

## File Formats

### Guidelines for formatting data

Research data is only useful if the data can be read and understood. Will you still be able to access your file in 20 years? Where possible use text files. Data will be usable even if formatting is lost. Consider exporting to CSV, XML, JSON free-form text.

Below are recommended and acceptable formats for tabular, geospatial, textual, image, audio, video, and documentation and scripts data.

#### Tabular data with extensive metadata (Variable labels, code labels, and defined missing values)

<p><b>Recommended Formats:</b>          SPSS portable format (.por)          Delimited text and command ('setup') file (SPSS, Stata, SAS, etc.)          Structured text or mark-up file of metadata information, e.g., DDI XML file</p>	<p><b>Acceptable Formats:</b>          Proprietary formats of statistical packages: SPSS (.sav), Stata (.dta), MS Access (.mdb/.accdb)</p>
--	--

#### Tabular data with minimal metadata (Column headings, variable names)

<p><b>Recommended Formats:</b>          Comma-separated values (.csv)          Tab-delimited file (.tab)          Delimited text with SQL data definition statements</p>	<p><b>Acceptable Formats:</b>          Delimited text (.txt) with characters not present in data used as delimiters.          Widely used formats: MS Excel (.xls/.xlsx), MS Access (.mdb/.accdb), dBase (.dbf), OpenDocument Spreadsheet (.ods).          (Use XLSX or ODS for spreadsheet data)</p>
--	---

#### Geospatial Data (Vector and raster data)

<p><b>Recommended Formats:</b>          ESRI Shapefile (.shp, .shx, .dbf, .prj, .sbx, .sbn optional)          Geo-referenced TIFF (.tif, .tiff)          CAD data (.dwg)          Tabular GIS attribute data          Geography Markup Language (.gml)</p>	<p><b>Acceptable Formats:</b>          ESRI Geodatabase format (.mdb)          MapInfo Interchange Format (.mif) for vector data          Keyhole Mark-up Language (.kml)          Adobe Illustrator (.ai)          CAD data (.dxf or .svg) Use SVG for figures (an open-standard vector format)          Binary formats of GIS and CAD packages</p>
--	--

#### Textual Data

<p><b>Recommended Formats:</b>          DOCX, RFT or ODT for textual data          Rich Text Format (.rtf)          Plain text, ASCII (.txt)</p>	<p><b>Acceptable Formats:</b>          Hypertext Mark-up Language (.html)          Widely used formats: MS Word (.doc/.docx)          Some software-specific formats: NUD*IST, NVivo and ATLAS.ti</p>
--	---

eXtensible Mark-up Language (.xml) text according to an appropriate Document Type Definition (DTD) or schema	
--	--

## Image Data

<b>Recommended Formats:</b> TIFF 6.0 uncompressed (.tif)	<b>Acceptable Formats:</b> JPEG (.jpeg, .jpg, .jp2) if original created in this format GIF (.gif) TIFF other versions (.tif, .tiff) RAW image format (.raw) Photoshop files (.psd) BMP (.bmp) PNG (.png) Adobe Portable Document Format (PDF/A, PDF) (.pdf) PDA is a standardised version of PDF
---	---

## Audio Data

<b>Recommended Formats:</b> Free Lossless Audio Codec (FLAC) (.flac) (Open and loses less than mp3 compression)	<b>Acceptable Formats:</b> MPEG-1 Audio Layer 3 (.mp3) if original created in this format Audio Interchange File Format (.aif) Waveform Audio Format (.wav)
---	--

## Video Data

<b>Recommended Formats:</b> MPEG-4 (.mp4) OGG video (.ogv, .ogg) motion JPEG 2000 (.mj2)	<b>Acceptable Formats:</b> AVCHD video (.avchd)
---	--

## Documentation and Scripts

<b>Recommended Formats:</b> Rich Text Format (.rtf) PDF/UA, PDF/A or PDF (.pdf) XHTML or HTML (.xhtml, .htm) OpenDocument Text (.odt)	<b>Acceptable Formats:</b> Plain text (.txt) Widely used formats: MS Word (.doc/.docx), MS Excel (.xls/.xlsx) XML marked-up text (.xml) according to an appropriate DTD or schema, e.g., XHTML 1.0
---	---

The UK Data Service provide guidance on file formats for data sharing, reuse, and preservation.

- [UK Data Service Recommend File Formats](#)

For help or advice with any data format query please email [researchdata@soton.ac.uk](mailto:researchdata@soton.ac.uk)