

Tabelle 1: Dateiformate: Information zur Archivierung

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

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

<sup>2</sup>Schema und Buchstabenkodierung explizit in der Datei angegeben

<sup>3</sup>mit eingebetteten Fonts

<sup>4</sup>ISO 8859-1 kodiert

<sup>5</sup>Inkl. Doctype Deklaraiion und externe Links möglichst vermeiden

<sup>6</sup>Spezialfonts sollten möglichst migeliefert werden

<sup>7</sup>Benutzte lizenzfreie Softwarepakete und Libraries sollen möglichst mitgeliefert werden

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

<sup>1</sup>Der ASCII Text ist langfristig nutzbar, die spätere maschinelle Lesbarkeit ist jedoch unsicher

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

<sup>3</sup>Binäre Dateien

<sup>4</sup>Binäre Dateien

		<i>langfristig</i>	<i>mittelfristig</i>	<i>ungeeignet</i>
<b>Rastergraphikformate</b>				
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 <sup>1</sup>	*.jpg, *.jpeg, *.jp2		✓	
Bitmap (BMP)	*.bmp		✓	
<b>Vektorgraphikformate</b>				
Scalable Vector Graphics (SVG) <sup>2</sup>	*.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)

<sup>1</sup>komprimiert

<sup>2</sup>ohne Javascript binding