

## Section 3: Archiving GIS Datasets#

### 3.3 Data-specific Documentation and Metadata#

The sections below summarise the additional documentation and metadata recommended for specific datatypes and processes as outlined in Section 2. Data creators are advised to see the relevant pages under Section 2 of this guide.

- 3.3.1 The Vector model
- 3.3.2 The Raster model
- 3.3.3 Attribute Data Models
- 3.3.4 Data Capture using a Scanner
- 3.3.5 Data Capture using a Digitiser
- 3.3.6 Data Capture using a Scanning-digitising Hybrid
- 3.3.7 Common Sources of Spatial Data
- 3.3.8 Common Sources of Attribute Data
- 3.3.9 Maps and Plans
- 3.3.10 Textual and Numeric Data
- 3.3.11 Purchased or Downloaded Digital Data
- 3.3.12 Aerial Photography
- 3.3.13 Satellite and Airborne Remotely Sensed Images
- 3.3.14 Terrestrial Survey
- 3.3.15 Satellite-based Survey (GPS)
- 3.3.16 Creating a GIS Database
- 3.3.17 Documenting the Dataset

#### 3.3.1 The Vector model

The following information should always be recorded when assembling, compiling and utilising vector data:

- The data type, Point, Line or Polygon
- Type of topology which the file contains
- Details of any automatic vector processing applied to the theme
- State of the topology in the file
- Projection system
- Co-ordinate system

#### 3.3.2 The Raster model

The following information should always be recorded when assembling, compiling and utilising raster data:

- grid size (number of rows and columns)
- grid resolution
- georeferencing information, e.g. corner co-ordinates, source projection.

#### 3.3.3 Attribute Data Models

When attempting to structure and organise a flexible attribute database the following factors are of critical importance:

- Naming conventions
- Key fields
- Character field definitions
- Grid references
- Validation
- Numeric data
- Data entry control
- Confidence values
- Consistency
- Documentation
- Dates

### **3.3.4 Data Capture using a Scanner**

- Details of the scanning device used, software driver and version
- All parameters chosen in the scanning process, such as the resolution setting of the device, the number of bits per pixel used
- Details of any pre-processing undertaken on the source mapsheet. This may include a range of options provided by the specific scanning software used
- Details of any post-processing undertaken on the data, such as noise reduction or sharpening with convolution filters, histogram equalisation, contrast adjustment

### **3.3.5 Data Capture using a Digitiser**

- Detail of the digitising device used, software driver and version
- The precision, usually specified as a quoted resolution or as lpi
- Details of any automatic vector processing applied to the theme (such as snap-to-nearest-node)
- Details of control points used to manage conversion from digitiser to real-world planar co-ordinate systems
- Errors incurred in the above transformation process (e.g. quoted RMS)

### **3.3.6 Data Capture using a Scanning-digitising Hybrid**

In the case of using both a scanner and a digitiser to capture data, for example during 'heads-up digitising', the full information above for both the digitising and scanning procedures should be recorded.

### **3.3.7 Common Sources of Spatial Data**

- Maps and Plans
- Textual and numeric data
- Purchased or downloaded digital data
- Aerial photography
- Satellite and airborne remotely sensed images
- Terrestrial Survey data
- Satellite-based (GPS) data

### **3.3.8 Common Sources of Attribute Data**

Below are some likely sources of attribute data which you may come across, and wish to re-use:

- paper based card indexes
- archaeological site and survey archives (including paper based records, finds databases)
- qualitative report texts and articles published in journals (paper based or on the Internet)
- microfiche archives
- geophysical interpretation data derived from interpreted geophysics plots

- aerial photograph interpretations which may include morphological analysis attribute data and photo source information
- typological databases or artefact type series
- data generated at a regional level for integrated large scale historic landscape studies, such as the English Heritage Open Fields Project
- local level archaeological databases (e.g. Sites and Monuments Records or Urban Archaeological Databases where they are held separately from SMRs)
- local museum site and finds databases
- local Record Offices
- national archaeological databases (e.g. such as the various UK National Monument Records or English Heritage's database of Scheduled Ancient Monuments)
- Gardens Trust surveys
- historic buildings surveys and databases maintained by local authorities
- metadata relating to data sets

### **3.3.9 Maps and Plans**

In general the following information should always be recorded:

- Publisher **and** copyright owner
- The map medium
- Scale of source map
- Name of the map and the map series
- Claimed accuracy for any specific map components
- All details of the map projection and co-ordinate system employed

### **3.3.10 Textual and Numeric Data**

When integrating textual and numeric data the following information should be recorded:

- The data source
- The precision of the quoted co-ordinates
- Have the quoted locations been verified and how
- Projection system/co-ordinate origin
- If derived from a source map, record details of the map-base used
- If derived from a survey programme, record details of the survey procedure

### **3.3.11 Purchased or Downloaded Digital Data**

Several standard formats and standards are of interest.

