Guidelines for English Heritage projects involving GIS

Background

As modernisation occurs and GIS becomes available to all members of English Heritage staff it becomes essential that the quality of data received from externally commissioned projects is known and a minimum standard for acceptable data is set.

This document therefore gives basic guidelines for projects involving GIS to ensure that English Heritage can gain maximum benefits from the data which is generated. Any further clarification / information can be found by contacting Alice Froggatt – Corporate GIS Co-ordinator (alice.froggatt@english-heritage.org.uk, 020 7973 3095).

Use of Base Mapping

As a member of the Ordnance Survey Pan-Government Agreement, English Heritage has an organisation wide license to use a wide range of Ordnance Survey digital data. Depending on the nature of the project, it may be possibly to supply base mapping for use in the project. Similarly English Heritage has a license to use the LandMark Information Group Historical (Pre-war) OS digital map data and can, in certain circumstances, also supply this for use. Further details regarding the use and supply of these products can be obtained from David Gander - webGIS Manager and EH OS Liason Officer (david.gander@english-heritage.org.uk, 020 7973 3094).

English Heritage recommends that modern OS map products are used for data capture wherever possible. If sources of data other than OS are used to capture data please make sure it is fit for purpose. Are the data collection standards known and acceptable? Is the data used to capture the dataset a suitable age? Does the data cover the whole area? What happens at boundaries? Do you have a license to use the data? Full details of data used for data capture should be recorded in the project methodology and metadata.

Remember: Garbage In – Garbage out!
The most advanced automation procedures will not improve an inferior map base.

File formats

We would prefer data to be supplied in the following formats:
(N.B. This list will increase with time as EH GIS develops.)

Vector Data	*Raster*
ESRI Geodatabase	Uncompressed tiffs with ESRI world files
ESRI Shapefile	ESRI Grid or Tin

If you wish to supply EH with a different type of file format please contact Alice Froggatt to discuss it's interoperability with EH's Corporate GIS.

Data Capture

There are a number of important points to consider when capturing spatial data.

Methodology

It is important that the methodology of data capture is carefully recorded; this includes details of pre-processing, methods of geo-coding and classification. It should also include decisions and assumptions made in data capture – e.g. all features were digitised along the centre line of source map features and it was assumed that all point features were recorded within an accuracy of 100m.

Spatial Data Types

Think carefully when deciding to capture a vector dataset as to which data type is most appropriate.

A **point** is the simplest graphical representation of an object. Points have no dimensions. The scale of viewing can determine whether an object is defined as a point or a polygon. In a large scale representation a building may be shown as a polygon whereas it may only be a point (symbol) if the scale is reduced. **Lines** connect at least two points and are used to represent objects which may be defined in one dimension. These one-dimensional objects may be roads, rivers, regional boundaries, fences or any kind of object that is fundamentally long and very skinny. Roads and rivers, on the other hand, may be either lines or areas, depending on the scale. **Polygons** are used to represent objects defined in two dimensions. Adding the dimension of height to our area features allows us to observe and record the existence of surfaces.

Non-spatial information about an object is recorded in the *attributes*.

Recording Attributes

- There is no limit to the number of attributes a dataset can store.
- Wherever possible do not leave attributes as NULL values. There should certainly be no fields where each feature has a NULL.
- All field names should be in BLOCK CAPITALS
- Field names should not contain spaces. USE_UNDERSCORES_INSTEAD.
- Each spatial feature must be assigned a unique reference code.
- Avoid using abbreviations or punctuation within names, unless they form part of the official name.
- Ensure all attribute names and descriptions have been spelt correctly.
- Code and format attributes consistently.
- Document any codes, acronyms or shorthand terms.
- Wherever possible use already established thesauri, stating in the project metadata which thesauri have been used.
- Use appropriate field types – e.g. Dates should be Date type.

Topology

There should not be any duplicate features.

All polygon data sets should be topologically clean and correct. Topology defines the spatial relationships between features.

- All polygon boundaries should be closed i.e. they must contain coincident start and finish vertices.
- Excluded areas within polygons must be voids, although accompanying "island" features may be separately identified.
- Each arc has a beginning and ending node (this defines direction).
- Arcs connect to other arcs at nodes (this defines connection).
- Connected arcs that surround an area form polygons (this defines area and perimeter).
- Arcs have a right and left side (this defines adjacency). Every feature (point, arc, polygon, region, etc) should have at least one attribute record.
- There should be no slivers, dangles, knots or cross-overs.
- Multi-part polygons are acceptable; they should have one set of attributes associated with them.

