

Landesarchiv Baden-Württemberg
Projekt „Konzeption für ein digitales Landesarchiv“
Metadaten für die Archivierung digitaler Unterlagen

AZ: V-7511.9-5/Kei, Lg, Nau
Stand: Dezember 2008

Inhalt

1	Vorbemerkung	1
2	Metadaten zur Beschreibung und Archivierung.....	2
2.1	Grundlagen	2
2.1.1	Prämissen.....	2
2.1.2	Bereiche und Ebenen der Beschreibung	2
2.1.3	Strukturierte Metadaten, Dokumentation und integrierte Metadaten	2
2.1.4	Veränderbarkeit der Metadaten	3
2.2	Metadaten für alle Archivalientypen.....	3
2.2.1	Allgemeine Metadaten	3
2.2.2	Strukturdaten	4
2.2.3	Objektdaten	4
2.2.4	Repräsentationsdaten.....	5
2.2.5	Dateidaten (nur Archivalien)	5
2.2.6	Dokumentationsdaten.....	6
2.2.7	Versionsdaten.....	7
2.3	Metadaten für Datenbanken	7
2.3.1	Verzeichnisstruktur und Metadatenfelder	7
2.3.2	Objektdaten	8
2.3.3	Repräsentationsdaten.....	8
2.3.4	Dateidaten (nur Archivalien)	8
2.3.5	Datenfelder	8
2.3.6	Exkurs: Kapselung von Feld-Metadaten und Primärdaten	9
2.3.7	Codelisten.....	9
2.4	Metadaten für digitale Bilder	10
2.4.1	Umfang und Zweck spezifischer Bild-Metadaten.....	10
2.4.2	Felder.....	10
3	Prozesse und Protokolle.....	10
3.1	Grundlagen	10
3.2	Felder	11
3.3	Prozesse und Tätigkeiten	11
4	Metadaten beim Ingest von Fachverfahren	12
5	Hashwerte	15

1 Vorbemerkung

Das vorliegende Papier stellt die für die Archivierung digitaler Unterlagen notwendigen Metadaten in ihren logischen Zusammenhängen dar. Die Ergebnisse wurden im Rahmen eines vom Land Baden-Württemberg geförderten Projekts vom Januar 2006 bis zum Dezember 2008 von Christian Keitel (Projektleitung), Rolf Lang und Kai Naumann erarbeitet. Aufgrund der raschen Entwicklungen im Bereich der digitalen Archivierung kann es sich dabei nur um eine Momentaufnahme zum jetzigen Stand (Dezember 2008) handeln, die der kontinuierlichen Anpassung an künftige Anforderungen und Forschungsergebnisse bedarf. Aus praktischen Erwägungen werden vier Metadaten-Bereiche voneinander abgegrenzt. Zunächst werden die für die Beschreibung und Archivierung notwendigen Metadaten beschrieben. Diese Metadaten erfahren nach Abschluss der Eingangsbearbeitung nur wenige Änderungen. Danach werden die sich häufiger verändernden da anwachsenden Metadaten aufgeführt, die für die einzelnen Protokolle und die in ihnen nachgewiesenen Prozesse benötigt werden. In einem dritten Bereich werden die Metadaten beschrieben, die während des Ingest-Prozesses anfallen und derzeit mit dem Tool IngestList erhoben werden. Alle hier dargestellten Metadaten werden derzeit in dem System DIMAG 1 gespeichert. Bei allen durch

das Tool IngestList verwalteten und in DIMAG archivierten Dateien wird automatisch der Hashwert erfasst. Die Angaben zu diesen Hashwerten bilden die vierte Metadaten-Gruppe.

2 Metadaten zur Beschreibung und Archivierung

2.1 Grundlagen

2.1.1 Prämissen

- Definition Metadaten und Primärdaten: Metadaten sind Daten, die andere Daten beschreiben. Die anderen Daten können Primärdaten (Archivalien) oder Metadaten sein. Als Archivalien bzw. Primärdaten werden die Daten angesehen, die eigentlich erhalten werden sollen. Diese Einordnung bedeutet, dass die Inhalte auch bei offenkundiger Fehlerhaftigkeit nicht korrigiert werden (z.B. falsches Entstehungsdatum in einem Scan bzw. auf dessen nicht mehr vorhandener Vorlage). Im Gegensatz dazu sollten die Metadaten bei offenkundigen Fehlern korrigiert werden.
- Es gibt einheitliche Verfahren zur Anlage der Metadaten.
- Komplexe Metadatenstrukturen erhöhen den Bearbeitungsaufwand und erschweren eine spätere Migration.
- Es sollten so wenige Felder wie möglich angelegt werden (Aufwand).
- Daten, die in der Behörde nur als Hilfsmittel verwendet wurden (also als Metadaten angesehen wurden), können in dem Archiv als Primärdaten klassifiziert werden. Bei einer Verwendung als Findmittlersatz (Metadaten) sollte eine zusätzliche Ablage als Primärdaten erwogen werden.
- Die Signatur dient zur dauerhaften menschlichen Benennung. Zur dauerhaften maschinellen Identifizierung und zum Metadatenaustausch sind ID-Nummern einzusetzen.

