Datei ist direkt einer Ablieferungs-Mappe zugeordnet. Eine Ablieferungs-Mappe besteht aus 0..* Dateien. Fall A, B und C können gleichzeitig in einem Dossier vorkommen. Fall B und C können gleichzeitig in einer Dossier-Mappe vorkommen. Fall D und C können gleichzeitig in einer Ablieferungs-Mappe vorkommen.	M

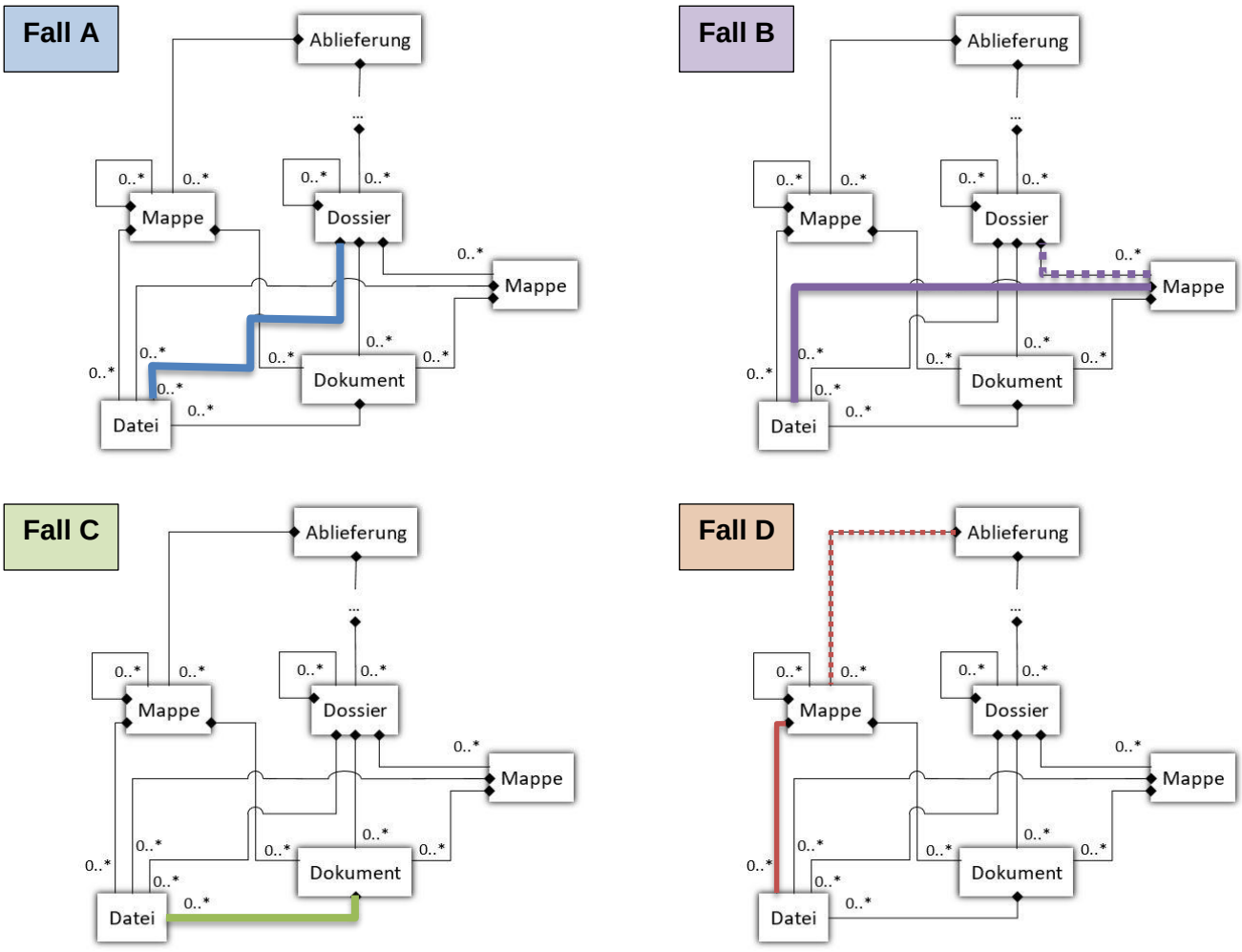


Abbildung 8: Möglichkeiten der Beziehungen zwischen Dossier, Mappe und Dokument zu Datei (vereinfachte FILES-Darstellung)

ID	Beschreibung Anforderung	M/K
M_4.4-3	Für die Beziehungen zwischen den Entitäten Ordnungssystem und Mappe zur Entität Ordnungssystemposition gibt es im FILES SIP zwei Fälle (vgl. Abbildung 9). Fall A (blau markiert): Eine Ordnungssystemposition ist direkt einem Ordnungssystem zugeordnet. (wie bei der Ausprägung GEVER) Fall B (violett markiert): Eine Ordnungssystemposition ist direkt einer Mappe zugeordnet. Ein Ordnungssystem besitzt 1..* Ordnungssystempositionen, ENTWEDER direkt wie im Fall A beschrieben ODER indirekt via Mappe (Fall B). Fall A und B können gleichzeitig in einem Ordnungssystem vorkommen.	M

