Specifications for transfer packages and their transmission to the German National Library using hotfolders

Version 2.0
Last updated: 28 May 2021

Editorial: Editorial team online publications

<urn:nbn:de:101-2021022201>

Table of contents

0 Introduction

The German National Library (DNB) is Germany's central archival library. Its collection mandate encompasses all texts, images and sound recordings published in Germany or in the German language, translated from German or relating to Germany that have been issued since 1913. The Law Regarding the German National Library (DNBG) and the Legal Deposit Regulation (PflAV) are the foundations on which our collections are based; they also regulate our obligation to collect non-physical media works, i.e. online publications.

Our collection of non-physical media works encompasses all texts, images and sound recordings accessible in the internet. These include both online publications with their printed equivalents and web-only media works.

More specifically, we collect books (e-books), journals (e-journals), daily newspapers (e-papers), university publications, sheet music, audio books, digitised works and selected websites published in the internet.

We offer the following options for the legal deposit of online publications:

- automated processes for transmitting larger quantities of publications, and
- web forms for transmitting individual publications

When depositing larger quantities of online publications, you can choose between hotfolder and OAI-PMH (Open Archives Initiative Protocol for Metadata Harvesting) processes. You can use either process to submit monographs, university dissertations and theses, audio books, sheet music and periodicals.

This document describes the process of depositing these works using the hotfolder process.

0.1 Additional information

You will find additional information about depositing online publications at:

https://www.dnb.de/EN/netzpublikationen

This web page also provides access to additional documents that you can use when depositing publications.

If you have any general questions about depositing online publications, please contact the online publication service at np-info@dnb.de.

If you have any technical questions about depositing online publications, please contact the technical service at np-technik@dnb.de.

The productive deposit of online publications is preceded by a test phase in which all the necessary arrangements are made by the depositor and the DNB. These arrangements relate to the structure of the transfer packages, the content and technical properties (e.g. freedom from restrictions) of the electronic publications, and the use of bibliographic metadata.

1 Basic structure

In order to transfer online publications through the hotfolder interface, it is first necessary to create a so-called "transfer package". This transfer package must always contain the digital publication itself together with the publication's bibliographic metadata in a format accepted by the DNB.

The metadata formats accepted are gradually being added to; you will find an updated list at:

http://www.dnb.de/metadataop

Digital publications are accepted in PDF, EPUB, TIFF, JPEG, PS and MP3 formats. The possibility of depositing publications in other file formats can be reviewed during the test phase.

The hotfolder process can be used regardless of the type of publication and the format of the digital material, which means that it can be used for both monographic and serial publications. A separate transfer package must be created for each digital publication. If you wish to deposit 15 e-books or 15 articles from a magazine, for example, you will have to send us 15 transfer packages through this interface.

The hotfolder has a directory on one of the DNB's own servers which depositors can use as the destination for the transfer packages.

2 Transmission

The packages are deposited at regular intervals as agreed, e.g. daily, weekly or monthly. The transfer packages deposited since the last request are retrieved in each case. The metadata supplied is imported directly into the German National Library's catalogue, where it serves as bibliographic information. The digital publications are stored on the archive server and linked with the bibliographic metadata.

After registration, all depositors are provided with a dedicated account with their own access data (user name and password).

2.1 Protocols

Two different transfer protocols are offered for transmitting the transfer packages:

- SSH File Transfer Protocol (SFTP)
- Web-based Distributed Authoring and Versioning (WebDav)

2.2 Deposits made by service providers

If you decide to deposit your publications using the hotfolder process, we will set up a hotfolder account for you and send you the access data. The hotfolder account will contain at least one deposit account. It is possible to set up several deposit accounts, e.g. for different types of publication or imprints.

If the deposit is made by a service company, the corresponding deposit account will be in their name. Here too, it is possible to assign several deposit accounts to different service companies.

2.3 Information on use

Transfer packages are normally sent using an SFTP client such as FileZilla or WinSCP.

During the upload processes, the file extension used with the transfer package must be changed to "tmp", or this file extension must be added. Once the file upload is complete, the depositor must reset the file extension. Any hidden files created during the deposit process must be deleted. It is also advisable to transfer checksum files before transmitting the relevant transfer packages.

Please make sure that UTF-8 character encoding is activated in the client used to transfer the file. If possible, please avoid using special characters or umlauts in the name of the transfer package. In exceptional cases where the use of special characters or umlauts is unavoidable, it must be ensured that the name of the transfer package is UTF-8 encoded, otherwise the DNB will be unable to process it.

3 The transfer package

Each transfer package contains the digital publication and the bibliographic information from the relevant metadata record. Both elements are transferred in a current container format. At present, the DNB accepts the following container formats:

- ZIP
- TAR

You are free to choose the file name for the transfer package provided the rules in section 3.1. are adhered to; this means for example that you can use the digital publication's ISBN in the transfer package's name.