2.1.2 Bereiche und Ebenen der Beschreibung

- DIMAG enthält zwei Bereiche: „Archivalien“ und „Übergeordnete Dokumentation“.
- Die Elemente, aus denen DIMAG besteht, werden durch Metadaten beschrieben.
- Das DIMAG kennt in den Bereichen „Archivalien“ und „Übergeordnete Dokumentation“ vier unterschiedliche Beschreibungsebenen (Struktur, Objekt, Repräsentation und Datei/Version).

2.1.3 Strukturierte Metadaten, Dokumentation und integrierte Metadaten

Metadaten erscheinen in DIMAG an drei Orten:

- strukturierte Metadaten: Metadaten, die für mindestens eine Beschreibungsebene einheitlich erhoben werden.
Die strukturierten Metadaten werden entweder für alle Archivalientypen oder nur bei manchen Archivalientypen erhoben (= formatspezifische Metadaten).
- Dokumentation: zusätzliche Angaben zu einem digitalen Objekt, die aus pragmatischen Gründen nicht weiter aufgespalten und in vordefinierte Felder verteilt werden (z.B. weitergehende Angaben zu dem in der Behörde verwendeten System, Handbücher, Bildschirmfotos etc.).
Dokumentation kann logisch als Metadaten begriffen werden, die praktisch wie Primärdaten zu behandeln sind. Dokumentation wird entweder als Digitales Objekt oder als einzelne Datei innerhalb eines Digitalen Objekts dargestellt.
 - Im ersten Fall ist die Dokumentation in sich gliederbar (z.B. mehrere Unterobjekte, Repräsentationen...). Formatbeschränkungen bestehen nicht. Dieser Typ von Dokumentation wird über das Feld „Archivalientyp“ ausgezeichnet.
 - Als Einzeldatei kann Dokumentation nur unterhalb einer Repräsentation eingestellt werden. Formatbeschränkungen bestehen nicht. Dieser Dokumentationsstyp wird über das Feld „Typ“ (Ebenentyp) ausgezeichnet.

- **Integrierte Metadaten:** Daten, die in die Dateien der Primärdaten integriert sind, z.B. Angaben im Header einer TIFF-Datei. Diese Angaben werden nur dann extrahiert (in den Bereich der strukturierten Metadaten überführt), wenn sie für Recherche-, Verwaltungs- oder Erhaltungsaufgaben unerlässlich sind. Sie werden auch extrahiert, wenn sie durch die Übernahme oder durch eine Migration verloren zu gehen drohen und archivwürdig sind.

2.1.4 Veränderbarkeit der Metadaten

Die Veränderungen in den Metadaten sollen in vertretbaren Maßen nachvollziehbar gehalten werden. Dieses Ziel wird durch die Mittel „Status“ und „Version“ erreicht. DIMAG kennt drei unterschiedliche Status:

- Status „In Bearbeitung“: Ältere Zustände werden überschrieben.
- Nachdem der Status auf „Abgeschlossen“ gesetzt wurde, wird bei der Veränderung der Metadaten automatisch eine neue Version erstellt und die ältere beibehalten.
- Es kann sein, dass Unterlagen definitiv nicht mehr benötigt werden (z.B. Dateien, die irrtümlicherweise archiviert wurden). Diese Dateien können dann einschließlich ihrer Metadaten gelöscht werden.

Statuswechsel können nur in einer Richtung erfolgen, d.h. von „In Bearbeitung“ zu „Abgeschlossen“. Der Status wird primär im Feld „StatusCode“ festgehalten, das Textfeld „Status“ deklariert die Bedeutung des Codes.

Die in einem automatisierten Übernahmeprozess mit dem Werkzeug IngestList festgehaltenen Prozessinformationen werden direkt nach der Übernahme nach DIMAG auf den Status „Abgeschlossen“ gesetzt, um einer Manipulation vorzubeugen.

2.2 Metadaten für alle Archivalientypen

2.2.1 Allgemeine Metadaten

Diese Metadaten fallen auf jeder Ebene an (Struktur, Objekt, Repräsentation und Datei). Das Feld „Titel“ gehört nicht zu den allgemeinen Metadaten, weil es auf der Ebene Primärdatei nicht vorkommt.

	System Anwender Pflicht Optional	Beschreibung	Eigenschaften
Signatur	S	Eindeutige Auszeichnung	
Signaturanteil	A/P	Anteil, der nur die aktuelle Ebene beschreibt.	Freitext
Beschreibung/Inhalt	A/O	Nähere Angaben zum Titel.	Freitext
Typ	S	Ebenentyp	
Status	S	vgl. Kap. 2.1.4	Freitext
StatusCode	A/P	vgl. Kap. 2.1.4	Ganzzahl
Ersteller	S	Ersteller des Objekts in DIMAG	
Erstellungsdatum	S	Datum der Einstellung in DIMAG	
Geändert durch	S	Urheber einer Änderung	
Änderungsdatum	S		
Versionsnummer	S	höchste Zahl ist aktuelle Version	
XML-Version	S	Stand des XML-Schemas	JJJJ-MM-TT
Quirks / Eigenheiten	A/O	Zeigt eine inhaltliche oder technische Unzulänglichkeit an, die auf Eigenschaften der Ausgangsdaten zurückgeht, z.B. Eintrag „grün“ im Feld „Geschlecht“ oder „nicht valides Tiff“.	Freitext
Rechte Level	A/P	Definiert das Rechtelevel, mit dem dieses Objekt gelesen bzw. bearbeitet werden kann.	Auswahlfeld, Vorauswahl „read_pub“

2.2.2 Strukturdaten

Strukturdaten beschreiben die oberhalb der Digitalen Objekte liegenden Einträge (Tektonik und Klassifikation). Sie sind weitgehend mit den Allgemeinen Metadaten identisch. Die Einträge können rekursiv angelegt werden.

