

KILIAN AMRHEIN¹, TIM HASLER², MARCO KLINDT¹,
ELIAS OLTMANNS², WOLFGANG PETERS-KOTTIG²

Digitale Langzeitarchivierung: Mustervorlage für Übernahmevereinbarungen, Version 2

¹Servicestelle Digitalisierung, Zuse Institute; amrhein@zib.de, klindt@zib.de

²Kooperativer Bibliotheksverbund Berlin-Brandenburg, Zuse Institute; hasler@zib.de,
oltmanns@zib.de, peters-kottig@zib.de

Zuse Institute Berlin
Takustr. 7
14195 Berlin
Germany

Telephone: +49 30-84185-0
Telefax: +49 30-84185-125

E-mail: bibliothek@zib.de
URL: <http://www.zib.de>

ZIB-Report (Print) ISSN 1438-0064
ZIB-Report (Internet) ISSN 2192-7782

Digitale Langzeitarchivierung: Mustervorlage für Übernahmevereinbarungen, Version 2



Kilian Amrhein¹, Tim Hasler², Marco Klindt¹, Elias Oltmanns², Wolfgang Peters-Kottig²

¹*Servicestelle Digitalisierung, Zuse Institute Berlin; amrhein@zib.de, klindt@zib.de*

²*Kooperativer Bibliotheksverbund Berlin-Brandenburg, Zuse Institute Berlin; hasler@zib.de, oltmanns@zib.de, peters-kottig@zib.de*

Die Servicestelle Digitalisierung (digiS) und der Kooperative Bibliotheksverbund Berlin-Brandenburg (KOBV) am Zuse Institute Berlin (ZIB) ermöglichen ihren Kooperationspartnern digitale Langzeitarchivierung gemäß den Vorgaben des Referenzmodells „Open Archival Information System“ (OAIS). Wesentliches Merkmal des Service für Kultur- und Wissenschaftseinrichtungen ist neben einer langfristigen Speicherung auf der Infrastruktur des ZIB vor allem die Bewahrung der Lesbarkeit oder genauer der Verstehbarkeit der archivierten Inhalte. Denn ein perfekt archivierter Datenstrom schützt nicht vor dem Veralten von Datenformaten. Um eine spätere Konvertierung obsoleter Formate überhaupt zu ermöglichen, werden im Zuge jeder Datenübernahme ins Langzeitarchiv zunächst umfangreiche technische, deskriptive und administrative Metadaten über die Informationsobjekte gesammelt. Dies umfasst auch Informationen über den rechtlichen Status der Daten, Identifikatoren, entscheidungsberechtigte Ansprechpartner und sonstige Kontextinformationen zu den Inhalten. Die Bedingungen für diese zusätzlichen organisatorischen und technischen Bearbeitungsschritte werden vor dem Start einer Datenübernahme mit dem Datengeber (Kooperationspartner) besprochen und in einer Übernahmevereinbarung schriftlich festgehalten. Insbesondere Absprachen über Struktur und Inhalte von Datenlieferungen werden auf diese Weise für alle Beteiligten transparent dokumentiert. Die Übernahmevereinbarung ergänzt die Kooperationsvereinbarung zwischen ZIB und institutionellem Partner, denn erfahrungsgemäß sind Details, die dort beschrieben sind, zum Zeitpunkt des eigentlichen Vertragsschlusses zumeist noch nicht bekannt. Für jeden Datengeber (im Einzelfall für jede Lieferung) wird der Inhalt der Übernahmevereinbarung vor dem Start des eigentlichen Übernahmeprozesses ins Langzeitarchiv individuell festgelegt.

Die nachfolgend dokumentierte Übernahmevereinbarung von KOBV und digiS ist in Anlehnung an Gestaltungsvorschläge von Übernahmevereinbarungen nach DIN 31645:2011-11 »Information und Dokumentation – Leitfaden zur Informationsübernahme in digitale Langzeitarchive« entwickelt und an die eigene Praxis im ZIB angepasst worden. Der Text ist als Mustervorlage konzipiert, die von anderen digitalen Langzeitarchiven an die Bedingungen der eigenen Einrichtung angepasst werden kann. Inhalte sind neben einer Auflistung allgemeiner Eigenschaften der Datenübernahme und der zu archivierenden Informationsobjekte konkrete Vorgaben zu notwendigen Metadaten einer Datenlieferung, Transferpaketformaten, Transferoptionen sowie die Beschreibung eines Transferprozesses. Im Anhang der Übernahmevereinbarung werden anhand von Beispielen die Optionen zur Strukturierung von Transferpaketen vorgestellt.