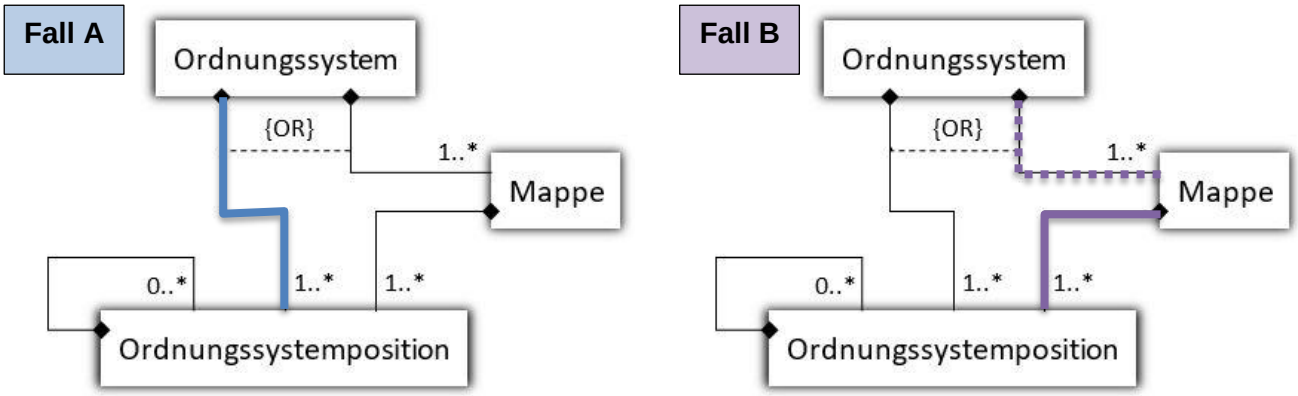


Abbildung 9: Möglichkeiten der Beziehungen zwischen Ordnungssystem und Mappe zu Ordnungssystemposition (vereinfachte FILES-Darstellung)

3.5 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	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. *Empfehlung* *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.*	M

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

3.6 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	Die XSD ist für das `metadata.xml` verbindlich einzuhalten. Das heisst, jedes metadata.xml in einem SIP muss gegenüber der XSD positiv validiert werden können. Als XSD ist die Schemadatei `arelda.xsd` aus der Beilage zur eCH-0160 Spezifikation Version 1.2.0 zu verwenden. Die XSD `arelda.xsd` von eCH-0160 Version 1.2.0 hat die interne Version 5.0 `schemaVersion="5.0"`	M
M_4.6-2	Die Schema Location soll auf das lokale Schema im SIP Ordner `/xsd`, verwiesen. Beispiel `xsi:schemaLocation="http://bar.admin.ch/arelda/v4 xsd/arelda.xsd"`	K

3.7 Aufbau Inhaltsverzeichnis im `metadata.xml`

ID	Beschreibung Anforderung	M/K

M_4.7-1	Alle Ordner und Dateien im Paket (Inhalte des Ordners `content/` und Ordners `header/`) müssen im Inhaltsverzeichnis des `metadata.xml` gemäss den Vorgaben mit allen Metadaten aufgeführt sein.
	Die einzige Ausnahme dieser Regel gilt für die Datei `metadata.xml` im Ordner `header/`. Diese Datei darf nicht im `metadata.xml` aufgeführt werden, da sie sonst auf sich selbst verweisen würde.
	
	Beispiel
	Aufbau des SIP (schematisch)
	`SIP_20090824_Bm/`
	`header/`
	`metadata.xml`
	`xsd/`
	`ablieferung.xsd`
	`….xsd`
	`content/`
	`d0001/`
	`d0002/`
	`p0001.pdf`
	
	Entsprechendes Inhaltsverzeichnis eines `metadata.xml`
		M

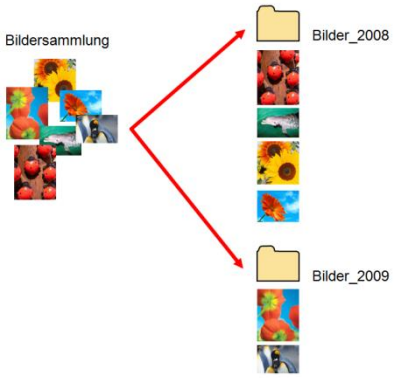
```
<?xml version="1.0" encoding="UTF-8"?>
<paket xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance" xmlns="
http://bar.admin.ch/arelda/v4" xsi:schemaLocation="http://bar.admin.ch/arelda/v4
xsd/arelda.xsd" xsi:type="paketSIP" schemaVersion="5.0">
    <paketTyp>SIP</paketTyp>
    <inhaltsverzeichnis>
        <ordner>
            <name>header</name>
            <originalName>header</originalName>
            <ordner>
                <name>xsd</name>
                <originalName>xsd</originalName>
                <datei id="_BAZ5gcEWEeGjN725uNwvBQ">
                    <name>ablieferung.xsd</name>
                    <originalName>ablieferung.xsd</originalName>
                    <pruefalgorithmus>MD5</pruefalgorithmus>
                    <pruefsumme>ca09882b4e76d8bd6f32309930b871a5</pruefsumme>
                </datei>
...
        <ordner>
            <name>content</name>
            <ordner>
                <name>DOS_01</name>
                <datei id="DAT0">
                    <name>Auftragsuebersicht.pdf</name>
                    <originalName>Auftragsuebersicht.pdf</originalName>
                    <pruefalgorithmus>MD5</pruefalgorithmus>
                    <pruefsumme>ea205477f15961fa22553b94c4152fd5</pruefsumme>
                </datei>
...
```

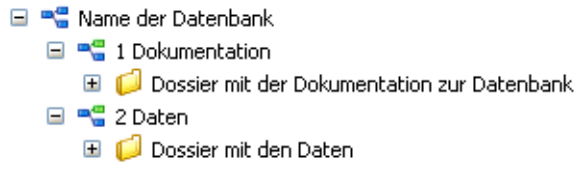
3.8 Mapping Metadaten des Ursprungssystems auf SIP-Metadaten

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* *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 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 und ggf. Mappe zusammengefügt. *Empfehlung* *Falls keine solche Ordnung bereits vorhanden ist, sollte sie aufgrund der Inhalte der Unterlagen erstellt werden.* *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.* *Für die Abbildung der Struktur können auch Mappen verwendet werden. Mappen haben im Gegensatz zu Dossier nur einen minimalen Metadatensatz und entspricht dem früheren «Reiter» in der Registratur.*	M

ID	Beschreibung Anforderung	M/K

	Beispiel
	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.
	 