	System Anwender Pflicht Optional	Beschreibung	Eigenschaften
Allgemeine Metadaten (siehe oben)			
Titel	A/P		

2.2.3 Objektdaten

Mit diesen Feldern werden die Digitalen Objekte beschrieben. Digitale Objekte sind die bestellbaren Einheiten. Die Einträge können rekursiv angelegt werden.

	System Anwender Pflicht Optional	Beschreibung	Eigenschaften
Allgemeine Metadaten (siehe oben)			
Titel	A/P		
Entstehungszeitraum	A/P	Zeitraum, in dem das Objekt bei der Ursprungsstelle entstanden ist.	
abweichend dokumentierter Zeitraum	A/P	Zeitraum, der in dem Objekt abgebildet wird.	
Provenienz	A/P	Obsolet, wenn IngestList verwendet wurde.	Freitext
Abgebende Stelle	A/O	Obsolet, wenn IngestList verwendet wurde.	Freitext
Übernahme	A/O	Zeitpunkt, Beteiligte und Umstände. Obsolet, wenn IngestList verwendet.	Freitext
Archivalientyp	A/P		erweiterbare Auswahlliste (DMS, Fachverfahren, Plan, Textdokument, Foto, Statistik)
Signifikante Eigenschaften	A/O	Eigenschaften, die auch nach Migrationen oder Konversionen erhalten bleiben sollten.	Freitext
Sperrfrist	A/P	Jahr, in dem die Sperrfrist ausläuft	Freitext
Benutzungshinweise	A/O	Hinweise zu Einschränkungen bei der Nutzung nach Archivrecht.	Freitext
Rechte	A/O	Hinweise zu weitergehenden Einschränkungen nach Urheber- oder Leistungsschutzrechten.	Freitext
Papierte Bestandteile: Referenz	A/O	Angaben zum papiernen Bestandteils (= Primärdaten) des Objekts.	Freitext
Papierdokumentation: Signatur	A/O	betr. Metadaten auf Papier, die wegen ihres Umfangs nicht gesammelt werden können.	Freitext

	System Anwender Pflicht Optional	Beschreibung	Eigenschaften
Papierdokumentation: Titel	A/O	Angaben zur papiernen Dokumentation (= Metadaten) zum Objekt. Betr. Metadaten auf Papier, die wegen ihres Umfangs nicht gesammelt werden können.	Freitext
Papierdokumentation: Inhalt	A/O		Freitext
Papierdokumentation: Lagerort	A/O		Freitext

2.2.4 Repräsentationsdaten

Eine Repräsentation ist die Erscheinungsform eines digitalen Objekts. Ein digitales Objekt kann mehrere Repräsentationen haben, die jeweils dieselben Informationen enthalten. Repräsentationen können nicht rekursiv angelegt werden.

	System Anwender Pflicht Optional	Beschreibung	Eigenschaften
Allgemeine Metadaten (siehe oben)			
Titel	A/P	Beschreibt die Art der Repräsentation, nicht den Inhalt.	Freitext
Struktur	A/O	z.B. Multimediadatei, Webauftritte, verschiedene Formate, verschiedene Bestandteile, funktionale Zusammenhänge	Freitext
Hardware-Umgebung	A/O	Dokumentiert die Hardware, die zum Bearbeiten dieser Repräsentation im Zeitpunkt ihres Entstehens benötigt wurde.	Freitext
Software-Umgebung	A/O	Dokumentiert die Software, die zum Bearbeiten dieser Repräsentation im Zeitpunkt ihres Entstehens benötigt wurde.	Freitext
Installationserfordernisse	A/O	Dokumentiert die Erfordernisse, die für das Öffnen des Objekts nötig sind.	Freitext
Basis-Repräsentation	A/O	Dokumentiert, welche Repräsentation als Basis für diese Repräsentation verwendet wurde.	Freitext

2.2.5 Dateidaten (nur Archivalien)

Die Dateidaten beziehen sich unmittelbar auf die Dateien mit Primärdaten. Die Dateidaten können nur unterhalb einer Repräsentation angelegt werden.

	System Anwender Pflicht Optional	Beschreibung	Eigenschaften
Allgemeine Metadaten (siehe oben).			
Original-Dateiname	S	Dateiname vor dem Einstellen in den Massenspeicher	
Dateiname	S		

	System Anwender Pflicht Optional	Beschreibung	Eigenschaften
Pronom Format	S	Formatname nach PRONOM.	Ohne Klammern, wenn eindeutig identifiziert. Mehrere Einträge möglich.
Pronom Persistent ID	S	Format-ID nach PRONOM.	Ohne Klammern, wenn eindeutig identifiziert. Mehrere Einträge möglich.
JHOVE Format	S	Formatname nach JHOVE	
Format Version	S	Formatversion laut JHOVE	
MIME-Type	S	Internet Media Type gemäß Standard Multipurpose Internet Mail Extensions (MIME) nach JHOVE.	
Zeichenformat	A/O		Auswahlfeld (ASCII, UTF-8, ...)
Dateigröße in Byte	S		
MD5-Wert	S	MD5-Wert der beschriebenen Primärdatei.	

