

Übernahmevereinbarung

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

Das Dokument beschreibt den Transfer von digitalen Daten des Datenproduzenten [INSTITUTION Y] in das digitale Langzeitarchiv [der INSTITUTION X] in Bezug auf den zwischen beiden Einrichtungen geschlossenen Vertrag [Datum, Vertragsnummer]. Mit dieser Übernahmevereinbarung erfolgt die Festlegung der organisatorischen Prozesse und technischen Details im Zuge des Ingest.

1. Zu archivierende Informationsobjekte

Die Übernahmevereinbarung betrifft den Transfer folgender Daten: [...] (kurze inhaltliche Beschreibung der Lieferung(en) inkl. geschätzter Anzahl, Formate, geschätzter Gesamt-Datenmenge)

Informationen über die technische Umgebung und rechtliche Beziehungen sind im Vertrag festgelegt.

Es gelten folgende Bedingungen für die Datenübernahme:

- Die digitalen Objekte werden soweit möglich nur in archivfähigen Formaten geliefert. [INSTITUTION X] gibt im Vorfeld der Datenübernahme Empfehlungen für archivfähige Formate.
- Eine Normalisierung der Dateien durch Konvertierung (Migration) ist im Zuge der Datenübernahme nicht vorgesehen – sie erfolgt ggf. beim Datenproduzenten.
- Die zu transferierenden Daten werden nur bis zum Umfang der vertraglich vereinbarten Datenmenge angenommen. Alle Daten werden in Abhängigkeit von der Datenmenge in einem oder mehreren Transferpaketen (vgl. Punkt 3) geliefert.
- 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 der Zielgruppe (Nutzer der Daten) berücksichtigt sein.
- Es werden ausschließlich Metadaten in das Langzeitarchiv übernommen, denen mindestens eine digitale Repräsentation eindeutig zugeordnet werden kann.
- Es werden bei einer Lieferung ausschließlich Transferpakete in das Langzeitarchiv übernommen. Eventuell zusätzlich übermittelte Dateien werden ignoriert.

2. Notwendige Metadaten der Datenlieferung

Die Metadaten, die eine Lieferung beschreiben, werden gemäß der nachfolgenden Aufstellung in einer Datei „submission-manifest.txt“ bereitgestellt. Diese Datei dient lediglich der Identifizierung der gelieferten Daten und der Ansprechpartner beim Datenproduzenten im Zuge des Lieferprozesses. Sie ersetzt nicht die notwendigen beschreibungen in den Metadaten des Transferpakets. Die Textdatei wird in Absprache zwischen Datenproduzenten und digitalem Langzeitarchiv mit folgenden Angaben erstellt:

SubmissionManifestVersion: 1.0
SubmittingOrganization: Name der abliefernden Institution
OrganizationIdentifier: z.B. ISIL-Nummer der abliefernden Institution (oder ähnliches, sofern vorhanden)
ContractNumber: Vertragsnummer (Verweis auf zugrundeliegende vertragliche Regelung zw. Abliefernder Institution und Dienstleister)
Contact: Name des administrativen Ansprechpartners in der abliefernden Institution, der bei Fragen oder Zugriffsanfragen kontaktiert wird
ContactEmail: Adresse von *Contact*, an die ein Übernahmeprotokoll geschickt werden soll
SubmissionIdentifier: Eindeutiger, von der abliefernden Institution vergebener Identifikator, der die Lieferung kennzeichnet. Erlaubte Zeichen: A-Za-z0-9_ - () @# . Beispiel: L_x42-2013
SubmissionDescription: Selbstgewählte inhaltliche Beschreibung der Lieferung
Rights: urheberrechtliche, datenschutzrechtliche, persönlichkeitsrechtliche Angaben und Embargozeiten; z.B. „public domain“, „freier Zugriff – Rechte vorbehalten“ oder konkrete Lizenzangaben; Namen berechtigter Personen
DataSourceSystem: Software, aus der Daten exportiert wurden (mit Versionsangabe)
MetadataFile: Name und Pfad zur Metadaten-datei (nach Klärung der Verzeichnisstruktur des Transferpakets)
MetadataSchema: Struktur/Format der Metadaten, z.B. dc, dcterms, datacite, lido, ead, mods
DataCurator: Name des Ansprechpartners, der bei Fragen zu dieser Lieferung kontaktiert/genannt wird (wenn abweichend von *Contact*)
DataCuratorEmail: Adresse *DataCurator*
DataEncoding: verwendete Dateiformate, z.B. TIFF

3. Beschreibung des Transferpaketformats

[*INSTITUTION X*] 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 **bag**¹ oder b) in Form eines mit **METS** beschriebenen Verzeichnisses oder c) als **ZIP**-Container übermittelt werden. In Bezug auf die Verzeichnisstruktur hat der Datenproduzent verschiedene Möglichkeiten (s. Anhang). [*INSTITUTION X*] unterstützt bei Bedarf den Datenproduzenten bei der Erstellung eines Transferpaketes.

