

## Section 2. The Life of Geophysical Data#

### 2.1 Introduction#

Although it may seem peculiar to consider a life-cycle of archaeological geophysical data, it has evolved as a useful metaphor to highlight the changes that geophysical data undergo subsequent to their collection in the field until they finally reach the Archive.

### 2.2 Data Acquisition#

Geophysical instruments are used to measure geophysical properties; a magnetometer, for example, measures the magnetic field (or some of its vector components) that is affected by buried structures. In a geophysical survey instruments are moved across the area of interest, either manually or by a vehicle (land, air or water based). Combining the measurements from all positions into a graphical map representation allows subsequent archaeological interpretation. Geophysical data acquisition therefore requires simultaneous recording of the instruments' measurement values and the location of the measurement. It is the latter that will be discussed here first.

#### 2.2.1 Gridded Data

The most common method of archaeological geophysical data acquisition relies on regular grids. They conceptually subdivide the survey area into a raster of many small rectangular cells of equal size (e.g. 0.25 m x 0.5 m) and a measurement is recorded for each cell. In practice this may be realised by walking along lines (e.g. 0.5 m apart) and recording at regular intervals (e.g. every 0.25 m). This geophysics grid then forms the 'geophysics coordinate system' to which all measurements are related, either by indicating line and point counts or by referring to it as a Cartesian coordinate system with perpendicular x- and y-axes. Importantly, the resolution of the grid (i.e. its subdivision into small cells) is fixed (Figure 1).

**Figure 1:** Geophysics coordinate system and grid resolution.

The relationship between positions in this geophysics grid and real location on the ground has to be clearly specified, as explained in the Appendix on Georeferencing Geophysical Data ([Appendix 2](#)), otherwise it is impossible to link identified geophysical anomalies to their location on the ground or to results derived from other sources (e.g. aerial photographs).

In addition to specifying the particular grid cell at which a measurement was taken the recording position within a grid has to be specified. For example for a 1 m x 1 m grid cell it is important whether the measurement was taken in the centre of the cell (i.e. 50% from the edge in x- and y-direction, Figure 2a) or whether it was recorded in the lower left corner (Figure 2b). Some practitioners for example prefer to walk over tapes that are laid out at regular intervals, which would usually mean that the recording is over the edge of a grid cell, others prefer to walk between tapes, resulting in measurements in the middle of the cell. Some instruments may have similar 'preferences': Geoscan magnetometers, for example, are designed such that they assist the operator with regular beeps for every metre walked, including start and finish (resulting in 21 beeps for a 20 m long line) but they actually record the measurements between these beeps in the centre of the grid cells. It is therefore important to note exactly where the readings are located in relation to the grid cells so that a good correlation between data and ground position can be achieved.

**Figure 2:** Two examples of recording positions.

It is the intention to always collect data in the exact grid position specified. However, this is seldom possible and the spatial measurement accuracy describes how close, on average, the measurement positions are to the locations indicated by the grid cell. It is useful to estimate this positional accuracy so that the reliability of spatial interpretation can be gauged. The final link between a recorded measurement and its true location on the ground is a combination of this positional accuracy and the georeferencing accuracy of the geophysics grid.

The major advantage of a regular grid is the even coverage of the ground in all locations, which makes comparisons between geophysical anomalies in different locations possible and avoids bias in data interpretation. If data are recorded with varying spatial resolution and then interpolated back to a regular grid (see below), the interpolation algorithm may introduce smooth edges or sharp corners where none really exist. Another advantage of recording on a regular grid is the ease with which the operator can locate their position, just by counting lines and recording points.