In der vorliegenden Version 2 der Übernahmevereinbarung wurde das Submission Manifest in aktualisierter Version 1.3 eingefügt, es erfolgten sprachliche Präzisierungen, und die Beispiele möglicher Verzeichnisstrukturen für Transferpakete wurden erneuert.

Übernahmevereinbarung

zwischen [INSTITUTION X, *Anschrift*], nachstehend Langzeitarchiv und
[INSTITUTION Y, *Anschrift*], nachstehend Datengeber,
gemeinsam Partner, genannt.

Das Dokument beschreibt den Transfer von digitalen Daten des Datengebers [INSTITUTION Y] in das digitale Langzeitarchiv [der INSTITUTION X] in Bezug auf den zwischen beiden Einrichtungen geschlossenen Vertrag (*Vertragsnummer, Datum*).

Mit dieser Übernahmevereinbarung erfolgt die Festlegung der organisatorischen Prozesse und technischen Details im Zuge des Ingest. Es gelten folgende Bedingungen für die Datenübernahme:

- Zusammengehörige Daten werden als strukturierte Datenpakete (digitale Objekte) geliefert. Diese Objekte werden im Langzeitarchiv als Einheiten verwaltet.
- Die Repräsentationen der Informationsobjekte werden soweit möglich nur in archivfähigen Formaten geliefert. Das Langzeitarchiv gibt im Vorfeld der Datenübernahme Empfehlungen für archivfähige Formate.
- Eine Transformation der Dateien durch Konvertierung (Migration) ist nicht im Zuge der Datenübernahme vorgesehen – sie erfolgt ggf. beim Datengeber.
- Die zu transferierenden Daten werden nur bis zum Umfang der vertraglich vereinbarten Datenmenge angenommen
- Alle Datei- und Pfadnamen sollen nach dem gleichen Schema aufgebaut sein und keine Leer- und Sonderzeichen enthalten. Ferner wird eine einheitliche Ordnerstruktur angestrebt.
- Metadaten werden soweit möglich als valide Metadatenformate geliefert. Die signifikanten Eigenschaften der Daten sollen gemäß den Erwartungen (zukünftiger) Nutzer der Daten berücksichtigt sein.
- Es werden ausschließlich Metadaten in das Langzeitarchiv übernommen, denen mindestens eine digitale Repräsentation eindeutig zugeordnet werden kann. Umgekehrt werden keine digitalen Repräsentationen übernommen, zu denen keine sie beschreibenden Metadaten vorliegen.
- Es werden bei einer Lieferung ausschließlich Transferpakete in das Langzeitarchiv übernommen. Eventuell zusätzlich übermittelte Dateien werden ignoriert.

1. Zu archivierende Informationsobjekte

Die Übernahmevereinbarung betrifft den Transfer folgender Daten: (*hier kurze inhaltliche Beschreibung der Lieferung(en) inkl. geschätzter Anzahl, Formate, geschätzter Gesamt-Datenmenge*). Informationen über die technische Umgebung und rechtliche Beziehungen der Partner sind nicht in diesem Dokument sondern im Vertrag festgelegt.

2. Notwendige Metadaten der Datenlieferung

Die Metadaten, die eine gesamte Lieferung beschreiben, werden in einem **Submission Manifest** zusammengestellt. Dieses Submission Manifest dient vorrangig der Identifizierung der gelieferten Daten und der Ansprechpartner beim Datengeber im Zuge des Lieferprozesses. Gemäß der nachfolgenden Aufstellung wird es je nach Lieferformat entweder innerhalb von METS (submission-manifest.xml) oder als separate Textdatei (submission-manifest.txt) bereitgestellt (vgl. Abschnitt 3 und Anhang). Das Submission Manifest ersetzt nicht die individuellen Beschreibungen der Objekte in den Metadaten des Transferpakets. Die Textdatei muss UTF8-codiert sein. Sie wird in enger Abstimmung zwischen den Partnern mit folgenden Angaben erstellt:

SubmissionManifestVersion: 1.3

SubmittingOrganization: Name der abliefernden Institution (Datengeber)

OrganizationIdentifier: z.B. ISIL-Nummer der abliefernden Institution (oder ähnliches)

ContractNumber: Vertragsnummer (Verweis auf zugrundeliegende vertragliche Regelung zw. Datengeber und Langzeitarchiv)

Contact: Name der verantwortlichen administrativen Person in der abliefernden Institution, die bei Fragen kontaktiert wird. Zudem Zieladresse des Übernahmeprotokolls nach Abschluss des Ingestprozesses (Format: Nachname, Vorname, *optional:* Titel, *optional:* Funktion in der abliefernden Institution)

ContactEmail: E-Mail-Adresse von Contact.

TransferCurator: Name der Person, die die Daten bereitstellt (wenn abweichend von *Contact*). Wird bei Rückfragen bzgl. des Transfers kontaktiert. Erhält Kopie des Übernahmeprotokolls. (Format: Nachname, Vorname, *optional:* Titel)

TransferCuratorEmail: E-Mail-Adresse von *TransferCurator*

SubmissionName: Eindeutiger, vom Datengeber vergebener Identifikator, der die Lieferung kennzeichnet. Erlaubte Zeichen: A-Za-z0-9_()#- (Beispiel: L_x42-2020)

SubmissionDescription: Selbstgewählte inhaltliche Beschreibung der Lieferung. Beispiel: „Digitalisierungsprojekt Korrespondenzen 1995, Teil 1 bis 5. Siehe auch Artikel zum Projekt von Mustermann (1995): XYZ...“

AccessRights: Angaben zu Nutzungs- bzw. Zugriffsrechten, die für alle Objekte gelten. Enthält ausschließlich einen der folgenden drei Begriffe:

- *institution*
(nur der Datengeber darf auf die archivierten Objekte zugreifen)
- *public*
(die archivierten Objekte können potentiell zugänglich gemacht werden. Es gelten die in den deskriptiven Metadaten der Transferpakete beschriebenen Lizenzen. Wenn dort nichts verzeichnet ist, gilt die im submission-manifest unter *License* aufgeführte Lizenz. [Erfordert einen Eintrag im Feld *License*])
- *embargoUntil YYYY-MM-DD*
(bis zum angegebenen Datum darf nur der Datengeber auf die archivierten Objekte zugreifen, danach gelten im Fall einer Veröffentlichung durch das Langzeitarchiv die in den deskriptiven Metadaten der Transferpakete beschriebenen Lizenzen. Wenn dort nichts verzeichnet ist, gilt die im submission-manifest unter *License* aufgeführte Lizenz. (Erfordert einen Eintrag im Feld *License*)

License: Nutzungsbedingungen für die potentielle Nachnutzung der Objekte, zu denen in den gelieferten Metadaten *keine* Lizenzangaben vorhanden sind. Die hier angegebene Lizenz dient als „Fallback-Information“ für das Langzeitarchiv falls keine sonstigen Lizenzangaben vorhanden sind. Es soll ausschließlich ein URI angegeben werden. Beispiel: Creative Commons Lizenzen.

RightsDescription: Enthält zusätzliche inhaltliche Erläuterungen für das Langzeitarchiv bezüglich der urheberrechtlichen, datenschutzrechtlichen oder persönlichkeitsrechtlichen Beschränkungen, insbesondere wenn individuelle Absprachen oder vertragliche Regelungen mit dem oder der RechteinhaberIn erfolgt sind. (Mögliche Statements unter rightsstatements.org). Ermöglicht die Erläuterung und Konkretisierung der unter *License* angegebenen Lizenzinformationen. Beispiel: „*Nach einer mündlichen Vereinbarung mit dem ursprünglichen Datengeber xyz von ca. 2005 dürfen die Digitalia veröffentlicht, aber nicht nachgenutzt werden, obwohl dies urheberrechtlich möglich wäre. Aus diesem Grund kann keine freie Lizenz vergeben werden, für die Werke gilt die unter License angegebene Rechteausszeichnung* <http://www.europeana.eu/portal/rights/rr-f>.“

