

# Dateien vorbereiten

Bitte beachten Sie die folgenden Informationen beim Vorbereiten Ihrer Publikationen und Forschungsdaten für den Upload.

- [Archivtaugliche Dateiformate](#)
- [PDF-Dateien erzeugen](#)
  
- [Empfehlungen für Publikationen und Forschungsdaten](#)
  - [1. Passwortschutz, Verschlüsselung und Komprimierung vermeiden](#)
  - [2. Sonderzeichen vermeiden](#)
  - [3. Dateierweiterung bewusst verwenden](#)
- [Weiterführende Empfehlungen für Forschungsdaten](#)
  - [1. Daten selektieren](#)
  - [2. Offene Formate wählen](#)
  - [3. Packen von Dateien in ZIP- oder Tar-Archive](#)
    - [3.1 Länge der Datei- und Ordnernamen beschränken](#)
    - [3.2 Grosse Datenpakete portionieren](#)
    - [3.3 Anleitung zum Erzeugen von Containern](#)

## Empfehlungen für Publikationen und Forschungsdaten

### 1. Passwortschutz, Verschlüsselung und Komprimierung vermeiden

Dateien sollten nicht passwortgeschützt, verschlüsselt oder komprimiert sein.

### 2. Sonderzeichen vermeiden

Verwenden Sie nur Buchstaben, Zahlen, Unterstrich (\_) und Bindestrich (-) in den Namen von Ordnern und Dateien, also keine Leerzeichen, Schrägstriche, Umlaute usw. Solche Zeichen erschweren die Kompatibilität, da sie je nach Betriebssystem unerwünschte Effekte erzeugen.

Problematische Zeichen sind:

`\ / ? : * " > < | : # % " { } | ^ [ ] ` ~` sowie Leerzeichen

Nicht-ASCII-Zeichen wie z.B. `¢ ™ ®`, Umlaute (`ä ö ü`), diakritische Zeichen wie `à é ô` usw.

Erlaubt sind folgende ASCII-Zeichen:

**ABCDEFGHIJKLMN OPQRSTUVWXYZabcdefghijklmnopqrstuvwxyz0123456789**

Zudem sind uns bei folgenden Zeichen bis jetzt keine Probleme bekannt:

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

### 3. Dateierweiterung bewusst verwenden

Die Dateierweiterung (z.B. `.txt`, `.pdf`) sollte mit dem tatsächlichen Dateiformat konsistent sein. Vermeiden Sie es, Dateien ohne Erweiterung zu speichern. Vermeiden Sie Sonderzeichen in der Dateierweiterung.

## Weiterführende Empfehlungen für Forschungsdaten

### 1. Daten selektieren

Wählen Sie nur Daten, die einen wissenschaftlichen Bezug haben und auch wirklich langfristig verfügbar bleiben sollen, für die Publikation/Archivierung aus. Vermeiden Sie Redundanzen wie ZIP-Files neben entpackten ZIP-Inhalten, mehrfache Backups oder temporäre Dateien. Private Daten gehören nicht in die Research Collection.

### 2. Offene Formate wählen

Im Hinblick auf eine langfristige Lesbarkeit sollten Sie darauf achten, dass Ihre Daten nach Möglichkeit in nicht-proprietären Formaten gemäss offenen, dokumentierten Standards gespeichert sind. Wenn die Daten länger als 10 Jahre nutzbar bleiben sollen, empfiehlt es sich, seltene Formate in gebräuchlichere zu konvertieren. Hinweise dazu gibt Ihnen die Anleitung zum Thema [Archivtaugliche Dateiformate](#).

### 3. Packen von Dateien in ZIP- oder Tar-Archive

Für die befristete Publikation bzw. Archivierung von grösseren Mengen heterogener Forschungsdaten ohne aktive Validierungs- und Erhaltungsmaßnahmen empfehlen wir zurzeit das Packen der Dateien in Containerformate (ZIP oder Tar). Dieses Verfahren erlaubt es, sämtliche zu einem Archivpaket gehörenden Dateien in einem Schritt hochzuladen und auch als Gesamtpaket wieder herunterzuladen. Dabei bleiben ursprüngliche Ordnerstrukturen gewahrt.