	
	Der Order content im SIP sieht dann folgendermassen aus:
	`content/`
	`Bilder_2008/`
	`Kaefer.tif`
	`Delfin.tif`
	`Sonnenblume.tif`
	`Orange_Blume.tif`
	`Bilder_2009/`
	`Mehrere_Blumen.tif`
	`Pinguine.tif`
	Modellhaft könnte diese Ordnung mit der Zuordnung zu den Dossiers folgendermassen dargestellt werden:
	Bildersammlung
	1 Bilder von 2008 bis 2009
	Bilder 2008
	Bilder 2009
	= Ordnungssystem
	= Ordnungssystemposition
	= Dossier	

ID	Beschreibung Anforderung	M/K
M_4.8-4	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. Beispiel Modellhaft könnte eine solche Ordnung folgendermassen dargestellt werden:  = Ordnungssystem = Ordnungssystemposition = Dossier	M

3.9 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	Die Angaben zu den Schutzfristen der Unterlagen im Paket sollten in die dazu zur Verfügung stehenden Metadaten im `metadata.xml` eingetragen werden. Bei den Metadaten handelt es sich um die folgenden: `<schutzfristenkategorie>`: Angabe des relevanten Artikels in den gesetzlichen Grundlagen (falls anwendbar) `<schutzfrist>`: Dauer der Schutzfrist in Jahren Das Metadatenpaar, das die Schutzfristen festhält, steht in drei Entitäten zur Verfügung: Ablieferung Ordnungssystemposition Dossier *Empfehlung* *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.*	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

3.10 Metadaten zu den Zeiträumen der Dossiers

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

3.11 Metadaten zu den Prüfsummen der Dateien

ID	Beschreibung Anforderung	M/K

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

3.12 Referenzierung der Zugehörigkeit zu Dossiers / Mappen in den Metadaten

ID	Beschreibung Anforderung	M/K

M_4.12-1	Das Metadatum `<dateiRef>` dient dazu, den Zusammenhang zwischen den Dateien im Paket und deren Zugehörigkeit zu den Dokumenten in den Dossiers / Mappen oder direkt zu einem Dossier / Mappe herzustellen. Dieser Zusammenhang wird durch die entsprechende Referenzierung der `<dateiRef>` zu der entsprechenden Datei im Inhaltsverzeichnis des `metadata.xml` festgehalten.
	Diese Referenzierung ist die eindeutige Zuweisung der Inhalte der Dateien zu den Dossiers / Mappen.
	`<?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:schemaLocation="http://bar.admin.ch/arelda/v4 xsd/arelda.xsd" xsi:type="paketSIP" schemaVersion="5.0">`
	`<paketTyp>SIP</paketTyp>`
	`<inhaltsverzeichnis>`
	`<ordner>`
	`<name>header</name>`
	...
	`</ordner>`
	`<ordner>`
	`<name>content</name>`
	`<ordner>`
	`<name>DOS 01</name>`
	`<datei id="DAT0">`
	`<name>Auftragsuebersicht.pdf</name>`
	`<originalName>Auftragsuebersicht.pdf</originalName>`
	`<pruefalgorithmus>MD5</pruefalgorithmus>`
	`<pruefsumme>ea205477f15961fa22553b94c4152fd5</pruefsumme>`
	`</datei>`
	...
	...
	`</ordner>`
	`</inhaltsverzeichnis>`
	`<ablieferung xsi:type="ablieferungFilesSIP">`
	`<ablieferungstyp>FILES</ablieferungstyp>`
	...
	`<dossier id="IdD-1">`
	`<titel>Allgemeines zum Test</titel>`
	`<erscheinungsform>digital</erscheinungsform>`
	`<entstehungszeitraum>`
	`<von>`
	`<datum>2006</datum>`
	`</von>`
	`<bis>`
	`<datum>2008</datum>`
	`</bis>`
	`</entstehungszeitraum>`
	`<aktenzeichen>1</aktenzeichen>`
	`<dateiRef>DAT0</dateiRef>`
	`<dateiRef>DAT1</dateiRef>`
	`<dateiRef>DAT2</dateiRef>`
	`<dateiRef>DAT3</dateiRef>`
	`<dateiRef>DAT4</dateiRef>`
	`<dateiRef>DAT5</dateiRef>`
	`</dossier>`
	...
	`</ablieferung>`
	`</paket>`
	Hinweis: In einem FILES-SIP kann eine Datei auch mehreren Dokumenten oder mehreren Dossiers / Mappen 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.
		M

ID	Beschreibung Anforderung	M/K

M_4.12-2	Das Metadatum `<dateiRef>` kann bei Bedarf mit vier verschiedenen Attributen ergänzt werden: `version, repraesentation, information` oder `reihung`
	
	Beispiel
	Die Datei kann mit einer Version ergänzt werden. Das Attribut `version` kann eine Nummer, ein Dateiname oder ein Entstehungsdatum sein:
	Eine neue Repräsentation (DAT22) einer Ausgangsdatei (DAT21) kann mit dem Attribut `repraesentation` mit der ID der Ausgangsdatei[1] ergänzt werden:
	`<dateiRef version="2">DAT21</dateiRef>`
	`<dateiRef repraesentation="DAT21">DAT22</dateiRef>`
	
	Beschreibende Informationen (z.B. PREMIS-Daten) können mit dem Attribut `information` und der ID der Beschreibenden Datei[2] verbunden werden.
	Bei Dokumenten mit diversen Dateien können mit dem Attribut `reihung` die bestehende Reihung oder Sequence mit Nummer, Dateinamen oder Entstehungsdatum abgebildet werden
	`<dokument id="IdDok-1">`
	`<titel>Lieferschein Glockenaepfel</titel>`
	`<erscheinungsform>digital</erscheinungsform>`
	`<dateiRef reihung="Seite1">DAT303</dateiRef>`
	`<dateiRef reihung="Seite2">DAT304</dateiRef>`
	`<dateiRef information="DAT303">DAT312</dateiRef>`
	`<dateiRef information="DAT304">DAT313</dateiRef>`
	`</dokument>`
		K

