

There are several important things to consider when depositing data; that the files are in the **correct format**; that proper **filenaming** conventions are used; and that they are accompanied by appropriate **documentation**.

File formats

Unfortunately, we do not have the software to deal with all the file types produced during the course of research, so it is important that data is provided in formats, preferably open standards, that we can access, migrate and preserve for the future.

- Do check that the files you are depositing are accepted by the ADS (see table overleaf).** If you are concerned about any of your data formats please contact us as soon as possible.
- Don't include duplicate, draft or spurious files within your deposit.** We aim to preserve and disseminate quality data but do not edit or proof read the contents of deposited files. We therefore ask the depositor to make sure that the data deposited is in its final form.

Filenaming conventions

Our UNIX operating system requires that certain filenaming conventions should be adhered to when transferring files. Files must have a file extension to help the ADS and future users of the resource determine the file type, these are normally 3 characters long and we recommend they are in lower case.

- Do use only alphanumeric characters (a-z, 0-9), the hyphen (-) and the underscore (_).** Both upper and lower case characters and numbers can be used in a filename but keep file names within your project consistent and ensure that supplied documentation accurately reflects the case of your filenames.
- Don't use spaces or full stops (.) within filenames.** Full stops should only be present where the filename is separated from the file extension e.g. `_doc` or `_pdf`. Spaces can usually be replaced with the underscore (`_`) character.
- Do use a consistent scheme and case when naming files.** A descriptive filename helps explain the contents of the file, for example `12102004_trench_1.tif` could be a digital photograph of trench 1 taken on 12/10/2004. A non-descriptive file name might be a unique id number allocated to an image within an accompanying image catalogue database. Non-descriptive filenames are acceptable but their content must be adequately described in accompanying metadata. Consistent use of case ensures that files can be reliably identified on case-sensitive operating systems such as UNIX where `report.doc` would be recognised as a different file to `Report.doc`

Documentation

(Full documentation requirements are available at <http://ads.ahds.ac.uk/project/userinfo/deposit.html>. Also see our **Guides to Good Practice** at <http://ads.ahds.ac.uk/project/goodguides/g2gp.html>)

In order for us to undertake proper archiving of your data, we need to have as much information about it as possible, so we would ask that you provide metadata for your files.

- Do include a list of the files you have sent.** This should list filename, file size, software package and version used to create the file and a short description of the file's contents. This documentation should be sent to us in digital form so we can preserve it alongside the data files themselves for future users of the resource (refer to the documentation columns overleaf).

If you have any queries please contact:

Catherine Hardman,
Archaeology Data Service,
Department of Archaeology,
University of York, King's Manor
York YO1 7EP
Tel. 01904 433982 Email: csh3@york.ac.uk