Anmerkung: Es fallen teilweise noch formatspezifische Metadaten an, deren Eingabefelder nach einem entsprechenden Eintrag in „Archivallientyp“ angezeigt werden. Vgl. S. 8, S. 10.

2.2.6 Dokumentationsdaten

Die Dokumentationsdaten beziehen sich unmittelbar auf die Einzeldateien, die Dokumentation enthalten.

	System Anwender Pflicht Optional	Beschreibung	Eigenschaften
Allgemeine Metadaten (siehe oben)			
Titel	A/P		
Original-Dateiname	S	Dateiname vor dem Einstellen in den Massenspeicher	
Dateiname	S		
Pronom Format	S	Formatname nach PRONOM.	Ohne Klammern, wenn eindeutig identifiziert. Mehrere Einträge möglich.
Pronom Persistent ID	S	Format-ID nach PRONOM.	Ohne Klammern, wenn eindeutig identifiziert. Mehrere Einträge möglich.
JHOVE Format	S	Formatname nach JHOVE	
Format Version	S	Formatversion laut JHOVE	
MIME-Type	S	Internet Media Type gemäß Standard Multipurpose Internet Mail	

	System Anwender Pflicht Optional	Beschreibung	Eigenschaften
		Extensions (MIME) nach JHOVE (0,1).	
Zeichenformat	A/O		Auswahlfeld (ASCII, UTF-8, ...)
Dateigröße in Byte	S		
MD5-Wert	S	MD5-Wert der beschriebenen Dokumentationsdatei.	

Anmerkung: Es fallen teilweise noch formatspezifische Metadaten an, deren Eingabefelder nach einem entsprechenden Eintrag in „Archivalientyp“ angezeigt werden. Vgl. S. 8, S. 10.

2.2.7 Versionsdaten

Die infolge Versionierung veralteten Stände von Metadaten werden als Dateien im XML-Format im Massenspeicher abgelegt, nicht jedoch in der Datenbank von DIMAG.

2.3 Metadaten für Datenbanken

2.3.1 Verzeichnisstruktur und Metadatenfelder

Für eine ausreichende Beschreibung von Datenbanken sind Angaben notwendig, die über die allgemeinen Metadaten hinausgehen. Diese Metadaten können strukturiert erhoben oder als weiterführende Dokumentation abgelegt werden. Bei der Abwägung, welche Informationen strukturiert, welche unstrukturiert erhoben und wie diese abgelegt werden sollten, sind mehrere Parameter zu beachten:

- **Aufwand:** Mit jeder manuellen Erhebung eines zusätzlichen strukturierten Metadatenfelds steigt der archivistische Aufwand.
- **Benutzung:** Nutzer von archivierten Datenbanken und statistischen Tabellen schätzen die Möglichkeit, mühelos Abfragen zu generieren und einzelne Informationen stets am selben Ort zu finden (vgl. National Digital Archive of Datasets des britischen Nationalarchivs).
- **Komplexität:** Mit der Zahl der zu einem Objekt gehörenden Metadatendateien steigt die Wahrscheinlichkeit späterer Zuordnungs- und Erhaltungsprobleme.
- **Links:** Es ist schwierig, physikalisch angelegte Verweise über lange Zeit zu erhalten.
- **Verzeichnung:** Nach ISAD (G) soll eine Information wenn möglich nicht bei allen Objekten einer Ebene wiederholt, sondern auf einem übergeordneten Verzeichnungsobjekt abgelegt werden.

Es ist nicht möglich, bei der Archivierung von Datenbanken allen Regeln gleichermaßen gerecht zu werden. Evident wird dies bei der Beschreibung der Felder: Die inhaltliche Beschreibung der Felder ist für alle Repräsentationen eines Objekts gleich, sie sollte demnach auf Objektebene abgelegt werden. Gleichzeitig können sich die technischen Eigenschaften der Felder von Repräsentation zu Repräsentation ändern. Diese Angaben sollten also auf der Ebene der Repräsentationen oder noch tiefer abgelegt werden. Das getrennte Ablegen der inhaltlichen und technischen Informationen zu einem Feld erhöht die Komplexität und macht interne Links erforderlich. Die Anforderungen müssen daher gegenseitig abgewogen und ausbalanciert werden.

Der vorliegende Ansatz legt datenbankspezifische Metadaten zur Datei zusammen mit den für alle Archivalientyp anfallenden Metadaten in einer Datei ab. Datenbankspezifische Metadaten zum Objekt oder zur Repräsentation fallen nicht an. Technische und inhaltliche Informationen zu den einzelnen Feldern einer Tabelle (d.h. einer Primärdatendatei) werden in

XML beschrieben und gemeinsam mit der Datei in einer Repräsentation abgelegt. Auch zu den Codelisten wird eine eigene Datei angelegt (vgl. Kap. 2.3.7).

Mehrere Tabellen können entweder mit je einer Felder-XML oder mit einer Felder-XML für alle Tabellen beschrieben werden.

