Geophysics Procedures

Version 1.92

Created date:	26 January 2012
Last updated:	24 August 2017
Review Due:	24 August 2018
Authors:	Jen Mitcham, Tim Evans, Kieron Niven, Ray Moore and Jenny O'Brien
Maintained by:	Digital Archivists
Previous version:	Live

1 Purpose of this document

This is primarily based on the latest version of the *Guides to Good Practice*.[1] More than any other datatype it is important to receive the data in the specified formats. Over the last 5 years we've spent a lot of time in trying to convert proprietary data to "preservation formats", but this process is fraught with pitfalls (see further discussion below).

[1] http://guides.archaeologydataservice.ac.uk/g2gp/Geophysics_App1

2 Formats

Offered format	Accepted	Preservation	Presentation	Notes
Resistivity/magnetometry (includes resistivity profiling, (caesium)gradiometer and magnetometer)				
Raw xyz data: **.txt, .csv**, **.asc** Contors: **.dat**, **.rep**	YES	Raw xyz data: **.txt**, **.csv**, **.asc** (in txt directory(?) Contors: **.dat**, **.rep**	Raw xyz data: **.txt, .csv**, **.asc** (in txt directory or zip if compressed due to size(?). Contors: **.dat**, **.rep**. (for groups of files, or where files are over 50MB, **.zip** archives of the formats listed above should be used for dissemination).	See footnote[2]
Geoplot Grid Files **.dat**, **.grd**, **.grs**, **.hdr**	NO			See footnote.[3]

[2] XYZ text files could more appropriately be called 'XYZ text files'. These are text files (usually only using ASCII coding) in which each text line contains the values of the coordinates (X and Y) as well as a measurement value. As these files were initially used with topographic heights as the measurement value ('Z') they became known as XYZ text files but any other data value could be used ('V'). The three values can either be separated by blanks, tabs or commas. If commas are used as separators the file format is often referred to as csv, meaning 'comma separated values'. It can be useful to add a header line to a XYZ text file that contains the names of the columns represented in subsequent lines (e.g. "X, Y, MagField"). Contors is a DOS programme written by John Haigh of the School of Computing and Mathematics, University of Bradford, for viewing geophysical data. The data are held in (*.DAT) files which are comma delimited ASCII files with 20 lines of data each holding 20 floating point numbers. Sets of *.DAT files hold contiguous blocks of data and their spatial relationships are defined in REP files.

[3] Geoplot makes use of a selection of proprietary file formats. As some of these elements are binary, we advise that the depositor export the raw xyz data into comma separated values format (CSV or TXT). Also, the creation of raster images for presentation purposes is recommended (TIF, PNG). Both of these can be generated with relative ease by using the Geoplot software.

Geoplot Plotmesh file **.plm**	NO			Master grid/mesh file, ASCII text but only of use with the geoplot grid files.
Other Geoplot Files **.sta**, **.his**, **.tem**, **.gip**, **.cip**	NO			
Geoplot Composite Files **.cmp**, **.cmd**, **.cms**	NO			Geoplot Grid files merged into a single composite. These *can* be the raw data but obviously make certain reprocessing (e.g. edge matching) difficult as would need to be broken back down into grids. Not recommended but OK (raw data) if there's nothing else.
GPR				
SEG-Y **.segy** Rev.1	YES	SEG-Y **.segy** Rev.1	SEG-Y **.segy** Rev.1 (Groups of files should be disseminated in **.zip** archives of the formats listed above).	See footnote[4]
.rd3 + **.rad** + **.grd**	NO			GPR: Processed MALA RD3 files.

[4] SEG-Y format is published by the Society of Exploration Geophysicists. It is an openly published binary data format for storing raw GPR data. SEG-Y was originally developed in 1973 but was revised in 2002. The spec is attached as a pdf. There are a number of free viewers for SEG-Y data files. These include GSEGYView 0.2, SeisVU and the SEGY Viewer by PETRA. This format is ideally suited for dissemination. It is strongly recommended to export GPR data to the seismic 'SEG-Y (revision 1)' format as defined by the Society for Exploration Geophysicists. Most GPR packages allow export of their proprietary data to SEG-Y, although not all seem to fully adhere to the SEG-Y standard so it may be best to check the output with one of the free SEG-Y readers available online (e.g. SeiSee, SeiView, SEGYViewer, GSEGYView).

.rd3 + **.rad** + **.mrk** + **.cor** + **.obm**	NO			GPR: Original MALA RD3 files.
.dat + **.par**	NO			GPR: Raw output from ReflexW software
.##r + **.##t** e.g. 12r, 12t	NO			GPR: Processed output from ReflexW software
ERI files				
Text file (AMNBV format, see below) **.txt**	YES	Text file **.txt**	Text file **.txt** (Groups of files should be disseminated in **.zip** archives of the formats listed above).	
Other (includes Alkali vapour, pulse induction)				
Raw xyz data: **.txt, .csv**	YES	Raw xyz data: **.txt**, **.csv**	Raw xyz data: **.txt**, **.csv** (Groups of files should be disseminated in **.zip** archives of the formats listed above).	
Other formats				
Surfer data files **.grd**	NO			Surfer is a general purpose mapping programme. Data files may be ASCII or binary. The ASCII version may be suitable for archiving.[5]
InSite data files	NO			May be an acceptable format, although some

[5] Since Surfer does not support rotated data grids, exporting to xyz does _not_ pose any dangers of numerical errors and there is absolutely no problem with ASCII-xyz, other than the size of the file. But that's no big deal. One could in fact argue that Surfer-ASCII is problematic for archiving as it uses over-length lines of data values and if somehow carriage-returns are inserted during its maintenance, things can break. So I would not consider it a preservation format but a proprietary format.