4 Struktur SIP

4.1 Rahmenvorgaben und Begrenzungen Grösse Paket

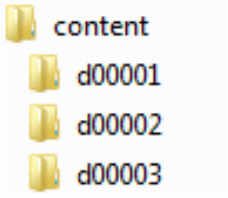
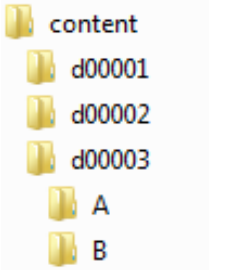
ID	Beschreibung Anforderung	M/K
S_5.1-1	Ein SIP sollte maximal 8 GB gross sein. *Empfehlung* *Es wird empfohlen, aus Gründen der schnelleren Übertragung und Vermittlung, die Grösse eines einzelnen SIP für die Ablieferung unter 2 GB zu halten.* *Dies kann durch eine regelmässige Ablieferung von Unterlagen an das Archiv und durch eine gute Aussonderungsplanung erreicht werden.*	K

[1] Mehrere Ausgangsdateien werden mit Leerschlag voneinander getrennt.

[2] Werden mehrere Dateien gleichzeitig mit weiteren Informationen beschrieben, werden die einzelnen Dateien mit einem Leerschlag voneinander getrennt.

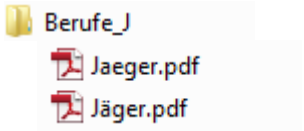
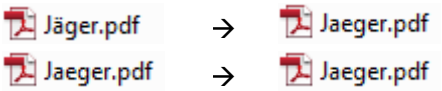
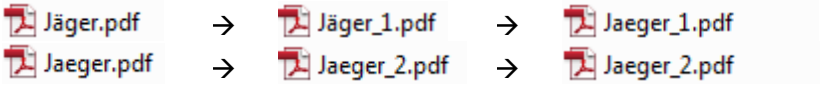
ID	Beschreibung Anforderung	M/K
S_5.1-2	Die abliefernde Stelle muss vor der Erstellung der Ablieferung und dem Transfer des Paketes mit dem Archiv Kontakt aufnehmen, wenn das SIP grösser als 8 GB ist.	M
S_5.1-3	Die abliefernde Stelle muss vor der Erstellung der SIP mit dem Archiv Kontakt aufnehmen, wenn die abzuliefernden Unterlagen aus Gründen der Grösse auf mehrere SIP (und demzufolge auch auf mehrere Ablieferungen) aufgeteilt werden müssen.	M

4.2 Rahmenvorgaben und Begrenzungen Anzahl Dateien im Paket

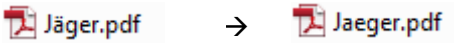
ID	Beschreibung Anforderung	M/K
S_5.2-1	Ein SIP darf nicht mehr als 1’000’000 Dateien enthalten.	M
S_5.2-2	Ein einzelner Ordner innerhalb des SIP sollte nicht mehr als 5'000 Dateien enthalten. *Empfehlung* *Die Anforderung sollte eingehalten werden, weil sonst Probleme mit der Darstellung des SIP auf dem Explorer auftreten können. Falls ein einzelner Ordner mehr als 5000 Dateien enthält, wird empfohlen, Ordner zur weiteren Unterteilung einzufügen.* Beispiel  Der Ordner `d00003/` enthält ursprünglich 9’000 Dateien.  Im Ordner `d00003/` werden zwei Ordner eingefügt, und die 9'000 Dateien so darauf verteilt, das 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.	K

4.3 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
S_5.3-2	Der erlaubte Zeichensatz zur Benennung von Ordnern und Dateien ist ein Subset von US-ASCII. Die folgenden Zeichen sind erlaubt: Buchstaben: A-Z a-z Ziffern: 0-9 Andere: ! # $ % ( ) + , - . = @ [ ] { } ~_ SPACE Die Zeichencodes der erlaubten Zeichen sind in den Zeichensätzen US-ASCII, ISO-8859 und UTF-8 identisch.	M
S_5.3-3	Wenn die Namen von Ordnern und Dateien nicht erlaubte Zeichen enthalten, müssen diese vor der Ablieferung an das Archiv normalisiert werden. *Empfehlung* *Im Anhang J finden sich Empfehlungen im Sinne einer Best Practice, die bei der Normalisierung von Namen befolgt werden können*.	M
S_5.3-4	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. Der ursprüngliche Dateiname muss entsprechend zuvor in den Metadaten festgehalten werden und geht dadurch nicht verloren. Beispiel Kollision bei Normalisierung:   Geeignete Ergänzung:  	M

ID	Beschreibung Anforderung	M/K

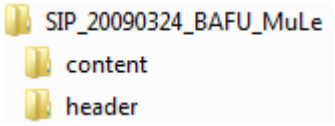
S_5.3-5	Werden existierende Dateien im Paket bei der Paketerstellung (z. B. aus technischen Gründen) umbenannt, müssen die originalen Dateinamen im Metadaten-Element originalName in der Entität Datei eingetragen werden.
	
	Beispiel
	Umbenennung Dateiname bei Dateien im Paket
	 
	
	Einträge in Metadaten-Elementen im metadata.xml
	`<dateiName>Jaeger.pdf</dateiName>`
	`<originalName>Jäger.pdf</originalName>`
		M

4.4 Aufbau eines SIP

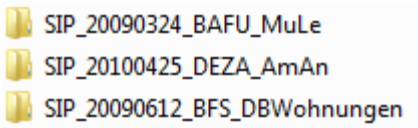
Die Basis des SIP ist eine Datei-Ordner-Struktur. Für diese gibt es sowohl Vorgaben über den Aufbau wie auch für die Benennungskonventionen (Inhalt und Aufbau der Namen) der einzelnen Ordner und Dateien. Als erste Übersicht folgen untenstehend die obersten Ordner eines SIP (Abbildung 10).

