



Preservation Handbook

Geographical Information Systems (GIS)

Author	Jo Clarke and Jenny Mitcham
Version	5 th draft v0.5
Date	June 2005
Change History	23 rd June: changes made in response to AHDS review



Definition

Geographical Information Systems provide a set of tools for capturing, storing, checking, integrating, manipulating, analysing and displaying data related to positions on the Earth's surface. Typically, a GIS is used for handling maps of one kind or another. These might be represented as several different layers where each layer holds data about a particular kind of feature (e.g. rivers, roads, geology, archaeological sites). Each feature is linked to a position on the graphical image of a map. As well as being spatially referenced, layers can also hold any number of other attributes, for example a layer relating to archaeological sites could include a name plus site type and period attributes for each individual site.

GIS can also be described as a set of processes, executed on raw data, to produce information which will be useful in decision-making

- A chain of steps leads from observation and collection of data through analysis
- An information system must have a full range of functions to achieve its purpose, including observation, measurement, description, explanation, forecasting, decision-making

It uses geographically referenced data as well as non-spatial data and includes operations which support spatial analysis

- In GIS, the common purpose is decision-making, for managing use of land, resources, transportation, retailing, oceans or any spatially distributed entities
- The connection between the elements of the system is geography, e.g. location, proximity, spatial distribution
- In this context GIS can be seen as a system of hardware, software and procedures designed to support the capture, management, manipulation, analysis, modelling and display of spatially-referenced data for solving complex planning and management problems

Although many other computer programs can use spatial data (e.g. AutoCAD and statistics packages), GIS includes the additional ability to perform spatial operations.

Additional Information

- What is GIS? – GIS.com
< <http://www.gis.com/whatisgis/index.html> > Last checked 25/02/2005
- GIS Guide to Good Practice - ADS
< <http://ads.ahds.ac.uk/project/goodguides/gis/> > Last checked 25/02/2005



Technical Environment

Geographical Information systems have been designed for most operating systems, including Windows, Unix and Linux. It is possible to accession most major GIS file formats, though ESRI formats are preferred as ESRI software is used internally to manage and migrate GIS data sets See ESRI website for description of products (<http://www.esri.com/products.html>). For preservation, the ArcInfo export format (.e00) is used.

Common File Formats

File extensions	Software	Notes
.e00	ESRI ArcInfo export	ArcInfo coverage export format <i>Preferred delivery and preservation format</i>
.poi and .lin and .pol and .txt	ESRI ArcInfo Ungenerate	A fairly simple means of transferring data to other GIS systems in an ASCII file format. ASCII files are of point, line or polygon data. The text file contains : delimited attribute data. <i>Acceptable delivery format</i>
.shp .shx .dbf .sbn and .sbx .fbn and .fbx .ain and .aih .prj .xml	ESRI ArcView	SHP - the file that stores the feature geometry. Required. SHX - the file that stores the index of the feature geometry. Required. DBF - the dBASE file that stores the attribute information of features. Required. SBN,SBX - the files that store the spatial index of the features. Optional. FBN, FBX - the files that store the spatial index of the features for shapefiles that are read-only. Optional. AIN, AIH - the files that store the attribute index of the active fields in a table or a theme's attribute table. Optional. PRJ - the file that stores the coordinate system information. Optional. XML - metadata for ArcInfo 8.0, for using shapefiles on the Internet. Optional. <i>Preferred delivery and dissemination format</i>
.mif .mid	MapInfo Interchange format	MIF – contains the graphics MID – contains textual data (optional) <i>Acceptable delivery format</i>
.tab .dat .map .id	MapInfo	Standard MapInfo file formats. <i>Problematic delivery format – ask depositor to supply MIF/MID or use Universal Translator to convert to shp file</i>
.ntf	National Transfer format	Ordnance Survey format (to be replaced with Geographical Mark up Language - GML) <i>Problematic delivery format</i>
.ddf	Spatial data transfer standard	SDTS is a large standard composed of smaller, more limited subsets that are Federally approved as part of the SDTS FIPS 173



		standard. <i>Acceptable delivery format</i>
.exp	Map Overlay Statistical System	MOSS export files contain polygon data extracted from the U.S. Department of Interior's MOSS public domain GIS. These consist of points, lines, or closed polygon loops (possibly with islands), and a 30-character attribute field referred to as the subject value. <i>Acceptable delivery format</i>
.vpf	Vector Product Format	VPF is a digital geographic vector-based format used by the U.S. Defense Mapping Agency for the distribution of its vector data sets. <i>Acceptable delivery format</i>