However, if a large area is to be surveyed, such counting also becomes difficult and it is therefore very useful to subdivide larger areas into smaller chunks (e.g. of 20 m x 20 m) that occupy only a limited space of the larger geophysics grid. These smaller chunks are given various names by different practitioners, for example by archaeological geophysics instrument manufacturers simply as 'grids', by remote sensing experts as 'tiles' and sometimes as 'sub grids'. In accordance with Aspinall et al. (2008) we use the term 'data grids' in this Guide. In addition to subdividing a large survey area into manageable chunks, which can then again be indexed through row and column counts data grids also have other advantages. Some magnetometers exhibit instrument drift which may not necessarily be constant over time. Monitoring such a drift for each individual data grid therefore allows more efficient processing and correction of this effect. Twin probe earth resistance measurements require the relocation of the remote electrodes when the cable length becomes insufficient for more distant recording positions. Undertaking such a survey over chunks of data grids makes these operations far more efficient. It should also not be underestimated that subdividing a large area greatly helps with the motivation of survey personnel ("only two more grids to go"). Data grids can be named according to the sequence in which they are recorded but if the same data grid is used with different instruments but in a different order (e.g. magnetometers and earth resistances meters) this may become confusing and naming by a combination of letters and numbers (similar to a computer spreadsheet) could be useful to reflect their position in the overall geophysics grid.

If data acquisition is undertaken over several data grids it is important to record this sequence, for example to monitor an instrument's drift and later correct for it. Obviously, it is essential to provide clear indications as to where each data grid is located in the overall geophysics grid. This can either be done through a map of the data grids, visualised as a mesh (also referred to as 'master grid') or by specifying the x- and y- coordinates of the corners of each data grid. Using this information all recorded measurements from all data grids can be assembled into one large data structure for the whole geophysics grid, often referred to as a 'composite' (i.e. data grids + mesh = composite, Figure 3).

**Figure 3:** Assembling data grids into a composite using a mesh.

### 2.2.2 Line Data

One alternative to the use of a regular grid is the continuous recording of readings while walking along straight lines. In the past systems were designed with analog and continuous recording on mechano-electrical strip charts (Clark 1996). Modern digital instruments have fast continuous digital recording of measurements (e.g. at a rate of 10Hz) or are triggered at selected spatial intervals using an odometer wheel or GNSS/GPS. These data can be represented as continuous line graphs, similar to a spreadsheet graph, or they are re-sampled to a grid with a resolution that is chosen to match the line data as well as possible.

### 2.2.3 Non-Gridded Data

Recording the measurements (geophysics data) as well as their location (e.g. using GNSS/GPS or a Total Station) during the survey allows to collect data without a predefined grid. If no further spatial guidance is used ('random walk') the resulting data density can be very uneven, covering some areas very densely while in other parts large gaps exist between the walked lines. Guidance systems are now available that show the already covered tracks so that the data collection can be somewhat adjusted.

The resulting data are usually re-sampled in X- and Y- direction to a regular grid so that grid-based processing software can be used for further data enhancement. The shortcomings of individual gridding algorithms should be carefully considered and the mechanisms for filling gaps between data tracks taken into account when interpreting the results.

A different approach is the direct use of these data without re-gridding. Sauerländer et al. (1999) demonstrated that representing such data as a two-dimensional Voronoi diagram is a very effective way of retaining their original spatial properties and visualising the density of data collection.

## **2.3 Storage in Instrument#**

### **2.3.1 Sequential**

The most basic and effective way of storing the measurement data in a geophysical instrument is as single measurement values in a long list that reflects the measurement sequence (e.g. in instruments from Geoscan and Bartington). Knowledge of sub grid dimensions (number of lines and their length) allows the correlation between each storage position in the list and the respective location in a data grid, leading to a display on the instrument of the current number for Grid, Line and Point, respectively.

### **2.3.2 Location Tagged**

Alternatively, measurement data and their location coordinates are stored together as a data pair for each measurement requiring more memory and a means to determine the location (e.g., in instruments by Geonics and Geometrics). If data are recorded on a regular grid the increment of 'positions' along a survey line for each subsequent measurement has to be specified (e.g. every 2 m) as well as the line increment (usually taken as being in the E/W direction). These increments are often defined in the geophysics coordinate system. If data are recorded along well defined lines with continuous sampling (e.g. at 10 Hz) markers ('fiducials') are usually inserted by the operator at specified intervals (e.g. every 5 m) and the collected data are subsequently re-sampled to a grid (see [Section 2.2](#)). In a non-gridded survey mode data acquisition is linked to location measurements from GNSS/GPS or a Total Station. This information can be stored with the geophysics data in real-time during the survey, requiring data transfer protocols to allow the communication of different instruments. Alternatively the geophysics and location data can be combined after the survey using a very accurate time stamp in both data sets which requires careful and repeated synchronisation of both instruments' internal clocks.