2.3.2 Objektdaten

Keine Zusätze

2.3.3 Repräsentationsdaten

Keine Zusätze

2.3.4 Dateidaten (nur Archivalien)

	System Anwender Pflicht Optional	Beschreibung	Eigenschaften
Anzahl Felder (Spalten)	A/O		Freitext
Anzahl Datensätze (Zeilen)	A/O		Freitext

2.3.5 Datenfelder

	System Anwender Pflicht Optional	Beschreibung	Eigenschaften
Feldname	A/P		Freitext, max. 32 Zeichen.
Beschreibung/Inhalt	A/O		Freitext
Feldtyp	A/O		Auswahl: Text, Ganzzahl, Fließkommazahl, Festkommazahl, Datum, Binär
Feldlänge	A/O	In Textfeldern: Maximale Länge der enthaltenen Information in Anzahl der Zeichen. ¹ Bei Festkommazahlen werden die Stellen vor und hinter dem Komma, durch Komma separiert, angegeben.	Zahl
Codierung	A/O	Syntax der benutzten Werte.	Freitext
Codes	A/O	Zuordnung Codeschlüssel-Klartext. Nur wenn Codeliste höchstens 10 Elemente umfasst.	Sub-Tabelle
Codeliste-Ref-Name	A/O	Zuordnung des Felds zu Codelisten innerhalb der Feldbeschreibung.	Freitext
Verknüpfungen	A/O	Verweise auf andere DB-Dateien mit Beziehungstyp (z.B. externe Codelisten, Fremdschlüssel).	Freitext
Bemerkungen	A/O	z. B. Aufbereitung, Quirks	Freitext

¹ Bei Textfeldern: Die maximale Länge muss nicht konkret an den jeweils vorhandenen Werten ermittelt werden, sondern kann aus den in der Ursprungs-Datenbank hinterlegten maximal möglichen Feldlängen übernommen werden. Bei Zahlenfeldern: Ein Verfahren zur Bestimmung der Feldlänge (in Bytes) erscheint derzeit überflüssig (in Zukunft zu überprüfen). Bei Datumsfeldern: Die Länge des Datumswerts wird einschließlich der Trennzeichen eingetragen.

2.3.6 Exkurs: Kapselung von Feld-Metadaten und Primärdaten

Nach dem oben beschriebenen Konzept werden zu einer Datenbankdatei (Tabelle) zugleich zwei Metadaten-dateien (zur Datei und zu den Feldern) angelegt. Um die Einheit von Primär- und Metadaten zu stärken und die Gefahr einer fehlerhaften Zuordnung zu verringern, könnten die Feld-Metadaten zusammen mit den Primärdaten in einer XML-Datei abgelegt werden. Alternativ können Meta- und Primärdaten auch in zwei unterschiedlichen Dateien abgelegt werden.

- **Meta- und Primärdaten in einer XML-Datei:** Diese Lösung verspricht v.a. dann Erfolg, wenn diese Datei unmittelbar aus dem System der Behörde heraus (also im Zuge des Exports) erzeugt werden kann. In diesem Fall verhindert sie eine spätere fehlerhafte Zuordnung von Primär- und Metadaten zuverlässig. Allerdings werden derzeit Metadaten und Primärdaten in unterschiedlichen Dateien übernommen. Für die Ein-Datei-Lösung spricht, dass eine Zuordnungsprüfung der Meta- und Primärdaten anhand eines normalen Parsers relativ schnell durchgeführt werden kann. Andererseits entstehen durch den Zwang, auch die Primärdaten in XML abzulegen, schnell sehr große Dateien.
- **Meta- und Primärdaten in zwei XML-Dateien:** Gegen diese Lösung spricht die Dateigröße. Besser wäre die Ein-Datei-Lösung.
- **Primärdaten in CSV unter Einfügung der Feldnamen in die erste Zeile:** Für diese Lösung sprechen die weitgehende Trennung von Meta- und Primärdaten, die kleine Dateigröße und der Umstand, dass die Inhalte auch bei fehlender Metadaten-datei rekonstruiert werden können (Voraussetzung: Inhalte sind nicht codiert). Nachteil: Die korrekte Zuordnung der CSV-Datei-Feldnamen zu der Feld-Metadaten-datei muss aufwändiger als bei der Ein-Datei-Lösung programmiert werden.
 - **Metadaten in XML:** Entspricht den sonstigen Metadatenauszeichnungskonventionen im Digitalen Archiv.
 - **Metadaten in CSV:** Widerspricht den Metadatenauszeichnungskonventionen; ist nicht flexibel.

Nach eingehender Diskussion sollen die Primärdaten künftig in CSV und die Feld-Metadaten in einer separaten XML-Datei archiviert werden.

2.3.7 Codelisten

Beschreibung: Codes geben Inhalte in verschlüsselter und verkürzter Form wieder. Beispielsweise steht im Feld „Geschlecht“ oft 1 (für männlich) oder 2 (für weiblich). Codes und Klartext werden einander in zweiseitigen Tabellen zugeordnet. Diese werden als Codelisten, Kataloge oder Lookup-Tabellen bezeichnet.

Auflösung in Klartext: Die Codes sollten nach Möglichkeit in Klartext aufgelöst werden. Dennoch müssen die Codelisten erhalten werden, um evtl. unkorrekte Zuordnungen auch zu einem späteren Zeitpunkt noch korrigieren zu können.