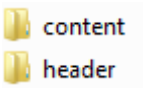
```
SIP_[Identifikatoren]/
      header/
            metadata.xml
            xsd/
                  arelda.xsd
                  ….xsd
                  ….xsd
      content/
```

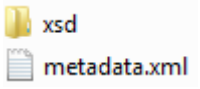
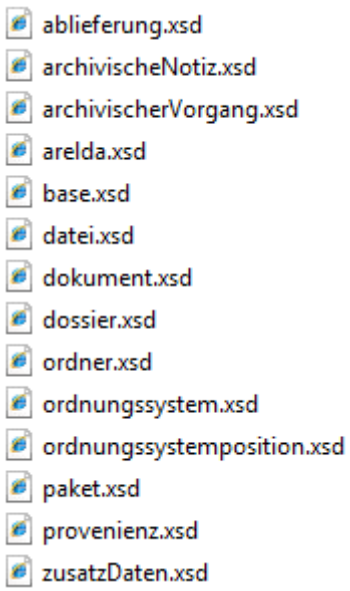
Abbildung 10: Grundstruktur eines Informationspaketes

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

<table>
<tr><th>ID</th><th>Beschreibung Anforderung</th><th>M/K</th></tr>
<tr><td>S_5.4-2</td><td>Der Name des Toplevel-Ordners beginnt bei einem SIP immer mit der Zeichenfolge <code>SIP_</code> gefolgt von bestimmten identifizierenden Angaben [Identifikatoren].

Empfehlung
Der Name des Toplevel-Ordners soll gemäss dem folgenden Muster aufgebaut sein:
<table><tr><td>SIP_</td><td>[Ablieferungsdatum]</td><td>_</td><td>[Name der abliefernden Stelle]</td><td>_</td><td>[Referenz]</td></tr></table>[Ablieferungsdatum]: Aufbau des Ablieferungsdatums: JJJJMMTT
[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.
[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.
Beispiel
Dargestellt sind 3 SIP, bei denen je der Toplevel-Ordner sichtbar ist.
  </td><td>M</td></tr>
<tr><td>S_5.4-3</td><td>Der Toplevel-Ordner enthält immer genau zwei Ordner. Diese Ordner haben die Namen <code>header/</code> und <code>content/</code>.
Es ist nicht erlaubt, andere Ordner oder Dateien direkt im Toplevel-Ordner zu platzieren.

Beispiel
Der Inhalt des Toplevel-Ordners sieht bei allen SIP folgendermassen aus:
  </td><td>M</td></tr>
</table>

ID	Beschreibung Anforderung	M/K
S_5.4-4	Der Ordner `header/` enthält die Metadaten und deren XSD-Dateien. Der Ordner `header/` enthält niemals Primärdaten. Der vordefinierte Ordner im Ordner `header/` eines SIP für die Ablieferung heisst: `xsd/` Es sind keine weiteren Ordner im Ordner `header/` erlaubt. Die vordefinierte Datei im Ordner `header/` heisst immer `metadata.xml`. Diese Datei ist die einzige Datei, die im Ordner `header/` erlaubt ist. In dieser Datei werden die Metadaten des Paketes gespeichert. Beispiel Der Inhalt des Ordners `header/` eines SIP, das abgeliefert wird.  	M
S_5.4-5	Alle XSD-Dateien die zur benutzten XSD-Version gehören, werden im Ordner `header/` in einem separaten Ordner mit dem Namen `xsd/` abgelegt. Die in metadata.xml referenzierte Schemadatei heisst zwingend `arelda.xsd` und wird als `schemaVersion="5.0"` referenziert Beispiel Alle XSD-Dateien für die Metadatenversion ARELDA v5.0 im Ordner `xsd/`.  	M
S_5.4-6	Der Ordner `content/` enthält die Primärdaten eines SIP. Das heisst, der Ordner enthält alle Dateien, aus denen sich die Inhalte der Dossiers / Mappen im SIP zusammensetzen.	M

ID	Beschreibung Anforderung	M/K

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

4.5 Pfadlängen und Hierarchien im SIP

ID	Beschreibung Anforderung	M/K

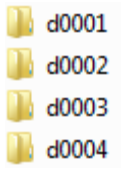
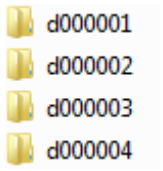
S_5.5-1	Die Pfadlänge zu jeder Datei und zu jedem Ordner innerhalb des Informationspaketes sollte weniger als 180 Zeichen betragen. Der Pfad beinhaltet dabei immer auch den Toplevel-Ordner. Auch die / müssen mitgezählt werden.
	Die Länge des Namens eines Ordners oder einer Datei sollte verkürzt werden, 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.
	
	Beispiel
	`SIP_20091220_EPA_hp/header/metadata.xml`
	Dieser Pfad besteht aus 39 Zeichen.
	`SIP_20091220_SBF_hp/content/d00001245/p00123453.pdf`
	Dieser Pfad besteht aus 51 Zeichen.
	`SIP_20091220_BFS_hp/content/orange_zettel/zettel_1.tif`
	Dieser Pfad besteht aus 54 Zeichen.
	
