

EMPFOHLENE DATEIFORMATE

Da das Hauptziel von SODHA die Verbreitung und Weiterverwendung von Forschungsdaten ist, und weil der [SODHA-Depositvertrag](#) keine Übertragung des Eigentums an hinterlegten Daten an das Staatsarchiv beinhaltet, verpflichtet SODHA nicht zur Umwandlung der hinterlegten Dateien in offene / nichtproprietäre Formate.

Allerdings ist Interoperabilität — eines der FAIR-Prinzipien — von grundlegender Bedeutung, um den Zugang und die Weiterverwendung der Forschungsdaten zu ermöglichen, weshalb wir die Nutzung der nachstehenden Dateiformate für die hinterlegten Dateien empfehlen. Wir raten den Hinterlegern an, Ihre Daten in **offenen Formaten** wie die auf der zweiten Seite dieses Dokuments aufgeführten zu hinterlegen. Offene Formate garantieren, dass Ihre Daten langfristig zugänglich sein werden, da sie nicht von Unternehmen abhängen, die in Konkurs gehen können (wodurch technologische Geheimnisse solcher proprietären Dateiformate verschwinden).

Falls Sie wissen oder befürchten, dass gewisse Teile Ihrer Arbeit durch eine Umwandlung in ein offenes / nichtproprietäres Format fehlerhaft werden oder verloren gehen, empfehlen wir Ihnen, sowohl die Originaldatei als auch eine Kopie in einem der nachstehenden Standards zu hinterlegen.

Hinweis: Bei **tabellarischen Dateien** ist eine Umwandlung ist nicht erforderlich. Tabellarische Daten können im SODHA-Datenarchiv in nahezu allen gängigen Dateiformaten (einschließlich .sav, .sas etc. von SPSS) hinterlegt werden, da das Dataverse-System bei der Einspeisung automatisch eine Version der Daten im Format .tab anlegt und den Benutzern so die Möglichkeit bietet, die Datei in diesem Format oder im Originalformat herunterzuladen.

Wir raten schlussendlich davon ab, komprimierte Dateien zu hinterlegen, da es zu Datenbeschädigungen mit schwerwiegenden Folgen kommen kann: Die Korruption eines einzigen Bits innerhalb einer Datei kann dazu führen, dass große Teile der Daten (wenn nicht alle) unlesbar werden.

↓ Die Liste der empfohlenen Dateiformate befindet sich auf Seite 2 ↓

Digitales Objekt	Empfohlenes Dateiformat
Kalkulationstabelle / tabellarische Daten	.csv – .tsv/.tab – .ods – .odf
Text	.pdf (PDF/A-2a vorzugsweise ¹) – .odt – .txt – vorzugsweise in UTF-8 Kodierung ²
Präsentation	.odp – .odt
Datenbank	.xml – .csv – vorzugsweise samt Dokumentation in .xml oder SQL über die Verbindungen zwischen Tabellen, Trigger etc.
Bild	.tiff (nicht komprimiert TIFF 0.6) – .jpg (JPEG oder JPEG 2000) – .png
Ton	.mp3 – .flac – .ogg – .oga
Video	.dpx
Karte, technische Zeichnung	.dwg – .svg
Geografische Daten	.gml (/ .xml/ .kml) – GeoTIFF (.tiff)
Webseiten	XHTML 1.0 – .css – .js – .jsp – .php

¹ PDF/A3 kann nicht-genormte Objekte enthalten, daher ist die langfristige Lesbarkeit von Dateien in diesem Format nicht unbedingt gewährleistet.

² UTF-8 gewährleistet, dass die meisten wenn nicht alle eventuell in den Daten enthaltenen Sonderzeichen (e.g. *é, ç, ü...*) beim Speichern nicht korrumpiert werden (beispielsweise in Word: *façade* wird im Text als „fa◊ade“ angezeigt).

Versionsverlauf dieses Dokuments

Datum	Versionsnummer	Vorgenommene Änderungen
30-07-2021	1.0	Erste Version