Felder und Codelisten: Ein Feld ist stets einer Codeliste zugeordnet, während eine Codeliste sich auf eine Vielzahl von Feldern beziehen kann, z.B. die Codeliste „Gemeindekennziffer“ auf die Felder „Wohnort“, „Arbeitsort“, „Urlaubsort“ etc. Da die Codes zudem eine andere Funktion als die übrigen Feldbeschreibungen haben, sollten sie nicht in der Datei der Feldbeschreibungen, sondern unabhängig davon geführt werden. Wenn die Anzahl der Codewerte geringer als 10 ist, können die Codes auch in den Metadaten zur Feldbeschreibung abgelegt werden.

Ablageort: Codelisten können sich auf mehrere Repräsentationen beziehen und sollten inhaltlich nicht verändert werden. Solche Codelisten sollten daher auf höherer Ebene abgelegt werden. Letzteres wäre z.B. bei Gemeindekennziffern denkbar.

Codelistendatei(en): Es sind zwei Modelle möglich: Die Codelisten eines Objekts können in einer Datei zusammengefasst oder jeweils pro Unterobjekt/Tabelle oder Feld in einer eigenen Datei angelegt werden. Für die erste Möglichkeit sprechen die leichtere Verwaltung (geringere Zahl an Dateien) und die bessere Übersichtlichkeit. Für die zweite Möglichkeit spricht

die leichtere Übernahme beim Export aus Datenbanksystemen, wenn diese keine Möglichkeit bieten, mehrere Codelisten aneinander zu hängen.

Format: Die Codelistendatei lässt sich gut in einer Tabelle mit zwei Feldern darstellen (Code, Eintrag in Klartext), sie wird im CSV- oder im XML-Format gespeichert.

2.4 Metadaten für digitale Bilder

Digitale Bilder können digital entstandene oder digitalisierte Pläne, Fotos oder Akten sein.

2.4.1 Umfang und Zweck spezifischer Bild-Metadaten

Von Interesse sind solche Metadaten, die für den Fortgang der Archivierung und insbesondere die Festlegung des nächsten Migrationszeitpunkts notwendig sind, d.h. einige allgemeine Angaben zum Format.

Detailinformationen zur technischen Umsetzung künftiger Migrationen sollen dagegen in den Bilddateien belassen und bei Bedarf extrahiert werden (integrierte Metadaten). Hier würde der nur vage zu umschreibende Bedarf den Aufwand nicht rechtfertigen. Diese Metadaten sollen also nicht vorsorglich aus den Bildern extrahiert und in die XML-Metadatendateien geschrieben werden. Ein weiterer Vorteil dieses Verfahrens ist, dass dabei die integrierten Metadaten mit den Primärdaten zusammen in einer Datei abgespeichert werden und dadurch fest mit diesen verbunden sind.

2.4.2 Felder

Die spezifischen Bild-Metadaten fallen auf Datei-Ebene an.

	System/ Anwender	Beschreibung	Eigenschaften
Scandatum	A/O	Datum des Scans; nur bei Scans auszufüllen.	Freitext
Kompression	A/O	Name des Kompressionsverfahrens	Freitext

3 Prozesse und Protokolle

3.1 Grundlagen

- Wesentliche Prozesse und Tätigkeiten werden in Protokoll-Dateien nachgewiesen. Diese Protokolle sollen dem Benutzer einen hinreichend präzisen Eindruck über den Umgang des Archivs mit seinen Digitalen Objekten geben. Sie erhöhen daher die Glaubwürdigkeit des Archivs (Anschein der Authentizität). Daneben können Protokolle auch zu Fragen des archivinternen Umgangs mit den Objekten herangezogen werden.
- Protokolle entstehen auf der Ebene jeden Digitalen Objekts. Zusätzlich wird für übergreifende Prozesse ein Gesamt-Protokoll geführt.
- Jedem Digitalen Objekt ist ein Protokoll (XML-Datei) zugeordnet. Es übernimmt dessen Dateinamen und ergänzt diesen um einen eigenen Zusatz.
 - Label-Name: Protokoll
 - Dateinamenzusatz: prot, z.B. DO_123.prot.xml

- Das Protokoll wird mit dem Erstellen eines Digitalen Objekts automatisch angelegt. Der erste Eintrag weist daher die Anlage des Objekts nach. Vorangehende Prozesse (z.B. Bewertung, Übernahme) werden im Protokoll nicht erfasst.
- Die Prozesseinträge werden jeweils an die bestehende Protokolldatei angefügt.
- Bestehende Einträge können nicht geändert werden.
- Die Unversehrtheit der Protokolldateien wird durch eine MD5-Datei belegt. Nach dem Anlegen eines neuen Eintrags wird jeweils der neue MD5-Wert ermittelt.
- Protokolle werden auch für rekursive digitale Objekte angelegt. Der Bearbeiter kann sich daher bei manuellen Einträgen (s.u.) überlegen, ob er im Protokoll zum über- oder demjenigen zum untergeordneten Objekt den Eintrag vornimmt.
- Die Prozesse können innerhalb des Massenspeichers oder außerhalb erfolgen (konkret: in DIMAG oder in IngestList).
- Die Einträge erfolgen automatisch oder manuell. An die automatisch vorgenommenen Prozesseinträge können (wenn sinnvoll) noch händische Anmerkungen hinzugefügt werden (z.B. warum gelöscht wurde). Die automatisch eingetragenen Informationen können nicht verändert werden.
- Die Prozesse Löschen und Verschieben können nicht sinnvoll auf Objektebene abgelegt werden. Sie werden daher in dem Gesamt-Protokoll abgelegt.