	Empfehlung
	Diese Anforderung sollte eingehalten werden, weil sonst Probleme beim Export des SIPs in ein Dateisystems auftreten können
		K

4.6 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 Abbildung 11 ersichtlich.

```
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
```

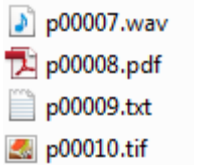
Abbildung 11: Struktur eines GEVER SIP

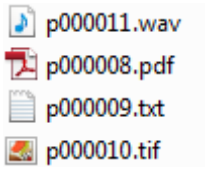
ID	Beschreibung Anforderung	M/K
S_5.6-1	Es wird im Sinne eines Best-Practice-Lösungsansatzes empfohlen, dass jedes Dossier und jedes Subdossier in der Ablieferung einem Ordner im Ordner `content/` oder tiefer entspricht.	K
S_5.6-2	Es wird im Sinne eines Best-Practice-Lösungsansatzes empfohlen, dass bei jedem Ordner, der einem Dossier oder einem Subdossier entspricht, die Ordnernamen mit dem Buchstaben `d` beginnen. Alle Ordner werden mittels einer Laufnummer durchnummeriert. Muster: d n n n d n n n n n n 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. Die Laufnummern (aus `d` und dem Integer) sind als Identifikatoren eindeutig und werden innerhalb des SIP nur einmal vergeben. Beispiel Ein Ordner, der einem Dossier oder einem Subdossier entspricht, wird wie folgt benannt: d n n n n  oder d n n n n n n  	K

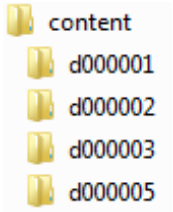
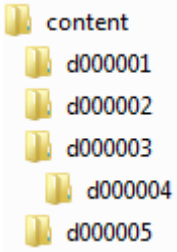
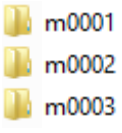
<table>
<tr><th>ID</th><th>Beschreibung Anforderung</th><th>M/K</th></tr>
<tr><td>S_5.6-3</td><td>Benennung von Dateien innerhalb des Ordners <code>content/</code>:
Es wird im Sinne eines Best-Practice-Lösungsansatzes empfohlen, dass bei jeder Datei der Dateiname mit dem Buchstaben <code>p</code> beginnt. 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.
Muster:
<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: ist die Abkürzung für (Primär-)Datei.
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.
[Dateiendung]: Die Dateiendung gehört zum Dateinamen. Sie muss nach dem Punkt ebenfalls angegeben werden.
Die Laufnummern (aus <code>p</code> und dem Integer) sind als Identifikatoren eindeutig und werden innerhalb des SIP nur einmal vergeben.

Beispiel
<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>


oder
<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>

</td><td>K</td></tr>
</table>

ID	Beschreibung Anforderung	M/K
S_5.6-4	Alle Ordner, die Dossiers aus dem GEVER-System entsprechen, sind mit Vorteil auf der obersten Ebene im Ordner `content/` verzeichnet. Die Ordnungssystemstruktur sollte nicht in Ordnern abgebildet werden. Beispiel Oberste Ebene → Ordner für Dossiers im Ordner `content/`.  Weitere Ebenen → Ordner für Subdossiers  Der Ordner `d000004` entspricht einem Subdossier.	K
S_5.6-5	Es wird im Sinne eines Best-Practice-Lösungsansatzes empfohlen, dass bei jedem Ordner, der einer Mappe oder einer Submappe entspricht, die Ordnernamen mit dem Buchstaben `m` beginnen. Alle Ordner werden mittels einer Laufnummer durchnummeriert. Muster: m n n n oder m n n n n n n m: ist die Abkürzung für (Sub-)Mappe 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. Die Laufnummern (aus `m` und dem Integer) sind als Identifikatoren eindeutig und werden innerhalb des SIP nur einmal vergeben. Beispiel Ein Ordner, der einer Mappe oder einer Submappe entspricht, wird wie folgt benannt: m n n n n  	K

4.7 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 Abbildung 12 ersichtlich.

```
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
```

Abbildung 12: Struktur eines FILES SIP

ID	Beschreibung Anforderung	M/K
S_5.7-1	Die ganze Ordnerstruktur der Primärdaten der Ablieferung ist im Ordner `content/` enthalten und abgebildet. Deshalb entsprechen die Ordner im `content/` nicht zwingend Dossiers und Subdossiers sowie Mappen uns Submappen. *Empfehlung* *Es wird im Interesse der Übersichtlichkeit empfohlen, in einem FILES SIP die Ordner, die Dossiers entsprechen, auf einer Hierarchiestufe darzustellen.*	M
S_5.7-2	Für die Benennung von Ordnern und Dateien innerhalb des Ordners `content/` 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.	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 / Mappen 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 Abbildung 13).	M