.dat, **.lst**				elements are binary.
Sensys Magnetometer files **.dgb**, **.disp**, **.prm**	NO			Can be exported to ASCII CSV.
GSSI Radan files **.dzt**	NO			
Documentation				
Rendered images: **.png** or **.tif** with assoc. world file .pgw or .tfw	YES	Rendered images: **.tif** (+ **.tfw)**	Rendered images: **.tif** (+ **.tfw**) or **.png** (+ **.pgw**)	Best to disseminate in the same format as they were received in

Notes on formats

Software

- Geoplot - Prop. Res, Grad, Mag.[6]
- ArchaeoSurveyor / TerraSurveyor - Prop.[7]
- Snuffler - Free. Grad and Mag only.[8]
  Geopsy - Free. Supports wide range of formats and exports.[9]

All geophysical survey data are digital. The geophysical data must be accompanied by ancillary data describing at least the location of the survey, survey conditions and any instrumentation used for survey. Most programmes designed to handle geophysical data can export the data in XYZ format. Depositors should use this facility and to supply copies of the data in this format.

3 Documentation / Metadata

- see table structure for metadata currently held by the OMS
- see current depositor metadata spreadsheet[10]
- The *Guides to Good Practice* also has a detailed section on Project and Technique metadata.[11] Rather than duplicate this verbatim here, please use the *Guides to Good Practice* as a guide to what we need. As with formats we should enforce a strict

[6] http://www.geoscan-research.co.uk/
[7] http://www.dwconsulting.nl/archeosurveyor.html
[8] http://www.sussexarch.org.uk/geophys/snuffler.html
[9] http://www.geopsy.org/
[10] http://archaeologydataservice.ac.uk/advice/FilelevelMetadata.xhtml#Geophysics%20and%20Remote%20Sensing
[11] http://guides.archaeologydataservice.ac.uk/g2gp/Geophysics_5

policy on Geophysics metadata - if we don't have the metadata we can't archive the data.

4 Accessioning checks

- Data is in the correct format (see table below). This is perhaps more important for Geophysics than any other datatype as we have almost no effective software to check content (see below).
- Metadata is present. Again, very important for reuse of this particular data type.
- Secondary data (Exported/derived data): it is common for data to either be exported or derived from the geophysics plot - for example image overlays (vector or raster) or images exported as georectified images for use within a GIS. Although associated with the dataset, these are not (unless specifically stated) the 'geophysics data'. Any use as documentation or metadata should be clearly identified as such by the depositor (for example a raster image with the survey site grid, text file of notes and so on). Otherwise, we should deal with the secondary data according to the appropriate procedures document.

Significant properties

Geophysics data comprises a given value for a specific geographic point (or cell), often just based on a local rather than national grid or UTM (although some instruments record geographic referencing in the data).

Significant properties can therefore include:

- The cell/point/scan values
- Technical details for the values (includes collection strategy, technical parameters included within metadata)
- Relationship to other files in the archive - for example a survey may comprise multiple files
- Georeferencing (if present). Note: a file should not rely on a shapefile for georeferencing. The extents of any survey should be included in the accompanying metadata.

As highlighted several times in the G2GP, migration of proprietary to preservation formats often involves the loss of metadata (often stored in headers and such). It is therefore paramount that metadata is present, accurate and stored with the data.

Images

Exported images are used frequently in geophysics, often it is a convenient way to record survey grid layout as documentation for the raw data. Images should be clearly identified as such, and should be dealt with according to conversion/storage guidelines below. In conversations with geophysicists, it seems that they despise JPG for any kind of export of a data plot - bear this in mind!

5 How to convert files

We should not need to convert geophysics data files. Documentation (text and images) should be dealt with according to the relevant procedures document. However, they should be stored as documentation as per procedures below.

Geophysics data can also be exported as other types of data - shapefiles, CAD and raster images. In this case, and if the files in question are not documentation or metadata, the data should be treated according to the data type.

6 Post-migration checking

Storage

Data should be stored in appropriately named folders, as described in the ADS Repository Operations. Any directory structure from the SIP should try to retained in the AIP. In some cases editing/restructuring may be required, any restructuring must be recorded in the Process table in the CMS.

On occasion, and because of the large number of composite files created by a survey, the dissemination versions will need to be zipped up. In this case disseminate in a sensible, logical fashion under their original file extension.

Otherwise, store data in one of the following directories:

```
/preservation
        /{original_structure}
                geophys_raw_data1.csv
                geophys_raw_data2.csv
                geophys_raw_data3.csv
                geophys_raw_data_rendered_image.tif
                geophys_raw_data_rendered_image.tfw

/preservation
        /{original_structure}
                gpr_raw_data1.segy

/dissemination/
        /{original_structure}
                geophys_raw_data.zip/
                        geophys_raw_data1.csv
                        geophys_raw_data2.csv
                        geophys_raw_data3.csv
                geophys_raw_data_rendered_image.zip/
                        geophys_raw_data_rendered_image.jpg
                        geophys_raw_data_rendered_image.jpw

/dissemination/
        /{original_structure}
                gpr_raw_data1.segy
```

Documentation

Documentation will often be in text and image form, all such should be stored in a 'documentation' sub-folder, for example:

```
/preservation/
        /{original_structure}
                /documentation
                        geophys_raw_data_metadata.csv
```

In some cases, documentation can be split to accompany specific/relevant parts of the archive (i.e. it doesn't all need to go in one folder), for example:

```
/preservation/
        /{original_structure}
                /radar_survey_05
                        /documentation
                                radar_survey_05_metadata.csv
                /radar_survey_06
                        /documentation
                                radar_survey_06_metadata.csv
```