Examples of transfer package file names:

```
9783958750685.zip
9783958750685.tar
abfallr_2016_01-jats-001.zip
SAD497392.tar
```

At the upper level, each transfer package contains a folder with the name "content"; this is where the digital publication is located. This folder name *cannot* be chosen at will.

The obligatory metadata must be provided in an XML file named "catalogue_md.xml". This file name *cannot* be chosen at will.

Example of transfer package structure – overview

```
9783958750685.zip
    catalogue_md.xml
    content
```

Example of transfer package structure – complete with digital object

```
9783958750685.zip
    catalogue_md.xml
    content
        9783958750685.pdf
```

3.1 "content" folder

All files belonging to the digital publication are stored in the "content" folder.

The following rules must be observed:

- The following restrictions apply when creating file names:
  - no umlauts
  - no special characters
  - not more than 128 characters
  - UTF-8 encoded
- For technical reasons, the "content" folder may not contain more than 4,999 files.
- The permitted file formats are PDF, EPUB, TIFF, JPEG, PS, MP3.
- The "content" folder may contain a ZIP or TAR container.
- It is not possible to use more than one container file at the top level in the "content" folder. Additional files can be transmitted in a subfolder. The name of the subfolder can be chosen at will.

Example of content in "content" folder: digital publication in one file

```
content
    978-3-644-50121-8.epub
```

Example of content in "content" folder: Example with book cover and body of book in separate files

```
content
    2711.pdf
    2711.jpeg
```

Example of content in "content" folder (extract): audio book with individual track files (64 files in all)

```
content
    9783837143294-Track-001.mp3
    9783837143294-Track-002.mp3
    9783837143294-Track-003.mp3
    9783837143294-Track-063.mp3
    9783837143294-Track-064.mp3
```

Example of content in "content" folder: digital publication with supplementary material (in container format)

```
content
   2711.pdf
   2711.zip
```

The file formats in the folder are reviewed automatically when the online publication is submitted. The DNB expects the quality of the file format to comply with a minimum standard. Container formats are reviewed as such; our processes do not encompass the data they contain.

The quality of the file formats is discussed in detail during the test phase (cf. 0.1 additional information) These discussions relate in particular to the possibility of unrestricted access to the online publications; the use of these publications may not be impeded by digital rights management. Compliance is checked during each import process. Additional information is available at

urn:nbn:de:101-2012102408 (only in German) and https://www.dnb.de/hotfolderop.

3.2 File "catalogue_md.xml"

Online publications must be deposited together with their metadata. Metadata can be delivered using various metadata standards, chosen according to how widely they are used in the respective communities. Information about the standards currently supported and how they can be used to deliver metadata to the DNB depending on the type of publication is available at

http://www.dnb.de/metadataop

The format and scope of the metadata are discussed with the depositor during the test phase. The DNB's requirements are oriented on a core set that specifies which fields are obligatory regardless of the format.

4 Transmitting checksums

The DNB offers optional reciprocal checksum validation to guarantee the authenticity of the online publications delivered. This involves the calculation of a checksum for each transfer package and – if applicable – for each file it contains; these can then be used to identify unwanted changes. The checksum process can be used just for the transfer package itself or for each of the files it contains.

The DNB currently uses the following hash processes for checksum validation:

- MD5
- SHA-1

If the DNB and the depositor intend to make use of reciprocal checksum validation, the processes mentioned above must be used to calculate checksums for the digital objects to be validated; these checksums must then be transmitted to the other party.

The checksums generated by the DNB for the transfer package are checked against the checksums delivered by the depositor with the transfer package using the checksum process selected by the depositor on adding the appropriate file ending to the checksum file. If the checksums are identical, the file is processed further for inclusion in the DNB's archive system. The same checksum process is performed for all parts of the transfer package if checksums are delivered at this level. The checksum file may only contain the checksum itself; other information such as file name repetition is not permitted.

Example of checksum delivery for the md5 process.

```
9783958750685.zip
9783958750685.zip.md5
    catalogue_md.xml
    catalogue_md.xml.md5
    content
        2711.pdf
        2711.pdf.md5
        2711.jpeg
        2711.jpeg.md5
```

5 Supplementary materials

Digital publications may be accompanied by supplementary materials such as answer sheets, questionnaires or research data.

The deposit of supplementary material is mandatory. Research data must be deposited if it is associated with a dissertation; it is subject to a separate deposit policy.

If you are planning to deposit supplementary material or research data, please send an e-mail to np-technik@dnb.de to discuss the details.

6 Examples

A test package is available for download at https://www.dnb.de/hotfoldertestdata (in German only)

Contact

If you have any technical questions about depositing online publications, please contact the technical service at np-technik@dnb.de.