3.2 Felder

Feldname	Eigenschaften	Beispiel/Erklärung
Prozess-Ende	automatischer Eintrag	2006-12-24-20:05
Prozess-Ausführender	automatischer Eintrag	Max Mustermann
Bezug	automatischer Eintrag oder Freitext	Signatur des Objekts, der Repräsentation oder der Datei
Prozess	Auswahlliste	Objekt anlegen
Nähere Angaben	automatischer Eintrag oder Freitext	StAL EL 413_14 DO 99 angelegt. Im erklärenden Text sollten die verschiedenen in Frage kommenden Stichpunkte genannt werden: Beteiligte (d.h. Handelnde und Verantwortliche), Grund, Hardware, Software, Schritte, Richtlinien

3.3 Prozesse und Tätigkeiten

Prozess	Entstehungsort	Ablageort	Erzeugung des Protokolleintrags	Beispiel/Erläuterung
Anlage Objekt	DIMAG	DO	automatisch	StAL EL 413_14 DO 99 angelegt.
Anlage Repräsentation	DIMAG	DO	automatisch	StAL EL 413_14 DO 99 R 2 wurde aus R 1 mit dem Tiff-Spezialkonverter der Firma TifTaf generiert.
Änderung Signatur	DIMAG	Gesamt-Protokoll	automatisch	Signatur von StAL EL 413_14 DO 99 in HStAS EA 77 DO 111 geändert.

Prozess	Entstehungsort	Ablageort	Erzeugung des Protokolleintrags	Beispiel/Erläuterung
Änderung Status	DIMAG	DO	automatisch	Status von „In Bearbeitung“ auf „Abgeschlossen“ geändert.
Änderung Metadaten	DIMAG	DO	automatisch	Metadaten von StAL EL 413_14 DO 99.xml geändert, Version 5 erstellt.
Löschen	DIMAG	Gesamt-Protokoll	automatisch	StAL EL 413_14 DO 99 R 5 gelöscht.
Migration	DIMAG	DO	manuell	Migration von StAL EL 413_14 DO 99 R 1 (tif) in R 2 (pdf-A). Eingesetzt wurde der Tiff-Spezialkonverter der Firma TifTaf.
Notiz	DIMAG	DO	manuell	Beim Erstellen einer neuen Repräsentation ist insbesondere auf ein funktionierendes Zusammenspiel der Felder aa und bb zu achten.
Überprüfung Unversehrtheit	DIMAG	DO	automatisch	Vergleich zwischen den Angaben in Bestandsaufnahme.xml von IngestList und den tatsächlich vorliegenden Daten.
Anlegen Dateiliste	IngestList	DO	automatisch	Legt eine Liste der zu übernehmenden Dateien an.
Formierung	IngestList	DO	automatisch	Ermöglicht die Entnahme einer neuen Bestandsaufnahme.xml auf der Basis einer bestehenden Bestandsaufnahme.xml.
Überprüfung Unversehrtheit	IngestList	DO	automatisch	Vergleicht neu ermittelte MD-5 Werte mit früheren Angaben.
Überprüfung Migration	IngestList	DO	manuell	Validiert das Migrationsergebnis mit den Ausgangsdaten.
Quell Angaben ergänzt	IngestList	DO	auf Nachfrage erstellt mit Ergänzungsmöglichkeit	IngestList ermöglicht es, manche Angaben wie SQL-Statements manuell einzugeben, andere wie Dateigrößen werden automatisch ermittelt. Im Anschluss besteht die Möglichkeit, diesen Prozess ins Protokoll eintragen zu lassen.
Datei Übertragung	IngestList	DO	automatisch	Übertragung der Dateien von IngestList per FTP nach DIMAG.

4 Metadaten beim Ingest von Fachverfahren

Während dem Ingest von Fachverfahren werden zahlreiche Metadaten erhoben, die später in DIMAG als eigene Datei abgespeichert werden (Dateiname: Bestandsaufnahme.xml). Manche Metadaten werden automatisiert aus dieser Datei extrahiert und in die entsprechenden Felder der oben unter 2. geführten strukturierten Metadaten eingefügt. Weitere Angaben können aus Bestandsaufnahme.xml manuell in die strukturierten Metadaten überführt werden. Zusätzlich werden beim Ingest-Prozess weitere Protokolldaten erhoben, die ausnahmslos ins Objektprotokoll in DIMAG übernommen werden (s.o. 3.).