Error

The problem of error devolves from one of the greatest strengths of GIS. GIS gain much of their power from being able to collate and cross-reference many types of data by location. They are particularly useful because they can integrate many discrete datasets within a single system. Unfortunately, every time a new dataset is imported, the GIS also inherits any errors. These may combine and mix with the errors already in the database in unpredictable ways. The key point is that even though error can disrupt GIS analyses, there are ways to keep error to a minimum through careful planning and methods for estimating its effects on GIS solutions.

Spatial Resolution

Spatial resolution refers to the area on the ground that data layer can identify.

Vector Data
Generally a line can not be drawn much narrower than about 0.5mm. Therefore, on a 1:20,000 scale map the minimum distance that can be represented (i.e. spatial resolution) is about 10 metres. As a general rule therefore, if the scale of capture is known then the spatial resolution can be estimated by dividing the scale of capture by 2000.

Scale of Capture	Spatial Resolution (metres)
1:10,000	5
1:25,000	13
1:50,000	25
1:100,000	50
1:250,000	125
1:625,000	312

Raster Data

For raster data derived from vector format the size of the pixels should not be smaller than the uncertainty in the vector. For raster data derived from Air-photo / satellite imagery, pixels should not be smaller than the resolution of the camera that recorded it.

It is important to make sure that the spatial resolution of the final dataset is not greater than those used to create the dataset. It is not sensible, for example, to digitise data from a 1:50,000 map and present it on a 1:2,500 scale map even if the GIS display is zoomed to 1:2,500 whilst digitising. For example, a dataset digitised from a 1:1,250 vector dataset and 1:10,000 raster map can not be used in a context greater than 1:10,000. Zooming in on a small scale data set does not increase its level of accuracy or detail. It is vital therefore that before data capture starts you consider what map scale the dataset is intended to be used at and what scale the source map is.

Some spatial operations may result in features which are smaller than the spatial resolution. For example, overlaying polygons may create 'slivers' which are, say, ten metres wide when the resolution of the data is 20 metres. These slivers should be included with their neighbours before the results of the overlay are used for further analysis. It is important to realize, however, that precise data--no matter how carefully measured--may be inaccurate.

Data precision

Data precision is the smallest difference between adjacent positions that can be recorded and stored. The GIS will carry much more precision through its calculations than are justified by the data's accuracy. The

results of these calculations should be rounded to a value appropriate to the uncertainty of the data for reporting.

Accuracy

Accuracy is the degree to which information on a map or in a digital database matches true or accepted values. In a GIS, it is necessary to consider horizontal and vertical accuracy with respect to geographic position, as well as attribute, conceptual, and logical accuracy. It is important to note that the level of accuracy required will vary from project to project.

Cascading

Cascading means that erroneous, imprecise, and inaccurate information will skew a GIS solution when information is combined selectively into new datasets. In a sense, cascading occurs when errors are allowed to propagate unchecked from layer to layer repeatedly.

Note, however, it would be a mistake to believe that highly accurate and highly precise information is needed for every GIS application. The need for accuracy and precision will vary radically depending on the type of information coded and the level of measurement needed for a particular project. Inaccurate and imprecise data isn't useless as long as it is correctly documented.

Projections

For projects based in terrestrial England, all data should be in the OSGB 1936 projection. For maritime projects GIS data can be in WGS84.

Units of measurement

All measurements should be metric. The metadata should state what the unit of measurement is for each field.

Metadata

Metadata is "data about data". It is used to describe the content, quality, condition, and other characteristics of data.

An EH GIS Metadata standard will be released shortly. In the meantime all GIS datasets should be supplied with UKGEMINI metadata. In addition data should be supplied with Exploration Metadata data for each GIS layer – that is a definition of all attributes.

Metadata Element	Definition
Field Name	Field name as it is in the GIS Data Layer
Field Alias	Alternative Name for Field Name
Field Definition	What the field contains, including any units of measurement
Field Type	Field Type e.g. String, Double
Field Width	Field Width

Data Delivery
Data should be supplied on CDROM and include documentation describing the files contained on the CDROM.