UK Polar Data Centre Data Format Guidance

When depositing data, the format used should ideally be an open, non-proprietary format in common usage by the research community.

Open formats are more likely to be readable by more than one application and therefore are available to the widest possible audience without the need for access to restrictive software. They also retain the best chance of being readable in the future.

Proprietary formats, used only by a specific software program or specific software version, are likely to present problems for future use. Someone without a license to the software may not be able to open the file at all.

Data that are stored in a proprietary format should be converted to an open format. For example, the preferred format for depositing tabular data is comma separated values (.csv) rather than Microsoft Excel (.xlsx). When converting files between formats, some information loss or corruptions may occur e.g. loss of formatting, special characters, links. Check the integrity of the converted file as thoroughly as possible after converting the format.

PDC's Preferred Formats:

As we accession a broad range of data this is not a complete list. Please get in touch with us at PDCServiceDesk@bas.ac.uk to discuss the most suitable format for a particular data type. Proprietary formats will be considered if there is no open alternative or if conversion would result in significant data loss. These will be dealt with on a case-by-case basis.

General File Formats		
Data Type	**Preferred Format**	**Common File Extension**
Text (document)	Text file (plain/ASCII)	.txt
	PDF, PDF/A	.pdf
	RTF	.rtf
Tabular data (spreadsheet)	Delimited text	.csv, .tsv, .txt
Database	XML	.xml
	JSON	.json
	Delimited text	.csv, .tsv, .txt
Gridded/Multidimensional	NetCDF	.nc
	HDF	.hdf
Geospatial	GeoTIFF	.tiff
	Shapefile	.shp, .shx, .dbf
	GeoPackage	.gpkg
Image	JPEG, JPEG2000	.jpeg, .jpg, .jp2
	TIFF	.tiff,
	PNG	.png
	SVG	.svg
	VTK	.vtk, .vtu, .vtp
Audio	MP3	.mp3
	FLAC	.flac
Video	MP4	.mpg, .mp4
	MKV	.mkv

Containers	TAR, ZIP, GZIP	.tar, .zip, .gzip, .tar.gzip

In addition to the above general formats, the following discipline-specific formats are accepted.

Other Discipline-Specific File Formats		
Data Type	**Preferred Format**	**Common File Extension**
Biodiversity data	Darwin Core Archive	DwC-A (.xml and .txt)
Oceanographic data	ASCII	.ascii
GPS data	Raw GPS (RINEX)	.yyt (yy=year, t=type)
	ASCII XYZ	.txt, .xyz
Seismic data	SEG-Y	.sgy
	SEG-2	.sg2
	SEED	.seed, .dataless, .miniseed, .mseed, .dless
Atmospheric data	GRIB	.grib
Multibeam Echosounder (bathymetric) data	Raw data: Format it was collected in along with any system file associated with acquisition	.all (for Kongsberg)
	Processed data: MBSystem file format but will accept other industry standard formats (e.g. CARIS)	.mbXX
	+ exported ASCII XYZ data	.xyz, .txt