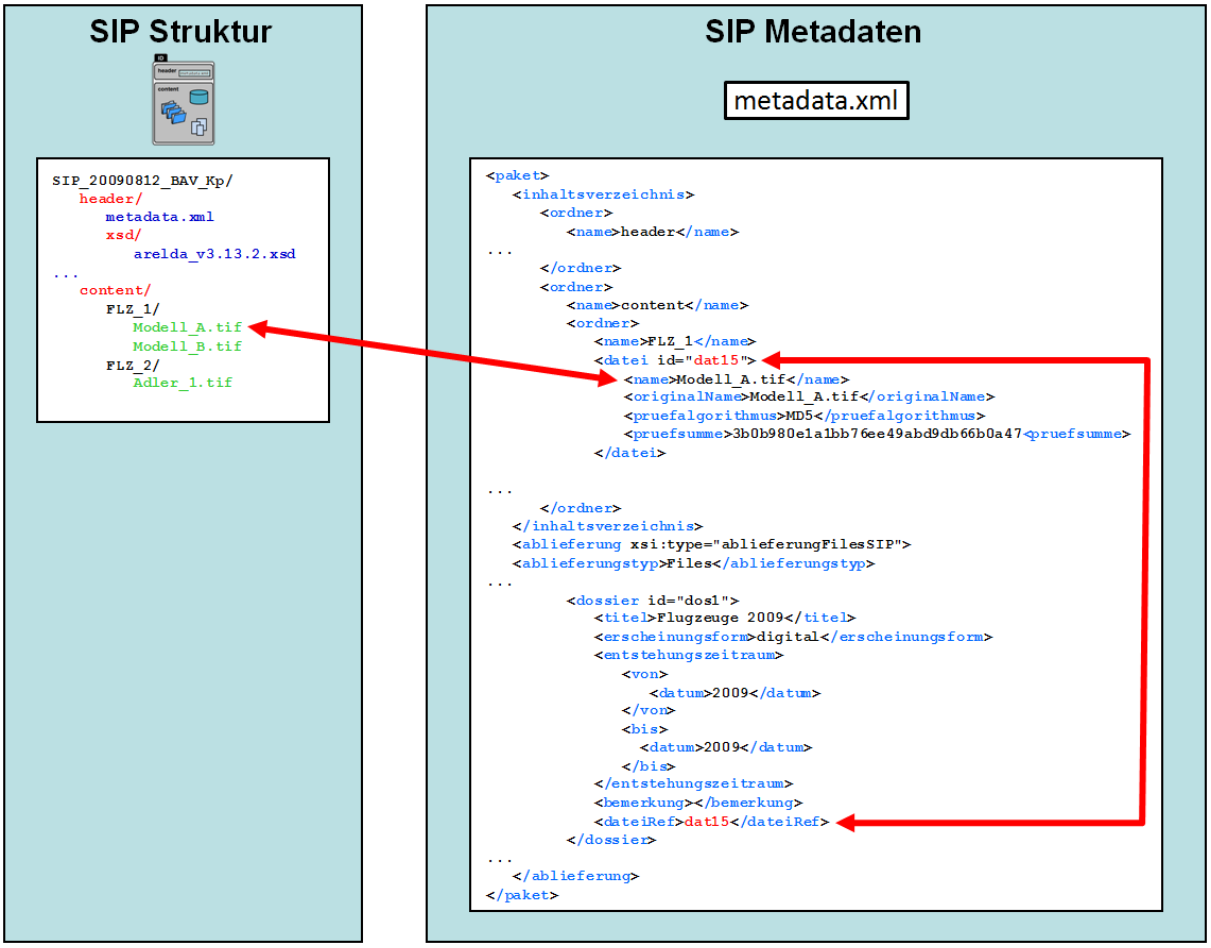


Abbildung 13: Struktur eines FILES SIP – Referenzierung der Datei aus dem Dossier zur Dateibeschreibung im Inhaltsverzeichnis

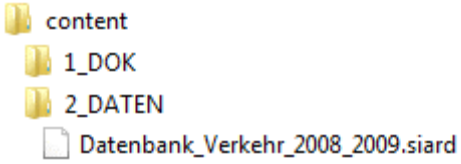
4.8 Aufbau eines FILES SIP mit integrierter Dokumentation

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 Abbildung 14 ersichtlich.

```
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
```

Abbildung 14: 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	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. Beispiel 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). 	M

5 Transfer

5.1 Format Package Container

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

6 Version und Gültigkeit der Spezifikation

Die vorliegende Version 1.2 ersetzt nicht die Version 1.0 und 1.1. Das heisst, dass die Version 1.0 und 1.1 nach wie vor gültig bleibt. Ältere Pakete zur Ablieferung digitaler Unterlagen (SIP) der Version 1.0 und 1.1 sind auch mit den XSD-Dateien der Version 1.2.0 valide und entsprechend ist die Version 1.2.0 rückwärtskompatibel.

Die Änderungen sind im Anhang E ersichtlich.

7 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.

8 Sicherheitsüberlegungen

Keine

9 Haftungsausschluss/Hinweise auf Rechte Dritter

eCH-Standards, welche der Verein **eCH** dem Benutzer zur unentgeltlichen Nutzung zur Verfügung stellen oder welche **eCH** referenzieren, 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.

10 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	◆ Change und Release Management von eCH-Standards
[EAD]	◆ 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 2.01 –Bundeskanzlei vom 11. Februar 2015 https://www.bk.admin.ch/bk/de/home/digitale-transformation-ikt-lenkung/ikt-vorgaben/standards/i017-gever_metadaten.html
[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
[xIsadg DataDictionary]	◆ Rahmenspezifikation für die digitale Archivierung (Projekt bentō): Descriptive Information: Data Dictionary & XML Schema Beschreibung (xIsadg_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

Christa Ackermann	Schweizerisches Bundesarchiv
Marguérite Bos	Schweizerisches Bundesarchiv
Georg Büchler	KOST
Markus Lischer	Staatsarchiv Luzern
Martin Lüthi	Staatsarchiv St. Gallen
Martin Kaiser	KOST, bis 31.05.2021
Lambert Kansy	Staatsarchiv Basel-Stadt
Nicole Martini	Schweizerisches Bundesarchiv
Claire Röthlisberger-Jourdan	KOST
Ralph Ruch	Staatsarchiv Schwyz
Tobias Wildi	Docuteam

Anhang C – Abkürzungen und Glossar

Abkürzung	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]

Abkürzung	Beschreibung
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).

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 abliefert. 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).

Begriff	Beschreibung
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.
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.

Begriff	Beschreibung
Schutzfrist	Über die Schutzfristen wird der Zugang zum Archivgut geregelt. Unterlagen, welche noch einer Schutzfrist unterstehen, dürfen (ausser von den 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 D – Änderungen gegenüber Vorversion

Folgende Änderungen von der Version 1.1 zur Version 1.2.0 vorgenommen.

Kapitel / ID / Dokument	Anpassung	RFC
Kap. Zusammenfassung	Es sollten mehrere Testpakete auf der KOST-Homepage zur Verfügung gestellt werden. In der Zusammenfassung wird darauf verlinkt.	2016-62
M_4.1-2	Definition der XSD Version und XML-Namespace	
M_4.1-3	Dringende Empfehlung auf Präfixe im XML-Namespace zu verzichten	2016-58
Kap. 4.2ff, XSD-Dateien & DataDictionary	Ergänzung mit der Entität Mappe, damit aktuelle FILES Ablagen und GEVER-Systeme besser abgebildet werden können.	2021-20
Kap. 4.6, XSD-Dateien & DataDictionary	Verweis auf das lokale Schema sowie deren Version 5.0	
M_4.12-2, XSD-Dateien & DataDictionary	Ergänzung des XML-Elements dateiRef mit vier optionalen Attributen "reihung", "information", "repraesentation" und "version.	2021-16
Anhang «Gültigkeit Kapitel»	Dieser Anhang war unnötig und wurde gelöscht.	
Anhang J	Anhang wurde ausgelagert in das Hilfsmittel eCH-0257, Metadatentransformationen zu eCH-0160, und enthält nur noch den entsprechenden Verweis.	
DataDictionary	Das DataDictionary stimmt nicht immer mit den XSD-Dateien überein und wurde entsprechen angepasst.	2020-62 2021-17