## **2.4 Download to and Storage on Computer#**

When the data are downloaded from the survey instrument to a computer, mostly through a serial/USB cable, the measurement data are saved together with other information about the survey, like time of survey, type of instrument, direction of traverse, zero drift, zigzag/parallel survey orientation etc. These form some of the survey's metadata (see [Section 3.3](#)). The data and metadata are usually stored in separate computer files, often in a proprietary format designed by the manufacturer of the hard- or software.

### **2.4.1 Sequential**

For instruments that store data as a long list of values on the survey instrument, the following steps are required during the data download.

- The download software needs to know the size of data grids. This information can either be entered by the user who knows in what size the data were originally collected (e.g. for Geoscan instruments) or by the instrument first sending this information to the downloading software to make it aware of it (e.g. for Bartington instruments).
- A stream of data is transferred to the computer, usually via a serial cable, when the 'Dump' procedure is initiated.
- On the computer, and usually after the full download, this stream is broken into appropriate data grids which are stored as smaller lists of data in individual files and related information (e.g. size of the data grids and their dimensions) are stored in metadata files to correlate each value in a data grid's file with the correct grid position.
- For line data (e.g. from a Scintrex or Geometrics magnetometer) the marker information is evaluated (e.g. "every 5 m") and the location for each recording is estimated and stored with the measurement value.

### 2.4.2 Location Tagged

If the location data are already stored in the geophysical instrument during the survey, the procedure is somewhat simpler.

- Measurement data and positions are transferred to the computer together.
- These data are stored in an unsorted list of coordinates and values, often referred to as 'XYZ text file'; this can also be in the form of a text file in which each line contains one data value, the survey line number, the reading number along the line, a marker indicator and a time stamp.
- The data are not broken into data grids at this stage.

If the position information comes from a non-integrated device (e.g. an auto-tracking Total Station), the data and position information have to be merged using the instruments' respective timestamp information. It is obviously important to know which coordinate system or datum is used for these coordinates.

## 2.5 Conversion to other Formats#

Each software normally uses a proprietary format to store data and metadata information and no common standard is currently available. For use of the data with other than the download software conversion is therefore required, which sometimes takes the form of exporting from one and importing into another package. When data are recorded on a regular grid it is often convenient to export the data either as a text file formatted like a spreadsheet so that each line of data values in the file corresponds with one survey line, or as a Z-file where the sequential list of stored data (see above) is written to a text file, one line at a time. When importing such data files into another software package all the metadata information has to be provided to the importing package, often by remembering the field procedure and entering this information manually. An alternative export format is as XYZ text files in which each text line contains the values of the coordinates (X and Y) as well as a measurement value ('Z'; see [Appendix 1](#)). The gridded layout of the data is thereby usually lost and the importing software therefore has to either re-grid the data (e.g. Surfer) or it has to estimate what the initial grid layout may have been. Re-gridding of data can also lead to small shifts in the data if the exact data position is unknown or was not used. For example if earth resistance data are recorded in the middle of each 1 m x 1 m grid cell, the first exported position should be at X=0.5, Y=0.5 and not at X=0, Y=0. Data and metadata should be stored with as much detail as possible to allow for conversion between different software systems. Geoscan's Geoplot and DWConsulting's Archaeosurveyor are exemplary in this respect and save a considerable amount of metadata with the respective measurement information.

## 2.6 Assembling data into Composites#

If a geophysical survey is split into smaller data grids (see above) their corresponding data files have to be re-assembled into an overall data entity. To achieve this a 'mesh' is used that indicates the relative position of each data grid in the overall survey area. It is hence a map of the data grid layout but does not contain measurement data itself. It often takes the form of a spreadsheet table and is also referred to as 'master grid' (Geoplot). Using this mesh the data grids are assembled to form a 'composite' for the whole survey area. This will also take the form of a gridded data structure with associated metadata but for a much larger area and allows data processing irrespective of the arbitrary data grid boundaries that were introduced during the survey.