#### 3.1 Länge der Datei- und Ordnernamen beschränken

Vermeiden Sie bei der Organisation Ihrer Daten überlange Pfade. Eine Kombination von langen Dateinamen mit fein abgestuften Ordnerstrukturen kann schnell zu Pfadlängen über 256 Zeichen führen. Diese sind bei der Weiterverarbeitung der Daten auf Windows problematisch. Sonderzeichen in den Dateinamen sowie das Auspacken von Containern in Unterordnern können die Pfade zusätzlich verlängern. Daher empfehlen wir im Sinne einer Faustregel eine maximale Pfadlänge von 200 Zeichen.

#### 3.2 Grosse Datenpakete portionieren

Grosse Datenmengen können sowohl beim Upload als auch beim späteren Download Schwierigkeiten bereiten. Ab welcher Menge die Probleme auftreten, ist von verschiedenen Faktoren abhängig, die wir nicht beeinflussen können (Browser, Internetverbindung).

Wir empfehlen eine **maximale Grösse pro ZIP- oder Tar-Datei von 10 GB**. Falls die Grösse Ihres Archivpakets diesen Wert überschreitet, teilen Sie es bitte in sinnvolle Untereinheiten auf und packen Sie diese in einzelne ZIP- oder Tar-Container. Mit einem Mehrfach-Dateiupload können Sie diese als eine Einheit hochladen. Bitte benutzen Sie für die Aufteilung nicht die Split-Funktion in WinZip.

**Die kumulierte Datenmenge pro Eintrag in der Research Collection sollte zudem auf 50 GB beschränkt werden.**

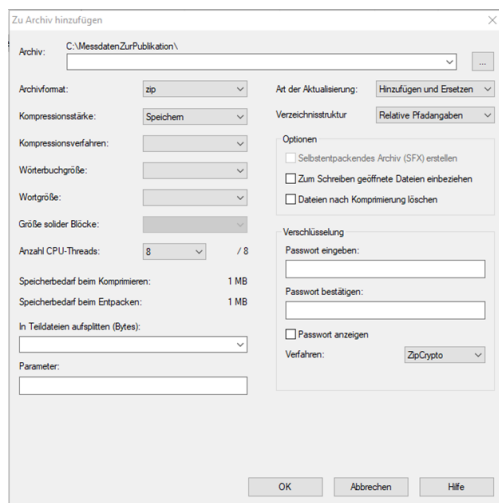
#### 3.3 Anleitung zum Erzeugen von Containern

- Erzeugen Sie Archive (Container Dateien) ausschliesslich mit der Endung .zip oder .tar (kein .7z, tar.gz, .rar usw.).
- Verwenden Sie innerhalb ihrer .zip oder .tar Datei keine weiteren Archiv-Dateien
- Erstellen Sie Container bitte ohne Komprimierung (also Kompressionsstärke "speichern" auswählen)
- Wenden Sie keine Passwort-Verschlüsselung an.

##### Windows

Windows-Nutzer sollten ZIP-Archive generieren (indem sie im [Tool 7-Zip](#) das Archivformat "zip" auswählen).

Dazu Files und Ordner selektieren, mit Rechts-Click «7-Zip» auswählen. Mit der Selektion von «Zu einem Archiv hinzufügen ...» gelangen sie zum Dialog im Bild unten. Zuoberst im weissen Feld können Sie den Namen des Archivfiles eingeben. Benutzen Sie für das Archivformat die Option «zip» und für die Kompressionsstärke die Option «Speichern».



##### Macintosh

In einer Mac-Umgebung sollte das Tar-Format gewählt werden.

Tar Container können entweder mit dem Befehl „tar“ auf der Kommandozeile (tar -cvf <archive\_name.tar> <folder\_to\_tar>) oder mit dem Programm [Keka](#) erzeugt werden. Letztere Methode besteht aus folgenden Schritten: Starten Sie das Programm Keka, unter «Preferences» selektieren Sie «Compression». Dort wählen Sie das Default Format «TAR» (wie im Bild unten). Danach ziehen Sie den zu verpackenden Ordner auf den Keka-Icon. Sie können nun den Namen des Archivfiles angeben um es von Keka erzeugen zu lassen.