Other GIS software exists, notably the raster packages GRASS and Idrisi. Native file formats from these software packages are not acceptable deposit formats. Depositors should export directly into one of the acceptable delivery formats listed above.



Ingest Checklist

Level 1 (Essential)

- General purpose of the GIS
- Coverages contained within the GIS
- Creator of the GIS
- Complete history of data ownerships and copyrights
- Co-ordinate system/projection used
- For each layer or coverage:
 - o Method of original data capture (Total Station Survey, etc.)
 - o Data source ('bought from Ordnance Survey', etc.)
 - o Scale/resolution of data capture
 - o Scale/resolution at which data are currently stored
 - o Root Mean Square error or other assessments of data quality
 - o Date the data were captured/purchased

Level 2 (Preferred)

- The computer hardware used
- The computer software used
- Purpose of data set creation, where known
- 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)

Level 3 (Best Practice)

- The date of any change/modification
- The reason for any change/modification
- The record numbers affected by the change
- Relationships to other resources; where, for example, you derive a new GIS data set by passing a mathematical filter or some other modification through an existing data set, you may wish to record the relationship formally between the original data and the new set

Inform Depositor

- How data will be stored
- That data lacking documentation may be difficult to preserve.



Preservation

Significant Characteristics

The significant characteristics of a GIS are:

- The spatial description of each feature on each layer (ie: as a series of xy co-ordinates)
- Content and purpose of each layer or coverage
- Format of that coverage (point, line, polygon, grid etc)
- Association of any attribute data with its spatial information
- Scale, accuracy, data source and data capture techniques
- Co-ordinate system used

Technique

The method used to check and archive GIS data depends on the file format in which the data is supplied and the GIS software you have available. The general principle is to import the files into a GIS, check that the files are not corrupt and that they contain the data that matches the documentation and then store them in a suitable archive format. The archive format also depends upon the delivery format.

The acceptable file formats can be grouped by import/archiving technique.

1. Formats which are archive formats (e00)
2. Other formats

Formats which are archive formats

1. Import into ArcInfo or ArcView (ArcView needs import71 function for e00 files) and check each layer against documentation.
2. Archive the delivered files with documentation

Other formats

1. Import into ArcInfo and check each layer
2. Export as e00 files

ArcInfo readable files:

Format	ArcInfo Import Command
ArcInfo export (e00)	IMPORT
Shapefile	SHAPEARC
Spatial data transfer standard (SDTS)	SDTSIMPORT
MOSS	MOSSARC
Vector Product Format (VPF)	VPFIMPORT



ArcView readable files:

Format

ArcInfo export (e00)

Shapefile

MapInfo Interchange Format (MIF) (using MIFSHAPE conversion utility)

Vector Product Format (VPF)

Validation of Exported Data

Validation of files can be difficult without having some idea of what the original coverages looked like. It can be helpful if the depositor supplies paper print-outs of the coverages so the created file can be visually compared. It is also important to check that the attribute data has translated correctly where appropriate. This is something that can be more easily forgotten but is essential to the interpretation of the layers and thus the future usability of the data. Check that field names are not truncated and that floating point numbers are not cut off at the decimal point.

Problems and Issues

- Lack of documentation causes difficulties in assessing the data.
- Documentation often missing:
 - Co-ordinate system used
 - Description of contents of each layer/coverage
 - Method of original data capture (Total Station Survey, etc.)
 - Data source ('bought from Ordnance Survey', etc.)
 - Scale/resolution of data capture
 - Root Mean Square Error or other assessments of data quality