- British Standard 7567 (NTF: National Transfer Format), the format used by Ordnance Survey for the supply and transfer of digital products
- The recommendations of the National Geospatial Data Framework (NGDF)
- SDTS (Spatial Data Transfer Standard), a United States Federal Information Processing Standard (FIPS)
- DLG (Digital Line Graph) format, used by the USGS for supply of vector information
- DRG (Digital Raster Graphics), is the description that the USGS gives for the distribution of scanned map sheets
- DXF (Digital eXchange Format) format, commonly used for transferring drawings between CAD (Computer Aided Design) systems

### **3.3.12 Aerial Photography**

To incorporate scanned and rectified aerial photographs into GIS databases the following information should be recorded:

- Full Photographic details
- Details of the scanning process
- Details of the rectification method(s) used
- The software employed including, where possible, specific parameters chosen
- Details regarding the ground control points (GCPs) used during the procedure
- Details of any post-processing undertaken on the data

### 3.3.13 Satellite and Airborne Remotely Sensed Images

To incorporate remotely sensed data into GIS databases the following information should be recorded:

- Data source
- Date image was captured
- Data resolution
- Details of any post-processing undertaken on the data
- Details of the rectification method(s) used
- The software employed including, where possible, specific parameters chosen
- Details regarding the ground control points (GCPs) used during the procedure

### 3.3.14 Terrestrial Survey

When integrating data themes which are derived from survey data, the following should be recorded:

- The source and estimated error of survey base station co-ordinates
- Details of the survey, including date time and purpose
- Details of the thematic organisation of the survey
- Make and model of instrument used
- Type of survey (contour, feature etc.)
- Estimated error terms for the co-ordinate pairs and (if appropriate) the z-co-ordinate
- Georeferencing information, overall accuracy of the survey data

### 3.3.15 Satellite-based Survey (GPS)

In integrating GPS data the following information should be recorded:

- The method used to locate stations: C/A or P code pseudorange measurements, carrier phase measurements and whether a single measurement or averaging (include time period) was used
- The software used for any co-ordinate transformation and associated error estimate
- The satellites used in obtaining fix and observed GDOP (Geometric Dilution of Precision)
- The nature of any differential correction undertaken + error estimates
- The broadcast differential: name of the service provider and the name and location of base station
- The local base station: instrument details, location (including error estimate) of base station
- Post-processing: the software used and the source of correction data

### 3.3.16 Creating a GIS Database

When combining and integrating information from a variety of sources the following points should be kept in mind:

- All spatial data must be recorded in the same co-ordinate system. Data which are recorded to some other system must be transformed/projected to the required co-ordinate system.
- All spatial data should be to the *same spatial* resolution, or scale. It is not possible to get meaningful results from the combination of spatial data recorded to a *scale* of 1:250, as might be the case for an excavation site plan, with road alignments recorded to a scale of

1:250,000. Spatial data recorded to *scales* of greater than around 1:10000 involve considerable generalisation of alignments to avoid features conflicting. Thi

- Non-spatial information to be combined, or integrated, must use the same field definitions, encoding regimes, etc. Where different schemes are used it will be necessary to convert or translate the data to the required scheme.

### 3.3.17 Documenting the Dataset

Information about where the data you use are acquired from is one of the most important things you can record whilst constructing and using a GIS. The following comprises a *non-exhaustive list* of the information you might wish to record during your everyday creation, collection, and use of data:

- Computer hardware used
- Computer software used
- Date the data were captured/purchased/whatever
- Who did the work
- Data source ('bought from Ordnance Survey', etc.)
- Scale/resolution of data capture
- Scale/resolution at which data are currently stored
- Root Mean Squared error or other assessments of data quality
- Purpose of data set creation, where known
- Method of original data capture (Total Station Survey, etc.)
- Purpose for which *you* acquired the data (might differ from the previous information where the data were *created* by someone else for one purpose, and bought from them by you for another)
- Complete history of data ownership/rights.