DataSourceSystem: Software, aus der Daten exportiert wurden (mit Versionsangabe)

MetadataFile: Name und Pfad zur Metadaten-datei im Transferpaket. Wenn es mehrere Metadaten-dateien mit unterschiedlichen Dateinamen pro Lieferung gibt, werden entsprechend Sternchen (*) als Platzhalter verwendet. Sind diese weiterhin in unterschiedlichen Verzeichnissen enthalten, werden auch für diese Verzeichnisnamen Platzhalter verwendet. Die Verzeichnistiefe wird hiermit ebenfalls bestimmt und ist pro Lieferung eindeutig. Beispiele: meta.xml oder */meta.xml (für gleichlautende Metadaten-dateien in unterschiedlichen Ordnern) bzw. */*.xml (für Dateien mit verschiedenen Dateinamen in unterschiedlichen Ordnern). Im zweiten Fall darf nur eine Datei mit dieser Endung im Verzeichnis enthalten sein.

MetadataFileFormat: URI des Metadaten-schemas (zum Verständnis von Struktur/Format der Metadaten wie dcterms, datacite, lido, ead, mods). Beispiele: <http://www.lido-schema.org> oder <http://www.loc.gov/standards/mods/v3/mods-3-6.xsd>

3. Beschreibung des Transferpaketformats

Das Langzeitarchiv nimmt ausschließlich Daten an, die mit Integritätsinformationen versehen sind. Diese werden idealerweise im Zuge der Erstellung eines Transferpaketes automatisch erzeugt. Transferpakete können grundsätzlich

- a) in Form eines mit **METS** beschriebenen Verzeichnisses
- b) in Form eines **bag**¹ oder
- c) als **komprimierter Container** (z.B.: ZIP) übermittelt werden.

In Bezug auf die Verzeichnisstruktur hat der Datengeber verschiedene Möglichkeiten (s. Anhang). Das Langzeitarchiv unterstützt den Datengeber bei der Erstellung eines Transferpaketes. Grundsätzlich wird die Übermittlung der strukturierenden Metadaten im METS-Format (Metadata Encoding and Transmission Standard) empfohlen. Die Zuordnung zu Intellektuellen Einheiten erfolgt hierbei im METS und es kann in diesem Fall auf andere Strukturierungen, wie in den anderen Beispielen gezeigt, vollständig verzichtet werden.

Ein Transferpaket beinhaltet in der obersten Verzeichnisebene der enthaltenen Informationsobjekte die Daten des Submission Manifest (wie oben gezeigt) mit den Liefermetadaten. Das Paket wird identifiziert durch den in dieser Datei enthaltenen *SubmissionName*. Dieser muss eindeutig sein in Bezug auf alle Lieferungen des Datengebers. Ein Transferpaket soll ausschließlich die zur Interpretation der Informationsobjekte notwendigen Daten und keine Referenzen zu externen Daten enthalten. Hierbei gelten Daten als extern, wenn sie nur für den Datengeber verfügbar sind. Referenzen auf persistent veröffentlichte Daten dürfen im Datenpaket enthalten sein.

Ein einzelnes Transferpaket besteht aus maximal 1,8 TB (Terabyte)² Daten. Der Datengeber soll in einem Transferpaket eine möglichst große Datenmenge zusammenzufassen, um die Zahl der Transferpakete gering zu halten.

4. Beschreibung der Transferoptionen

Zum Transfer der Daten in das Langzeitarchiv stehen zwei Optionen zur Verfügung:

- Internet: Der Transfer des vom Datengeber erstellten Transferpaketes erfolgt mittels sftp (secure file transfer protocol) oder scp (secure copy) in einen Transferbereich des Langzeitarchivs oder er wird vom Langzeitarchiv per scp aus einem vom Datengeber verwalteten Speicherbereich heruntergeladen.
- Externe Festplatten: Die Transferpakete werden auf externen Festplatten geliefert und die Daten von MitarbeiterInnen des Langzeitarchivs manuell in den Transferbereich übertragen. Integritätsinformationen werden vor der Übertragung durch den Datengeber erzeugt.