Ein Transferpaket beinhaltet in der obersten Verzeichnisebene der enthaltenen Informationsobjekte eine Datei „submission-manifest.txt“ (wie oben gezeigt) mit den Liefermetadaten. Das Paket wird identifiziert durch den in dieser Datei enthaltenen *SubmissionIdentifier*. Dieser muss eindeutig sein in Bezug auf alle Lieferungen des Datenproduzenten. Ein Transferpaket soll ausschließlich die zur

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

Interpretation der Informationsobjekte notwendigen Daten und keine Referenzen zu externen Daten enthalten. Hierbei gelten Daten als extern, wenn sie nur für den Datenproduzenten verfügbar sind. Referenzen auf veröffentlichte Daten dürfen im Datenpaket enthalten sein.

Ein Transferpaket besteht aus maximal 1,5 TB (Terabyte)² Daten. Der Datenproduzent verpflichtet sich, in einem Transferpaket eine möglichst große Datenmenge zusammenzufassen, um die Zahl der Transferpakete gering zu halten.

4. Beschreibung des Transfers und seiner technischen Implementierung

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

- Internet:

Der Transfer des vom Datenproduzenten erstellten Transferpaketes erfolgt mittels scp (secure copy) oder über WebDAV per HTTPS in einen Transferbereich [der *INSTITUTION X*] oder er wird per scp aus einem vom Produzenten verwalteten Speicherbereich von [der *INSTITUTION X*] heruntergeladen.

- Externe Festplatten:

Die Transferpakete werden auf externen Festplatten geliefert und die Daten von Mitarbeitern [der *INSTITUTION X*] manuell in den Transferbereich übertragen. Integritätsinformationen werden vor der Übertragung durch den Datenproduzenten erzeugt.

5. Beschreibung des Validierungsprozesses

Im Transferbereich [der *INSTITUTION X*] werden die übermittelten Daten anhand der vom Packprogramm erzeugten Prüfsumme auf Übertragungsfehler untersucht. Im Fehlerfall werden alle Dateien verworfen und eine erneute Übertragung initiiert. Bei fehlerfreier Übertragung werden die Daten in das Bandspeichersystem des digitalen Langzeitarchivs übertragen und anschließend die Transferkopie gelöscht. Transferpakete werden nur als Ganzes übernommen.

[*INSTITUTION X*] erstellt ein Übernahmeprotokoll, das die übernommenen Daten und ihre Prüfwerte auflistet. Die Korrektheit des Übernahmeprotokolls ist von beiden Parteien zu bestätigen.

6. Zeitplan der Datenlieferung

Der Produzent liefert die Daten möglichst zeitnah nach Abschluss des Vertrages.

Der Transferprozess ist nach Bestätigung des Übernahmeprotokolls beendet.

² 1 TB (Terabyte) = 10¹² Byte

ANHANG: Mögliche Verzeichnisstrukturen für Transferpakete

Bei der Erstellung eines Transferpaketes hat der Datenproduzent die Möglichkeit, Daten unterschiedlich strukturiert zusammenzustellen. Im Folgenden werden die Alternativen für die Verzeichnisstruktur beschreiben. Sie stellen einen Kompromiss dar zwischen größtmöglicher Flexibilität für den Datenproduzenten und Minimierung des Aufwands für das digitale 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- oder pdf-Dateien) bestehen und bilden zusammen mit den zugehörigen Metadaten jeweils eine Intellektuelle Einheit (IE). Das Konzept der Intellektuellen Einheit dient dazu, dass zusammengehörige Inhalte auf unbestimmte Zeit verständlich beschrieben werden, identifizierbar, verwaltbar und nutzbar sind.

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 Ordernamen sollten generell nicht zu lang sein.

Im optionalen Verzeichnis „submissionDocumentation“ können *zusätzliche* inhaltliche Informationen zu den im jeweiligen Transferpaket enthaltenen Intellektuellen Einheiten liegen, die (zukünftigen) Datennutzern helfen können, die Objekte zu interpretieren. Diese können 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.

