FILE FORMATS RECOMMENDED BY SOHDA

Because the main goal of SODHA is to allow for dissemination and reuse of research data, and because the SODHA Deposit Agreement does not transfer ownership of deposited data over to the State Archives, SODHA does not impose the conversion of deposited data files into open / non-proprietary file formats.

However, interoperability – one of the FAIR principles – is key to enabling access to and reuse of research data, which is why we encourage you to consider the following formats for your files. That is why we encourage depositors to deposit their data in **open formats** such as the ones listed on the second page of this document. Open formats also guarantee that your data will be available in the long run, as they do not depend on companies that could go bankrupt (and take such technological secrets as proprietary file formats to the grave).

But if you know or suspect that some aspects of your work might be lost or even corrupted upon saving it into an open / non-proprietary format, we recommend that you deposit both the original file and a copy formatted according to one of the standards listed below.

Please note: You do not need to do this for **tabular data files**. Tabular data can be deposited in the SODHA Dataverse in almost all commonly known tabular data file formats (including SPSS's .sav, .sas…) owing to the fact that Dataverse automatically creates a .tab version of the data upon ingest and then lets users download the file in either this format or the original one.

Finally, we recommend that you do not deposit compressed files, as corruption issues can be extremely damaging to them: the corruption of just one bit inside the file can make large parts of the data (if not all of the data) unreadable.

Types of digital objects	Recommended formats
Spreadsheets / Tabular data	.csv – .tsv/.tab – .ods – .odf
Texts	.pdf (PDF/A-2a ideally [1]) – .odt – .txt – ideally in UTF-8 encoding [2]
Presentations	.odp – .odt
Databases	.xml – .csv – ideally along with documen-tation in .xml or SQL about the relations between tables, triggers, etc.
Images	.tiff (non-compressed TIFF 0.6) – .jpg (JPEG or JPEG 2000) – .png
Sounds	.mp3 – .flac – .ogg – .oga
Videos	.dpx
Maps, technical designs	.dwg – .svg
Geographical data	.gml (/.xml/.kml) – GeoTIFF (.tiff)
Webpages	XHTML 1.0 – .css – .js – .jsp – .php

[1] PDF/A3 can contain non-normalized objects, so the long-term readability of files in this format cannot always be guaranteed.

[2] UTF-8 guarantees that most if not all special characters (e.g. *é*, *ç*, *ü*…) that might appear in your data will not be corrupted upon saving (which is what happens when, for example, the word *façade* appears as "fa�ade" in a text).

Version history of this document

Date	Version number	Changes made
07-09-2020	1.0	First version
03-02-2021	2.0	New version history section, formatting
03-03-2021	2.1	Brackets removed from the full name of SODHA
06-05-2021	2.2	Font of the document changed from Arial to Calibri