

# Preparing your files

Please observe the following guidelines when you prepare your publications and research data for submission.

- [File formats for archiving](#)
- [Creating PDF Files](#)
  
- [Recommendations for publications and research data](#)
  - [1. Avoid password protection, encryption and compression](#)
  - [2. Avoid special characters](#)
  - [3. Proper use of file extensions](#)
- [Further recommendations for research data](#)
  - [1. Data selection](#)
  - [2. Choose open formats](#)
  - [3. How to package data into ZIP or tar archives](#)
    - [3.1 Limit the length of file and folder names](#)
    - [3.2 Split large data packages](#)
    - [3.3 Instructions on creating container files](#)

## Recommendations for publications and research data

### 1. Avoid password protection, encryption and compression

Your files should not be password protected, encrypted or compressed.

### 2. Avoid special characters

Avoid special characters in names of files and folders. These characters hamper compatibility because they lead to undesired effects that depend on the operating system.

Avoid the following characters:

`\ / ? : * " > < | : # % " { } | ^ [ ] ` ~` as well as blanks

Non ASCII characters such as `¢` <sup>TM</sup> `@` , umlauts (`ä ö ü`), diacritics such as `à é ô` etc.

The following ASCII characters are permitted:

**ABCDEFGHIJKLMN OPQRSTUVWXYZabcdefghijklmnopqrstu vwxyz0123456789**

We are currently not aware of problems with the following characters:

`! $ & ' ( ) + , - . ; = @ _`

### 3. Proper use of file extensions

File extensions (such as `.txt`, `.pdf`) should be consistent with the file format. Avoid saving files without file extensions or using special characters in the file extension.

## Further recommendations for research data

### 1. Data selection

We recommend to carefully select the data, such that the archived data is of scientific relevance and worth archiving over the long term. Please remove unneeded data and avoid storing identical files in several places, such as storing ZIP-Files and their unzipped contents, multiple backups or temporary files. Private information does not belong into the Research Collection.

### 2. Choose open formats

To allow for long-term readability of your files, non-proprietary file formats that follow open and properly-documented standards should be preferred. If you plan to archive your data for more than 10 years, it is recommended to convert unusual file formats into more popular formats. Please consult the chapter [File formats for archiving](#) for further information on this topic.

### 3. How to package data into ZIP or tar archives

We currently recommend packing the data into ZIP or tar container files in order to archive large collections of heterogeneous research data sets (without active validation and preservation measures) over a limited time period. Using container files has the advantage that all files in an archival package are uploaded (and downloaded) in a single batch. Furthermore, the folder structure remains unchanged. Since uncompressed \*.zip and \*.tar files are well standardised formats, they can be unpacked in the long term.

#### 3.1 Limit the length of file and folder names

Avoid overly long path lengths in your folder structure. Long file names combined with a detailed folder hierarchy may lead to path lengths exceeding 256 characters, which causes some issues for Windows users. Such containers cannot be completely unpacked with WinZip. Effective path lengths are further increased when special characters are used in file names and when container files are unpacked within subfolders. We thus recommend using path lengths of less than roughly 200 characters.

#### 3.2 Split large data packages

Large data sets can lead to difficulties uploading your data and also when data are downloaded using the viewer. We have no influence on several factors that cause these difficulties (such as browser and internet connection).

We recommend using **ZIP or tar files not exceeding a maximal size of 10 GB**. If your archival package exceeds this size, please split it into meaningful subunits and use one ZIP or tar container for each subunit. You will then be able to upload all your container files in a single batch. Please do not use the split feature of WinZip when splitting your data.

Additionally, the cumulative amount of **data attached to one item in the Research Collection should not exceed 50 GB**.

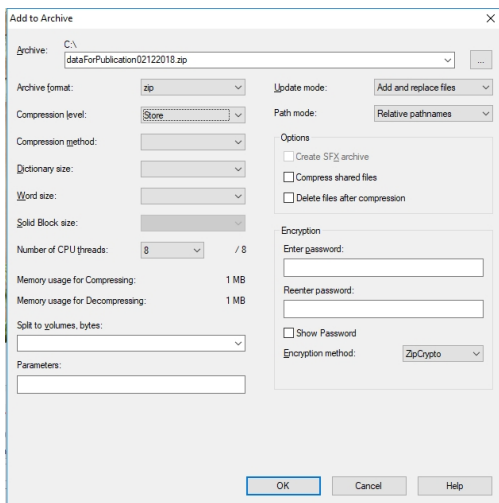
#### 3.3 Instructions on creating container files

- Create archives (container files) only with extensions .zip or .tar (do not use .7z, tar.gz, .rar, and so on).
- Do not use archives within your ZIP or tar files.
- Please create your archives without any data compression (compression level "store")
- Avoid encrypting your files with a password.

##### Windows

On Windows operating systems, you should create ZIP archives by using the [software tool 7-Zip](#).

You start by selecting your files and folders in Windows. Then, click your right mouse button to open a menu. Select the software tool "7-Zip", and "Add to archive ...". A dialog box opens as shown below. In the white field on top, you write the name of your archive file. Use the option "zip" and the compression level "store".



##### Macintosh

On a Macintosh computer, you should create tar archives.

You may create tar container files either by using the command line (`tar -cvf <archive_name.tar> <folder_to_tar>`) or by using the software [Keka](#). If you choose the second option, you start the program Keka and select "Compression" in "Preferences". You then select the default format "TAR" as in the dialog box shown below. You finally drag your folder onto the Keka icon and fill in the name of your archival file. You should select the option "Exclude Mac resource forks (e.g. .DS\_Store)" in the Keka settings. This prevents the inclusion of macOS specific hidden files.