Der Datenproduzent kann sich für eine der folgenden Varianten entscheiden, wobei [INSTITUTION X] bei der Auswahl bei Bedarf beratend unterstützt:

1. Alle Dateien der verschiedenen IE in einem Verzeichnis

- Primärdateien (Bilder, Filme etc.) und Metadateien mit gleicher Benennung (Beispiel Datum 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 nicht für XML-Primärdaten.
- /submissionDocumentation/ (optional) – als Unterverzeichnis des Hauptverzeichnisses, gilt nur für die gesamte Datenlieferung.
/submission-manifest.txt – im Hauptverzeichnis

Beispiel:

```
transfer_1/
|-- Iv1102a.tif
|-- Iv1102a.xml
|-- AIG10_DTS_log.pdf
|-- AIG10_DTS_log.xml
|-- AIG10_DTS_log_r.png
|-- AIG10_DTS_log_r.xml
|-- BrocktonOval.jp2
|-- BrocktonOval.xml
|-- QE_Park_sunset.jp2
|-- QE_Park_sunset.xml
|-- submissionDocumentation
    |-- setup.png
    |-- handoverDocu.txt
|-- submission-manifest.txt
```

oder

2. Einzelne Unterverzeichnisse je IE

Die einzelnen Dateien einer IE werden in jeweils einem Unterverzeichnis zusammengefasst.

- /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

Beispiel:

```
transfer_2/
|-- object_001
    |-- AIG10_DTS_log.pdf
    |-- AIG10_DTS_log_r.png
    |-- AIG10_P_and_S-wave.pdf
    |-- DataCite.xml
    |-- dif.xml
    |-- escidoc.xml
|-- object_002
    |-- DIF.xml
    |-- DataCite.xml
    |-- data.csv
|-- object_003
    |-- submissionDocumentation
        |-- setup.png
        |-- handoverDocu.txt
    |-- DIF.xml
    |-- DataCite.xml
    |-- data.csv
|-- submission-manifest.txt
```

oder

3. Einzelne ZIP-Container für einzelne IE

Die einzelnen Dateien einer IE werden in jeweils einer ZIP-Datei zusammengefasst.

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

Beispiel:

```
transfer_3/
|-- object_001.zip
|-- object_002.zip
|-- object_003.zip
|-- submission-manifest.txt
```

oder

4. Metadaten in CSV

Die Metadaten werden in einer CSV-Datei, deren Format bei [INSTITUTION X] zu erfragen ist, abgespeichert.

- /BezeichnerName/objects/***.*** – Primäre Objektdaten
- /BezeichnerName/submissionDocumentation/ (optional)
- /BezeichnerName/metadata/metadata.csv – Format beim ZIB erfragen. Diese Datei muss im Feld *MetadataFile* in der Datei *submission-manifest.txt* entsprechend spezifiziert werden; ferner gilt : *MetadataFileFormat: csv*
/submission-manifest.txt

Beispiel:

```
transfer_4/
|-- object_001
|   |-- metadata
|   |   |-- metadata.csv
|   |-- objects
|   |   |-- AIG10_DTS_log.pdf
|   |   |-- AIG10_DTS_log_r.png
|   |   |-- AIG10_P_and_S-wave.pdf
|   |   |-- P_und_S-wave_r.png
|   |   |-- dif.xml
|   |-- submissionDocumentation
|   |   |-- AIG10_Fig_4.png
|   |   |-- DataCite.xml
|   |   |-- escidoc.xml
|   |   |-- logging_truck.png
|-- object_002
|   |-- metadata
|   |   |-- metadata.csv
|   |-- objects
|   |   |-- BrocktonOval.jp2
|   |   |-- QE_Park_sunset.jp2
|   |-- submissionDocumentation
|   |   |-- DataCite.xml
|-- object_003
|   |-- metadata
|   |   |-- metadata.csv
|   |-- objects
|   |   |-- DIF.xml
|   |   |-- data.csv
|-- submission-manifest.txt
```

Beispiel-CSV für object_002:

filename	dc.title	dc.contributor	dc.date	notes	dcterms.isPartOf	dc.rights
objects/ BrocktonOval.jp2	Stanley Park in December	Don Langfield, photographer	1992/12/04	Originally part of series entitled "Winter in Vancouver"	Riley Studios collection	Copyright held by Riley Studios
objects/ QE_Park_sunset.jp2	Sunset in Queen Elizabeth Park	Don Langfield, photographer	1994/07/13		Riley Studios collection	Copyright held by Riley Studios

oder

5. Struktur in METS

- Datencontainer (Bilder, Filme etc.) und Metadatencontainer werden entsprechend ihrer IE in einer METS-Struktur verzeichnet. Diese METS-Datei muss im Feld *MetadataFile* in *submission-manifest.txt* genau spezifiziert werden. Der Speicherort der jeweiligen Dateien ergibt sich relativ zu dieser METS-Struktur. Die Erzeugung von METS-Dateien erfolgt in enger Absprache zwischen den Partnern.
- /submission-manifest.txt