
Time to move on?

Archaeology's lingering love for the Shapefile

Kieron Niven

30th March 2022



Overview

- Geospatial data at ADS
- What do we get, and what do we do with it
- Metadata we collect
- “What’s the problem”? - Wider developments

Not going to look at...

- Raster data
 - Survey data such as lidar. Specifically look at vector feature data.
-

Geospatial data at ADS

Geospatial data deposited
since 1999

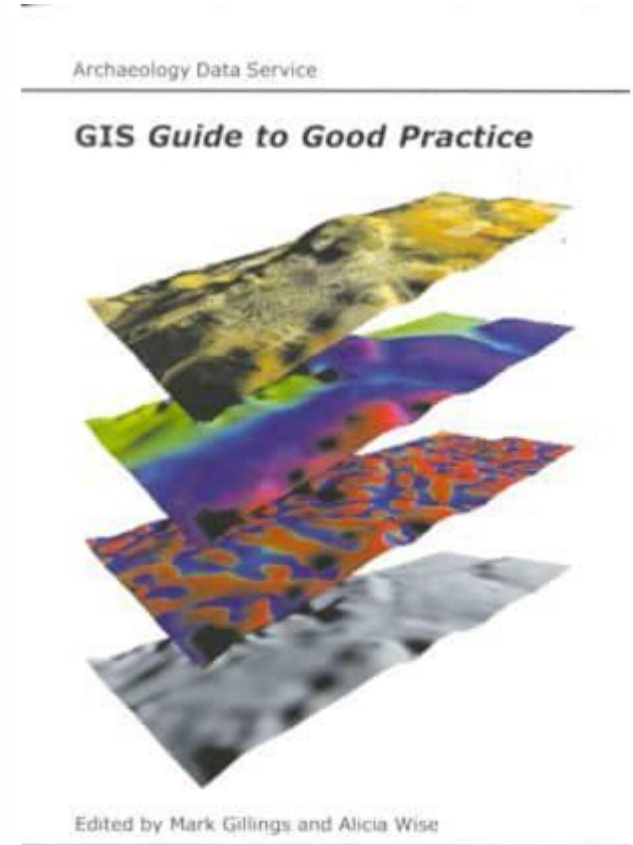
MOLA Royal Opera House
excavations



Geospatial data at ADS

One of the earliest Guides to Good Practice - 1999 on GIS and geospatial data.

ESRI-dominant approach in terms of formats



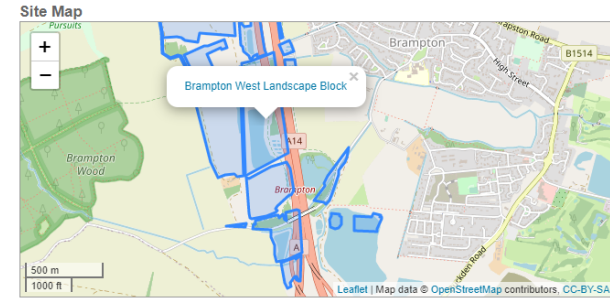
Geospatial data at ADS

Around 5000 GIS files currently in the ADS archive

From a variety of sources and of various types:

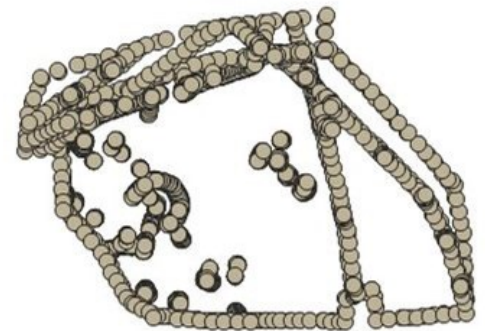
- site-based commercial / development data (common)
- large scale infrastructure datasets (e.g. road development schemes)
- through to thematic research datasets (such burial databases etc.)

Size of the datasets themselves can vary massively from a few kB to much larger.



Site List

- A14 Cambridge to Huntingdon, Cambridgeshire; Alconbury Landscape Block
- A14 Cambridge to Huntingdon, Cambridgeshire; Bar Hill Landscape Block
- A14 Cambridge to Huntingdon, Cambridgeshire; Brampton South Landscape Block
- A14 Cambridge to Huntingdon, Cambridgeshire; Brampton West Landscape Block
- A14 Cambridge to Huntingdon, Cambridgeshire; Conington Landscape Block
- A14 Cambridge to Huntingdon, Cambridgeshire; Fenstanton Gravels Landscape Block
- A14 Cambridge to Huntingdon, Cambridgeshire; River Great Ouse Landscape Block
- A14 Cambridge to Huntingdon, Cambridgeshire; West of Ouse Landscape Block



Geospatial data at ADS: Files & Formats

Approach largely unchanged.

Types of data we ingest largely unchanged.

Formats:

Shapefile and **GML** for deposit
Preserve as GML
Disseminating as Shapefile (zipped)

Worth noting that we've never had data deposited as GML.

Geospatial data at ADS: Metadata

Metadata - a key consideration from the outset.

GIS: Guide to Good Practice (1999):

- Clear sections on metadata
- Largely DC focussed core metadata...
- ...with additional documentation specified for sources, accuracy, etc. and ancillary metadata.