## 2.7 Processing#

A variety of computer algorithms are available to improve and enhance the measured geophysical data and they are often all referred to as data processing. However, since some require earlier steps and different metadata it is conceptually clearer if they are subdivided in steps for data improvement, data processing and image processing.

### 2.7.1 Data Improvement

If information on the process of data acquisition is available, some common surveying errors can be rectified with appropriate software. For example, if data were collected on adjacent lines, drift of an instrument's reference value can often be reduced. Such processing can however only be applied if data are available in a format that reflects the surveying procedure (e.g. if data are saved as separate data grids) and if information on the surveying strategy is given. Data improvement usually is applied to individual data grids as these are the units in which data were collected and hence delimit defects during survey. Typical examples include drift correction, destaggering, destriping (see Figure 4) and grid balancing.

**Figure 4:** Magnetometer data before and after applying destriping data improvement.

### 2.7.2 Data Processing

Once all data grids have been improved and a seamless composite produced this can be processed to deal with effects of the data and not with defects created during their acquisition. For example, if a very high reading produced by metal debris (fallen from a tractor) creates a magnetometer 'spike' across two data grids, the despiking procedure has to be applied after the grids were corrected for drift and been balanced. It is also possible to apply processing algorithms that are pertinent to the geophysical nature of the acquired data, for example using high-pass filtering for earth resistance data over a geological background variation (see Figure 5) or computing a reduction-to-the-pole for magnetometer data.

**Figure 5:** Earth resistance data before and after application of a high-pass filter.

### 2.7.3 Image Processing

Data are just numbers and have to be visualised to analyse and interpret them. In the past this often took the form of contour diagrams but now greyscale raster images are the most commonly used method. To do this a 'transfer function' is used that determines how each data value is converted to a display colour. In this process very high and low values are usually 'clipped' (i.e. mapped to the same display colour) and overall the process is irreversible. Once the image pixels are calculated, the real measurement data cannot be reconstructed. This creates problems if subtle information is to be extracted as the range of for example display grey shades (e.g. 0 to 255) is much coarser than the full data range (e.g. -20.05 to +500.07 nT). When data are converted to images for display or

file storage important information is therefore lost. Images are extremely useful for evaluating data but they are no replacement for the recorded measurements. These images can only be treated with standard image processing tools that cannot take into account the geophysical nature of the underlying data (see the [Guide to Raster Images](#)). Although of little use for geophysical analysis of the data, image processing functions in standard image software (e.g. embossing, see Figure 6) may still be useful to convey relevant information to users.

**Figure 6:** Image of magnetometer data before and after embossing.

## 2.8 Georeferencing and data combination#

To compare geophysical data with other data sources (e.g. aerial photographs, earthwork surveys, adjacent excavations) or data recorded on different alignments, all data have to be converted to a common map coordinate system and datum (e.g. WGS84, OS Coordinates, site grids). This requires detailed spatial information for all data sources.

Geophysical data that were collected on regular grids or along lines are usually defined in a rectangular and homogeneous coordinate system, the geophysics coordinate system, which is defined irrespective of any topographic changes. To achieve the best possible fit with other data that are usually projected onto a horizontal map plane they may have to be warped to adjust for topographic distortions. However, in many cases a simple restricted affine transformation (shifting, rotating, but no stretching) is sufficient. The accuracy of the georeferencing determines the accuracy with which a particular geophysical anomaly can be located on the ground, for example for subsequent excavation. It is important to remember the difference between the mathematically defined rectangular geophysics grid coordinates (lines and measurement points) and the projection and warping onto physical map coordinates (see Figure 7). More information on coordinate systems and coregistration is provided in [Appendix 2](#).

