

Tabelle 1: Dateiformate: Information zur Archivierung

Textformate		<i>langfristig</i>	<i>mittelfristig</i>	<i>ungeeignet</i>
PDF/A	*.pdf	✓		
Unformatierter Text ¹	*.txt, *.c, *.cpp, *.m, ...	✓		
XML ²	*.xml, *.xsd, *.html, ...	✓		
PDF ³	*.pdf		✓	
Unformatierter Text ⁴	*.txt, *.c, *.cpp, *.m, ...		✓	
HTML ⁵	*.html, *.htm		✓	
Word	*.docx		✓	
PowerPoint	*.pptx		✓	
LaTeX, TeX ⁶	*.tex		✓	
Programmcode ⁷	*.c, *.rb, *.go, ...		✓	
Word	*.doc			✓
PowerPoint	*.ppt			✓

Empfohlene Kovertierungen

- Möglichst Word und PowerPoint Dateien zu PDF/A-1b Dateien konvertieren
- LaTeX and TeX möglichst zu PDF/A konvertieren
- Die Qualität von Konvertierungen muss sorgfältig visuell überprüft werden. Besonders zu beachten sind Formeln, Sonderzeichen, Umlaute, spezielle Fonts, Schreibfehler, Auswählen und Suchen in Text, Tabellen, Farben, transparente Objekte, Kommentare, Vektorgrafiken sowie mehrfache Zeichenebenen.

¹kodiert aus ASCII, UTF-8, oder UTF-16 mit Byte Order Mark (BOM)

²Schema und Buchstabenkodierung explizit in der Datei angegeben

³mit eingebetteten Fonts

⁴ISO 8859-1 kodiert

⁵Inkl. Doctype Deklaraiion und externe Links möglichst vermeiden

⁶Spezialfonts sollten möglichst migeliefert werden

⁷Benutzte lizenzfreie Softwarepakete und Libraries sollen möglichst mitgeliefert werden

		<i>langfristig</i>	<i>mittelfristig</i>	<i>ungeeignet</i>
Spreadsheets und Tabellenformate				
Comma oder Tab-begrenzte Textdateien	*.csv	✓		
Excel (Containerformat)	*.xlsx		✓	
Excel (Binärformat)	*.xls			✓

Empfohlene Kovertierungen

- Für wichtige Objekte (wie z. B. Figuren) sollte möglichst auch eine Kopie als separate Datei abgespeichert werden
- Tabellen können zu ASCII Text-Dateien konvertiert werden (*.csv)

		<i>langfristig</i>	<i>mittelfristig</i>	<i>ungeeignet</i>
Workspace Speicherung für Matlab, R oder S-Plus				
Textdateien für S-Plus ¹	*.sdd		✓	
Matlab ²	*.mat		✓	
Matlab ³	*.mat			✓
R-Dateien ⁴	*.R, *.RData			✓

Empfohlene Kovertierungen

- Matlab Dateien (*.mat) als v7.3 MAT-Dateien abspeichern weil es dadurch einm HDF5-basierten Standard folgt (HDF5 ist ein offener Standard für Tabellen, Metadaten und komplexe Datenstrukturen)
- Der R-Workspace sollte mit dem Paket rhdf5 in HDF5 gespeichert werden. Die S-Plus Funktion data.dump produziert eine Datei welches mit der R-Funktion data.restore gelesen werden kann
- Für komplexe Datenstrukturen ist es meist nicht sinnvoll den Workspace als ASCII zu speichern, weil dies auf schwer lesbare Dateien führt
- Wichtige Tabellen im Workspace sollten zusätzlich als CSV-Datei gespeichert werden

¹Der ASCII Text ist langfristig nutzbar, die spätere maschinelle Lesbarkeit ist jedoch unsicher

²in HDF Format speichern, denn nichttriviale Matlab ASCII files können mit load nicht gelesen werden

³Binäre Dateien

⁴Binäre Dateien

		<i>langfristig</i>	<i>mittelfristig</i>	<i>ungeeignet</i>
Rastergraphikformate				
Tagged Image File Format (TIFF)	*.tif	✓		
Portable Network Graphics (PNG)	*.png	✓		
JPEG2000	*.jpeg, *.jpg	✓		
TIFF komprimiert	*.tiff		✓	
Graphics Interchange Format (GIF)	*.gif		✓	
JPEG/JFIF	*.jpg, *.jpeg		✓	
JPEG2000 ¹	*.jpg, *.jpeg, *.jp2		✓	
Bitmap (BMP)	*.bmp		✓	
Vektorgraphikformate				
Scalable Vector Graphics (SVG) ²	*.svg	✓		
Graphik InDesign	*.indd			✓
Illustrator	*.iat			✓
Encapsulated Postscript	*.eps			✓

Empfohlene Kovertierungen

- Vektorgraphikdateien werden langfristig eher schwieriger zu öffnen sein als Rastergraphikdateien (Bitmaps, GIF, usw.). Auch das Einbetten von Vektorgrafiken in PDF-Dateien ist fehleranfällig.
- Dateien in speziellen Vektorgraphik Formaten, wie InDesign (*.indd) oder illustrator (*.ait) sollten möglichst auch als baseline TIFF, PDF/A, SVG oder JPEG Datei abgespeichert werden.
- Die Qualität der Konvertierung muss sorgfältig visuell überprüft werden (Schärfe, Auflösung, Farben, halbdurchsichtige Objekte, Beschriftungen)

¹komprimiert

²ohne Javascript binding