5. Beschreibung des Transferprozesses

Im Transferbereich des Langzeitarchivs werden die übermittelten Daten anhand der beim Datengeber erzeugten Prüfsummen auf Übertragungsfehler untersucht. Im Fehlerfall werden alle Dateien verworfen und eine erneute Übertragung initiiert. Bei fehlerfreier Übertragung und Validität von Format und Verzeichnisstruktur der Transferpakete werden die Daten in das Langzeitarchiv übertragen und anschließend die Transferkopie gelöscht. Transferpakete werden nur als Ganzes übernommen. Aus den Transferpaketen werden im Langzeitarchiv nach definierten Verfahren Submission Information Packages (SIP), Dissemination Information Packages (DIP) und Archival Information Packages (AIP) erzeugt. Nachdem die AIP in den Speicherbereich des Langzeitarchivs übertragen wurden, ist der Transfer technisch abgeschlossen. Das Langzeitarchiv erstellt abschließend ein Übernahmeprotokoll, das die übernommenen Daten und ihre Prüfwerte auflistet. Nach Abschluss der Datenübernahme in das Langzeitarchiv wird das Übernahmeprotokoll dem Datengeber bereitgestellt. Der Transferprozess ist nach Bereitstellung des Übernahmeprotokolls auch organisatorisch abgeschlossen. Die Verantwortung über die Bewahrung der archivierten Objekte liegt nun beim Langzeitarchiv.

¹ vgl. BagIt der Library of Congress, <http://tools.ietf.org/html/draft-kunze-bagit-09>

² 1 TB (Terabyte) = 10¹² Byte

ANHANG: Mögliche Verzeichnisstrukturen für Transferpakete

Bei der Erstellung eines Transferpakets hat der Datengeber die Möglichkeit, Daten unterschiedlich strukturiert zusammenzustellen. Im Folgenden werden die Alternativen für die Verzeichnisstruktur beschrieben. Sie stellen einen Kompromiss dar zwischen größtmöglicher Flexibilität für den Datengeber und Minimierung des Aufwands für das Langzeitarchiv.

In einem Transfer kann eine beliebige Anzahl von Objekten mit Metadaten enthalten sein, sofern das zuvor genannte maximale Datenvolumen eines Transferpaketes nicht überschritten wird. Die Objekte können aus einer oder mehreren Dateien (digitalen Repräsentationen, z.B. tiff-Dateien) bestehen und bilden zusammen mit den zugehörigen Metadaten jeweils eine Intellektuelle Einheit (IE). Eine IE besteht zum Beispiel aus den Digitalisaten eines Buches mit zugehörigen beschreibenden und strukturierenden Metadaten. Das Konzept der Intellektuellen Einheit dient dazu, dass zusammengehörige Inhalte auf unbestimmte Zeit verständlich beschrieben bleiben und identifizierbar, verwaltbar sowie nutzbar bleiben.

Jede IE für sich kann beim Transfer nahezu beliebig bezeichnet werden, jedoch darf diese Bezeichnung keine Leer- und Sonderzeichen enthalten. Innerhalb eines Transfers werden in der Regel mehrere IE enthalten sein, jede IE erhält dann einen eigenen Bezeichner, der keinem festen Schema folgen muss.

Nachfolgend sind einzelne *Dateien* nach diesem Prinzip bezeichnet: *****.*****
Die ersten drei Sterne symbolisieren dabei den Dateinamen, die letzten drei die Dateiendung. Die Dateinamen können dabei ebenfalls beliebig gestaltet sein (jedoch ohne Leer- und Sonderzeichen), die Dateiendung muss dem jeweiligen Dateiformat entsprechen, z.B. ab45234.tif, 20348.tif, m29384.xml. Wenn in einer Variante ein bestimmtes Dateiformat verlangt wird, ist die Dateiendung entsprechend dargestellt, z.B. *****.xml**. Schrägstriche (/) kennzeichnen jeweils Unterverzeichnisse. Datei- und Ordnernamen sollten generell nicht zu lang sein.