**Figure 7:** Mathematical grid coordinates and actual warped layout.

## 2.9 Interpretation#

Geophysical data require interpretation to make them more widely accessible. To assign 'meaning' to a very large assemblage of data points (a 1 ha square of magnetometer data recorded at 0.25 m x 1 m resolution already contains 40,000 measurements) an interpreter needs to understand the nature of geophysical data and their possible causes, as well as the archaeological context and the resulting restrictions on reasonable feature attributions. Data interpretation is hence subjective and the caveats of postprocessual theory apply. It may be tempting to escape this uncertainty by restricting data interpretation to the delineation of geophysical anomalies, hoping that somebody else will make a judgment about their meaning. However, such a separation of geophysical and archaeological analysis will not harness the full potential of the carefully collected data. The archaeological interpretation of geophysical data may change over time as new information is revealed elsewhere and it is hence necessary to keep the measured data for possible future re-interpretation (see [Section 3.1](#)).

Data interpretation should be accompanied by clear diagrams that highlight relevant geophysical anomalies, possibly by delineating them with an outline polygon. They should be annotated with labels that can be used in the textual description; simply referring to an anomaly as "the strong magnetic anomaly in the north of the survey area" is far too ambiguous. Many different labelling schemes are currently in use and have their individual merits. They may even include site numbers and indicators of the geophysical method (e.g. [1m3]) to make them unambiguous in the report of a large survey area.

The textual description and interpretation can follow two different styles. The first style has two clearly separated stages, starting with a geophysical interpretation of anomalies according to the geophysical characteristics of the data (e.g. "Earth resistance anomaly [r1] could be caused by a confined low resistivity feature of ca. 2 m diameter at approximately 1.5 m depth.", or "The positive magnetic anomaly [m2] is very broad and appears to be caused by a deep feature."). The second stage is then the identification of these geophysical anomalies as archaeological features (e.g. "Anomaly [r1] can be interpreted as a buried pit of dimension..."). This requires prior knowledge about the site, its geology and related archaeology. Also the overall spatial morphology (square, rectangular etc.) and extent of anomalies may help with their archaeological interpretation in line with similar approaches found in aerial photographic interpretation. The second textual style combines the geophysical and archaeological interpretation for each feature and may often omit the geophysical analysis. This is far easier to read, especially for a busy reader who just wants to know "what has been found", but it is more difficult to follow for experts who may wish to understand why a particular interpretation was made.

## 2.10 Reporting#

All geophysical surveys should aim for a report that details the data, their processing and interpretation. Reports should be written in accordance with accepted guidelines. In Britain, it is recommended to use those laid out by English Heritage (David et al. 2008). A report will very rarely contain all the information that has been collected and analysed in a geophysical survey project. It is hence essential to understand that although the report is an important product derived from the geophysical work, it is not the sole outcome of an investigation and is not sufficient on its own to form the Archive.

## 2.11 Archiving#

This is the last stage of one cycle of geophysical data usage, before they are re-evaluated and re-used in their next life-cycle. All geophysical data and associated and supporting information (for example including spatial references) are taken together and form the Archive. Relevant field notes and paper records can nowadays easily be converted to electronic documents and the Archive can therefore be entirely digital. This will be discussed in more detail in the rest of this Guide. If possible, keeping the paper records in a physical archive is a welcome additional backup in case of unforeseen data emergencies.

Archiving can be considered as the combination of creating the Archive and depositing it with an Archiving Body. In the simplest case the Archiving Body may be the geophysical contractor or academic researcher themselves but it could also be a dedicated and specialised archiving service like the ADS (Archaeology Data Service) or tDAR (the Digital Archaeological Record). Exactly what happens to the Archive depends on the Archiving Body and scenarios range from the extremely risky "DVD in a drawer" to distributed servers with ongoing refreshing and migration of files (see [Section 3.4](#)), possibly making the data accessible online.

There is no point keeping data if they cannot be used again. Although this statement seems trivial, it has crucial consequences. Archiving is meaningless unless data files are made available in formats that can still be used in a few years time and are accompanied by documentation that allows them to be put in a meaningful archaeological context (e.g. describing how GPR transects were collected so that they can be assembled into timeslices). Looking at one's own data from one year ago is usually a good reminder why detailed documentation is essential. The creation of the necessary metadata and data documentation is covered in subsequent chapters.