DC Terms



Additional
Documentation

Geospatial data at ADS: Current Metadata

Approach, again, is largely unchanged.

Most recently updated metadata templates continue along the same lines:

Integration of DC terms
data-type specific elements

Largely parallel external
developments i.e
ISO > INSPIRE > UK GEMINI

TERM	MEANING
File name	This is the name of the file.
Title	A short title for each set of GIS files which should describe what it depicts.
Description	A longer description for the GIS which should describe what is depicted in more detail and will be used by those searching your collection.
Creator	The person and/or organisation responsible for the creation of the GIS.
Copyright holder	The copyright holder for the drawing. This can be either an individual and/or an organisation.
Period of Creation	The start and end date for the creation of the GIS.
Location	If applicable provide a list of locational terms for the GIS. Each term should be accompanied by an identifying type. New identifiers can be added where necessary. Each distinct term should be entered on a new cell/row.
	Where appropriate provide locational grid references for your GIS. These can be a single

So what's the issue?

Is it time to move on?

Developments

GIS software has moved on:
more available
better capabilities



GeoServer



Key for archaeology is the growth and development of free, open source options i.e. QGIS.



Viable alternative to ESRI software - capabilities and formats.

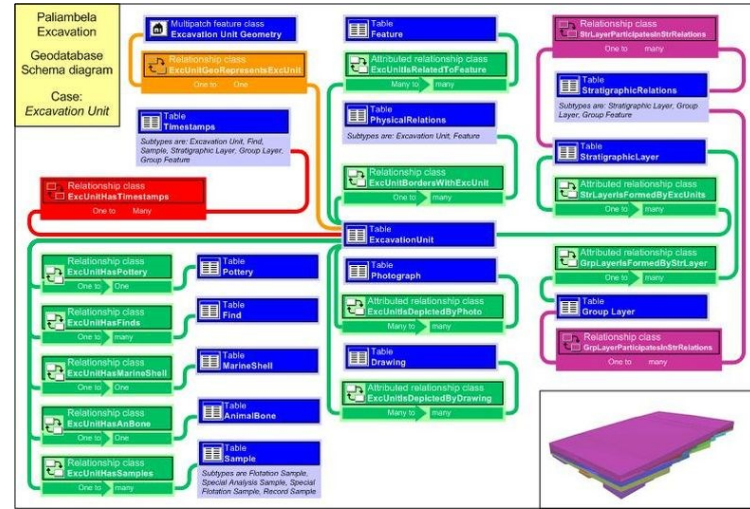
Native formats increasingly becoming an option (i.e. OGC GeoPackage).



Developments

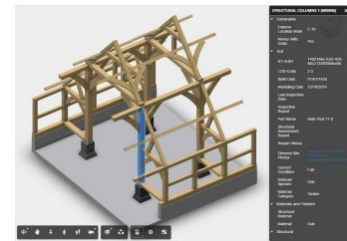
Also seen complexity of geospatial systems increase.

- 2012 **ACE case study** - issues of breaking down and exporting complex geodatabases for preservation.
- problems in exporting 3D and scenes.
- **Growth of BIM**



BIM for Heritage

Developing a Historic Building Information Model



It's not me, it's you

Growing awareness that Shapefile is a limited format

<http://switchfromshapefile.org/>

Shapefile is a bad format

Why is Shapefile so bad? Here are several reasons why the Shapefile is a bad format and you should avoid it.

- No coordinate reference system definition.
- It's a multifile format.
- Attribute names are limited to 10 characters.
- Only 255 attributes. The DBF file does not allow you to store more than 255 attribute fields.
- Limited data types. Data types are limited to float, integer, date and text with a maximum 254 characters.
- Unknown character set. There is no way to specify the character set used in the database.
- It's limited to 2GB of file size. Although some tools are able to surpass this limit, they can never exceed it.
- No topology in the data. There is no way to describe topological relations in the format.
- Single geometry type per file. There is no way to save mixed geometry features.
- More complicated data structures are impossible to save. It's a "flat table" format.
- There is no way to store 3D data with textures or appearances such as material definitions. There is



Developments

Where data can be exported to established formats such as Shapefile, this leads to:

- extra work breaking down dataset
- extra work documenting the process
- potential barrier to reuse through greater complexity & no. of files.

More files + processes = more documentation

Developments



And then we have the ongoing development of metadata standards, not a problem in itself but...

Our current approach is 'granular' - how to apply it to more complex systems/datasets without making the process of depositing more complicated.

So what do we do?

Moving on...

Consider our aims

- effectively ingest, preserve and disseminate spatial data
- Keep to a limited number of reliable formats

Is Shapefile still fit for purpose?

- DPC 2021 Guidance Note: SHP is ubiquitous,
- recommended as a standard by LC.
- Other similar repositories (SND, DANS) accept multiple formats.

Adopt a more complex approach reflecting complexity of data?

Greater degree of assessment?

Moving on...

More radical solutions e.g. WFS - provide a range of options?

As always, solution starts with consulting the designated community:

- how they're creating data
- assess reuse opportunities and barriers.

Look at developments elsewhere (through events like this).

Archaeology Data Service

<http://archaeologydataservice.ac.uk>

Department of Archaeology

The King's Manor

University of York

York, YO1 7EP

THANK YOU

Kieron Niven
kieron.niven@york.ac.uk