XML-Name in IngestList	Beschreibung	Automatisiert übernommen in strukturierte Metadaten, s.o. 2.
Angaben pro Abgabe/Digitale Objekt		
<datum-erstellung>	Datum der Erstellung	
<person-erstellung>	Ersteller	

XML-Name in IngestList	Beschreibung	Automatisiert übernommen in strukturierte Metadaten, s.o. 2.
<abgebende-stelle>		
<beschreibung>		
<quirks>	Zeigt eine inhaltliche oder technische Unzulänglichkeit an, die auf Eigenschaften der Ausgangsdaten zurückgeht, z.B. Eintrag „grün“ im Feld „Geschlecht“ oder „nicht valides Tiff“. Bezieht sich auf die Ebene des Digitalen Objekts.	
<verzeichnis>	Absoluter Pfad auf dem abgebenden System.	
<format-register-name>	Verweis auf das verwendete Formatregister (derzeit: PRONOM)	
<verwendete-droid-signatur>	Signatur der Formatbeschreibungsdatei der DROID-Software zur Ermittlung des PRONOM-Formats	
<anzahl-dateien>	Zahl der zu übernehmenden Dateien.	
<anzahl-fehler>	Fehler beim Einleseprozess (Zahl)	
<fehler-text>	Beschreibung der Fehler beim Einleseprozess	
<erstellt-mit: Programm>	Verweis auf IngestList	
<erstellt-mit: Version>	Verweis auf IngestList-Version	
Angaben pro Primärdatendatei		
<datei><name>	Dateiname vor dem Einstellen in den Massenspeicher	ja: Original-Dateiname
<pfad>	Relativer Pfad auf dem abgebenden System.	ja
<anzahl-bytes>	Größe der Datei in Bytes.	ja: Dateigröße in Byte
<aenderungs-datum>	Datum der letzten Änderung der Datei	
<md5>	MD5-Werte der Primärdatendatei.	
<pronom-file-format>	Formatnamen nach PRONOM.	ja: Pronom Format
<pronom-format-id>	Format-IDs nach PRONOM.	ja: Pronom Persistent ID
<pronom-classification>	Beschreibt die Qualität der Erkennung (z.B. sicher, unsicher...)	
<pronom-mime-type>	MIME-Typen laut DROID	
<quirks>	Zeigt eine inhaltliche oder technische Unzulänglichkeit an, die auf Eigenschaften der Ausgangsdaten zurückgeht, z.B. Eintrag „grün“ im Feld „Geschlecht“ oder „nicht valides Tiff“.	ja: Quirks / Beschreibung auf Ebene der Datei
<beschreibung>	Nähere Angaben	ja: Beschreibung/Inhalt

XML-Name in IngestList	Beschreibung	Automatisiert übernommen in strukturierte Metadaten, s.o. 2.
<zeichenformat>	Erkennung der Zeichensätze durch IngestList.	
<anzahl-semicolon-erste-zeile>	Anzahl der Semikolons der ersten Zeile.	
<anzahl-lf>	Anzahl des Zeichen „Line-feed“.	
<anzahl-cr>	Anzahl des Zeichens „Carriage Return“.	
<anzahl-tab>	Anzahl des Zeichens „Tabulator“.	
<anzahl-pipe>	Anzahl des Zeichens „Pipe“.	
<anzahl-komma>	Anzahl des Zeichens „Komma“.	
<anzahl-semi-colon>	Anzahl des Zeichen „Semikolon“ (ganze Datei!)	
<anzahl-single-quote>	Anzahl des Zeichens „Einfaches Anführungszeichen“.	
<anzahl-double-quote>	Anzahl des Zeichens „Doppeltes Anführungszeichen“.	
<view-table-name>	Ursprünglicher Tabellename.	
<anzahl-spalten>	Anzahl der Spalten in der ursprünglichen Datenbank.	
<anzahl-zeilen>	Anzahl der Zeilen in der ursprünglichen Datenbank.	
<export-sql>	Zum Export aus der ursprünglichen Datenbank verwendetes SQL-Statement.	
<describe-sql>	Datenbankbeschreibung von ORACLE.	
<jhove-format>	Von JHOVE ermitteltes Format.	ja: JHOVE Format
<jhove-version>	Version des erkannten Formats (z.B. 1.4 für PDF).	ja: Format Version
<jhove-wellformed>	Zahl der laut JHOVE im Sinne der Formatbeschreibung wohlgeformten Dateien.	
<jhove-valid>	Zahl der laut JHOVE im Sinne der Formatbeschreibung validen Dateien.	
<jhove-mime>	Internet Media Type gemäß Standard Multipurpose Internet Mail Extensions (MIME) nach JHOVE.	ja: MIME-Type
<jhove-compression>	Verwendetes Kompressionsverfahrens laut JHOVE.	
<jhove-image-length>	Länge (Pixelzahl) eines Rasterbildes.	
<jhove-image-width>	Breite (Pixelzahl) eines Rasterbildes.	
<jhove-color>	Farbraum laut JHOVE (z.B. RGB).	

XML-Name in IngestList	Beschreibung	Automatisiert übernommen in strukturierte Metadaten, s.o. 2.
<jhove-bitspersample>	Zahl der für einen Punkt in einem Rasterbilde verwendeten Bits laut JHOVE.	
<jhove-pdfa>	Erkennung als PDF/A laut JHOVE.	
<jhove-cnt-pages>	Anzahl der Seiten eines PDFs laut JHOVE.	
<jhove-cnt-images>	Anzahl der Bilder laut JHOVE.	
<jhove-cnt-chars>	Anzahl der Zeichen bei Textdateien laut JHOVE.	
<jhove-line-endings>	Verwendete Zeilentrenner laut JHOVE.	

5 Hashwerte

Zu allen durch IngestList übernommenen Primärdatendateien wird der Hashwert nach dem MD5-Verfahren erhoben und in der Beschreibungsdatei abgelegt. Außerdem wird in IngestList von der Beschreibungsdatei und in DIMAG von allen abgelegten Meta- und Primärdatendateien der Hashwert erhoben und in einer zweiten Datei abgelegt.