Kapitel / ID / Dokument	Anpassung	RFC
XSD-Dateien	Erstellen des Attributs «reihung» damit bei Bedarf die Reihenfolge festgehalten werden kann. Zudem soll «order» soll durch «reihung» ersetzt werden.	2021-18
XSD-Dateien	Fokusierung auf das Technische und Wesentliche zur Definition korrekter SIP. • Reduktion der Annotationen (Beschreibungstext im DataDictionary vorhanden) • AIP und DIP Teile entfernt; diese sind irrelevant für das SIP	

Tabelle 2: Änderungen gegenüber Vorversion

Anhang E – Abbildungsverzeichnis

Abbildung 1: Abgrenzung Ablieferungsobjekt SIP 9

Abbildung 2: Ablieferungskategorien SIP 10

Abbildung 3: OAIS Modell mit Kennzeichnung SIP 11

Abbildung 4: Aufbau SIP – Metadaten und Primärdaten 12

Abbildung 5: Konzeptuelles Datenmodell– Anwendungsbereich Ablieferungstyp GEVER ... 18

Abbildung 6: Möglichkeiten der Beziehungen zwischen Dossier und Mappe zu Dokument (vereinfachte GEVER-Darstellung) 18

Abbildung 7: Konzeptuelles Datenmodell– Anwendungsbereich Ablieferungstyp FILES 19

Abbildung 8: Möglichkeiten der Beziehungen zwischen Dossier, Mappe und Dokument zu Datei (vereinfachte FILES-Darstellung) 20

Abbildung 9: Möglichkeiten der Beziehungen zwischen Ordnungssystem und Mappe zu Ordnungssystemposition (vereinfachte FILES-Darstellung) 21

Abbildung 10: Grundstruktur eines Informationspaketes 32

Abbildung 11: Struktur eines GEVER SIP 36

Abbildung 12: Struktur eines FILES SIP 40

Abbildung 13: Struktur eines FILES SIP – Referenzierung der Datei aus dem Dossier zur Dateibeschreibung im Inhaltsverzeichnis 41

Abbildung 14: Struktur eines FILES SIP mit integrierter Dokumentation 42

Anhang F – Tabellenverzeichnis

Tabelle 1: Pakettypen 11

Tabelle 2: Änderungen gegenüber Vorversion 51

Anhang H – 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 einige 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)
...	...	...	...	...	...	...	...
titel	Kurze Beschreiung des im Dokument behandelten Gegenstandes	m	text4		-	ja	Eigenschaft (Feld XXXX)
autor	Verfasser eines Dokuments	k	text4		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	enu-meration	digital, nicht digital		fix	fixer Eintrag: digital Begründung: alle Unterlagen sind digital (keine analogen Teile)
dokumenttyp	Aussehen des Dokuments, das zu erhalten ist.	k	text3		-	nein	Begründung: kann für vorliegende Dokumente nicht spezifisch angegeben werden.
...	...	...	...	...	...	...	...

Anhang I – 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.

I.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 Zќrcher	
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

I.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 (I.2.1). Mit einigen zusätzlichen Restriktionen für Dateinamen (I.2.2), die durch die Betriebssysteme gegeben sind, werden nachfolgend Normalisierungstabellen zur Umwandlung der Dateinamen vorgeschlagen (I.2.3ff).

I.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).

I.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) ! $ & ' ( ) * + , ; = (subcomponent 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.

I.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	~

I.2.2.2 Nicht erlaubte Zeichen

Nicht erlaubt sind diejenigen Zeichen, die per Definition in Betriebsystemen 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*	

US-ASCII / ISO-8859-1	Unicode	Unicode Description	Symbol
...	...		
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	\
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.

I.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 I.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	)

Von...				Nach...		
US-ASCII / ISO-8859-1	Unicode	Unicode Description	Symbol	Unicode	Unicode Description	Symbols
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
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	~

I.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	*ƒ*	0x66	f
0x84	„	0x27	'
0x85	…	0x2E 0x2E 0x2E	...
0x86	†	0x5F	_
0x87	‡	0x5F	_
0x88	ˆ	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	---

Von...		Nach...	
CP-1252	Symbol	US-ASCII	Symbols
0x98	˜	0x7E	~
0x99	™	0x54 0x4D	TM
0x9A	š	0x73	s
0x9B	›	0x27	'
0x9C	œ	0x6F 0x65	oe
0x9D		0x5F	_
0x9E	ž	0x7A	z
0x9F	Ÿ	0x59	Y

I.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-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	_

Von...				Nach...	
ISO-8859-1	Unicode	Unicode Description	Symbol	US-ASCII	Symbols
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	_
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

Von...				Nach...	
ISO-8859-1	Unicode	Unicode Description	Symbol	US-ASCII	Symbols
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
0xF7	U+00F7	Division sign	÷	0x5F	_
0xF8	U+00F8	Latin Small Letter O with stroke	ø	0x6F	o

Von...				Nach...	
ISO-8859-1	Unicode	Unicode Description	Symbol	US-ASCII	Symbols
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.

I.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 J – Beziehung zu weiteren archivischen Standards

Das Hilfsmittel eCH-0257, Metadatentransformationen zu eCH-0160, beschreibt diverse *Crosswalks* beziehungsweise Metadatentransformationen zwischen der archivischen Ablieferungsschnittstelle [eCH-0160] und den massgeblichen archivischen Standards wie dem Verzeichnungsstandard [ISAD(G)] oder seiner Kodierung in [EAD].