	File Format		Documentation (software, version and platform; mandatory, if relevant)
CAD (Vector graphics)	Preferred	AutoCAD – DWG SVG	<ul style="list-style-type: none"> AI, CDR, SVG – Relationship to other documents, caption DWG, DXF – AutoCAD/DXF version, significance of conventions (layers, colours, linetypes, hatch styles, symbols, etc.), relationship to other files (databases, object libraries, etc.)
	Accepted	Adobe Illustrator - AI DXF CorelDraw - CDR	
Databases	Preferred	Access – MDB OpenDocument Database - ODB Delimited text	<ul style="list-style-type: none"> A data dictionary i.e. a list of all tables and their fields, including the data types and field sizes for text fields, relationships between tables. Row counts for each table. For delimited text the delimiters and qualifiers.
	Accepted	Dbase - DBF Paradox - DB, DBQ	
GIS	Preferred	ESRI Shapefile - SHP + SHX + DBF Geo-referenced TIF Image - TIF + TFW	<ul style="list-style-type: none"> Method of capture (Total station survey etc) Data source (Purchased from OS etc) Scale/resolution of data capture Scale/resolution at which data is stored Assessment of data quality (Root mean square error etc) Date of capture/purchase
	Accepted	ArctInfo Ungen ArctInfo Interchange - E00 ESRI Grid MapInfo Interchange Format – MIF + MID Spatial Data transfer standard - DDF MOSS - EXP Vector product Format - VPF	
Images	Preferred	Uncompressed Baseline TIFF v.6 -TIF	<ul style="list-style-type: none"> Caption mandatory. Also consider depositing a database or table of metadata, possible fields include: title, photographer, date taken, period, monument type, object type, method, country, district, parish.
	Accepted	Portable Network Graphics – PNG Joint Photographic Expert Group – JPG Graphics Interchange Format - GIF Bit-Mapped Graphics Format - BMP PhotoCD - PCD Photoshop (Adobe) – PSD	
Movies	Preferred	MPEG1 & 2 – MPG, MPEG MPEG4 – MPG4	<ul style="list-style-type: none"> Name & version of video codec, video dimension (in pixels), frame rate (fps) and bit rate. Name and version of audio codec - sample frequency, bit-rate & channel information. Length (hours, minutes, seconds) of file and size. Copyright clearances where required (e.g. oral history). Caption and short description for each movie file.
	Accepted	DivX – DIVX, AVI	
Spreadsheets	Preferred	CSV Microsoft Excel – XSL OpenDocument Spreadsheet - ODS OpenOffice.org Calc – SXC	<ul style="list-style-type: none"> Purpose & content of spreadsheet & worksheets. Content of each column and row if not obvious. Data type and scale for each column. Key for any codes within data. Column and row counts. Documentation of any extra features the spreadsheet may contain, i.e. formulae, macros, charts, comments and any significant characteristics to be preserved.
	Accepted	Lotus 1-2-3 - 123, WK4, WK3, WK1, WKS	
Statistics	Preferred	Delimited text	<ul style="list-style-type: none"> Source(s) of the data, collection methodology Purpose of data Details of tables and samples (columns) Number of rows Type and scale of variables Full description of any coding used Analyses performed on data
	Accepted	SPSS - SAV, POR, SPO SAS - SAS7DBAT, SAS Microsoft Excel – XLS OpenDocument Spreadsheet - ODS SYLK - SLK Microsoft Access - MDB xBase - DBF	
Texts	Preferred	Word – DOC OpenDocument Text - ODT	<ul style="list-style-type: none"> DOC, WPD – software, version and platform SXW, RTF – software, version TXT – text encoding HTML, XHTML – Software used in creation, doctype with HTML schema XML – Text encoding, DTD or Schema SGML – Text encoding
	Accepted	RTF WordPerfect -WPD TXT HTML, XHTML, XML, SGML	
Virtual Reality		X3D VRML Java3D QTVR	<ul style="list-style-type: none"> Original data files that make up the model (image files, CAD models and so on) where available. A video 'fly-by' render of the original VR world to preserve the look and feel.
Geophysics	Preferred	Raw xyz data: TXT, CSV Rendered images: TIF	<ul style="list-style-type: none"> Raw data: location of the survey, conditions and instrumentation Images: details of data processing and interpretation
	Accepted	Rendered images: JPG, PNG Raw data: DAT REP (Contours)	
Audio	Preferred	Waveform Audio WAV Audio Interchange AIF	<ul style="list-style-type: none"> Bit rate (kbps) and sampling frequency range (KHz) if applicable. NOTE: mp3 and ogg both have the option of variable bit rate compression Codec used (where appropriate) and documentation of conversion process Length of recording (mins and secs) Copyright clearances (very important for audio files, especially oral history) Transcriptions of interviews etc (where appropriate)
	Accepted	SUN au AU	

Note: 'Preferred' indicates file formats that the ADS prefer to accept, 'Accepted' indicates that these file types are suitable but that we may require additional documentation with the file. Please contact us if you plan on depositing 'Accepted' file formats.