Es ist möglich, im **optionalen Verzeichnis „submissionDocumentation“** *zusätzliche* inhaltliche Informationen zu den im jeweiligen Transferpaket enthaltenen Intellektuellen Einheiten abzulegen, die bei einer (zukünftigen) Nutzung helfen können, die Objekte zu interpretieren. Diese werden ausdrücklich *nicht* die beschreibenden Metadaten der IE ersetzen; sie ergänzen die IE lediglich um solche Informationen, die in die eigentlichen beschreibenden Metadaten nicht aufgenommen werden können. Diese Daten können auch unstrukturiert vorliegen. *Beispiele*: Bilder vom Experimentaufbau; Skizze vom Fotosetup; Emails mit Informationen, die auf die Herkunft der Daten/Objekte schließen lässt.

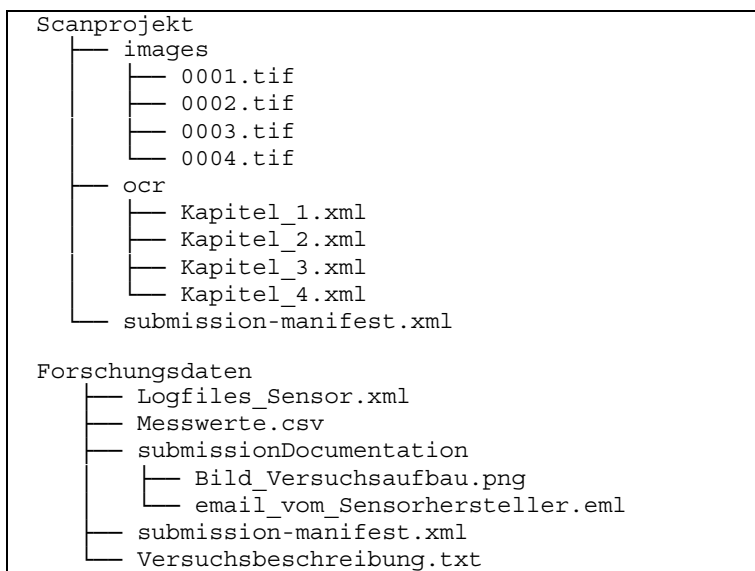
An den Inhalten des Verzeichnisses submissionDocumentation werden im Langzeitarchiv keine Erhaltungsmaßnahmen durchgeführt.

Der Datengeber kann sich bezüglich der Verzeichnisstruktur für eine der folgenden Varianten entscheiden, wobei das Langzeitarchiv ihn bei der Auswahl bei Bedarf beratend unterstützt:

1. Struktur im METS ordnet jede Datei einer Intellektuellen Einheit zu

Datencontainer (Bilder, Filme etc.) und Metadatencontainer werden entsprechend ihrer IE in einer vom Langzeitarchiv vorgegebenen METS-Struktur verzeichnet. Diese METS-Datei wird als submission-manifest.xml in der obersten Dateiebene mitgeliefert. Der Speicherort der jeweiligen Dateien ergibt sich relativ zu dieser METS-Struktur. Die Erzeugung von METS-Dateien erfolgt in enger Absprache zwischen den Partnern.

Beispiele:



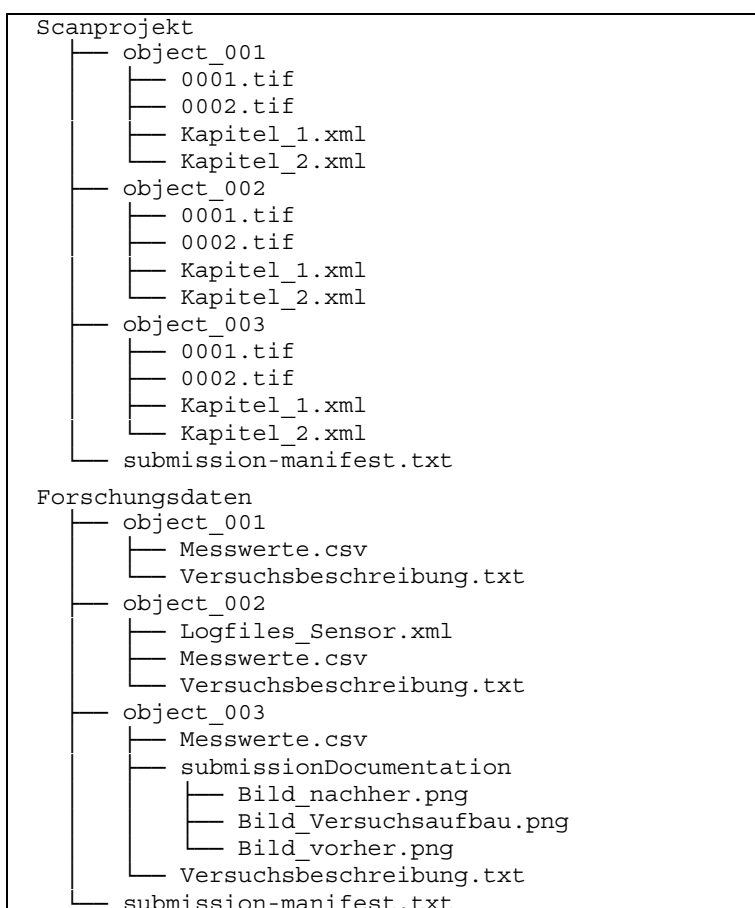
oder

2. Einzelne Unterverzeichnisse je IE

Die einzelnen Dateien einer IE werden in jeweils einem Unterverzeichnis zusammengefasst. Die Lieferung erfolgt als Container (ZIP) oder bag.

- /BezeichnerName/***.*** – Primärdaten (beliebige Anzahl an Dateien, mindestens eine)
- /BezeichnerName/***.xml – Datei mit Metadaten (insbesondere wenn auch die Primärdaten in XML abgelegt werden, muss diese Datei im Feld *MetadataFile* in der Datei *submission-manifest.txt* genau spezifiziert werden)
- /BezeichnerName/submissionDocumentation/ (optional)
- /submission-manifest.txt

Beispiele:

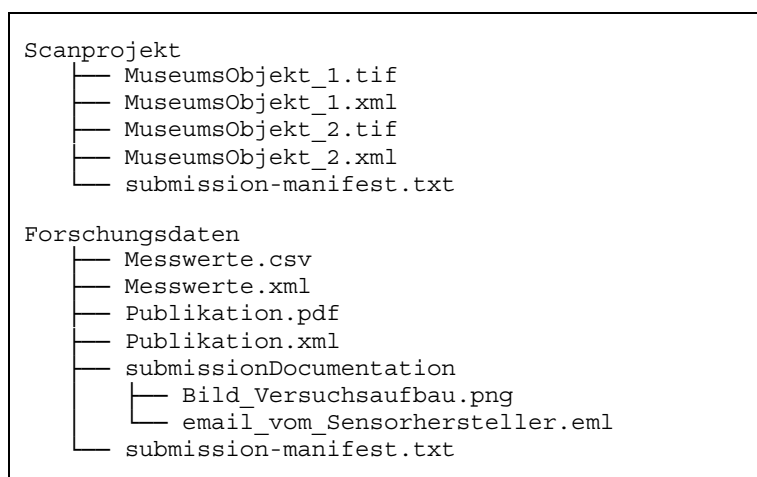


oder

3. Alle Dateien der verschiedenen IE in einem Verzeichnis

- Primärdateien (Bilder, Filme etc.) und Metadateiendateien mit gleicher Benennung (Beispiel lv1102a.tif, dann lv1102a.xml usw.) in einem Hauptverzeichnis. Dies ist nur möglich bei Intellektuellen Einheiten, die aus einer Digitalisat- und einer Metadatei mit jeweils unterschiedlicher Dateiendung bestehen, also z.B. nicht für XML-Primärdaten, wenn die Metadaten ebenfalls in einem XML-Format vorliegen. Die Lieferung erfolgt als Container (ZIP) oder bag.
- /submissionDocumentation/ (optional) – als Unterverzeichnis des Hauptverzeichnisses, gilt nur für die gesamte Datenlieferung.
- /submission-manifest.txt – im Hauptverzeichnis

Beispiele:



oder

4. Einzelne ZIP-Container für einzelne IE

Die einzelnen Dateien einer IE werden in jeweils einer ZIP-Datei zusammengefasst. Die Lieferung erfolgt als Container (ZIP) oder bag.

- BezeichnerName.zip – Einzelne Zip-Container mit Daten und Metadaten als Einheiten
- /submission-manifest.txt

Beispiele:

