Utilising Third-Party Datasets: an archaeological perspective

Dr. John Naylor

Developing a Digitisation Project Workshop
9 March 2005

Viking and Anglo-Saxon Landscape and Economy
Project aims:

• Aim 1: to map national distributions of metalwork types AD700-1000 and to compare these with distributions of early medieval coinage, and landscape factors, in order to understand the visibility, recovery and archaeological distribution of early medieval ‘productive sites’.

• Aim 2: to characterise the finds assemblages of individual known sites, graphing percentages of coins and other object types in order to examine change through time and to derive ‘fingerprints’ which will help define a hierarchy of settlement types.

• Aims 3 & 4: fieldwork and interpretation
The importance of third party data to VASLE

- Large amounts of data are from a wide geographical area (covering much of UK).
- Work can begin immediately subject to any required cleaning and/or enhancement of the data.
  - Deletion of unwanted data (i.e. outside the date range of the project),
  - Presentation of data in a standardised form, providing additional information to records where applicable/appropriate.
The importance of third party data to VASLE

• Data collection from scratch prohibitive, and extent of project aims would become unfeasible.
• The nature of the data collection is also difficult, with levels of trust an important issue.
  – Locally-based individuals, e.g. FLOs, build confidence of local metal-detecting community in order to secure data.
  – In reality we probably would not get the data in any other way.
  – Much of the data would otherwise go unpublished or be lost.
The importance of third party data to VASLE

- Large, ready-made datasets, much of which has either never been analysed or only in smaller subsets.
- Provide geographical scope well beyond any current archaeological surveys.
The Datasets: Early Medieval Corpus (EMC)

- Corpus of all finds of coinage dated from 410-1180AD
- Data composed all sources but casual finds, often metal-detectorists, a major component.
- Collated by small team of specialists at the Fitzwilliam Museum, Cambridge.
The Datasets: Portable Antiquities Scheme (PAS)

- Began 1997 based around five areas
- Extended across England and Wales in 2003- 32 separate areas (often based on counties) with staff generally based at local museum.
- Members of the public report their finds to these contacts and all data fed onto global database, subject to checking by specialist staff.
- Much data available is not yet checked.
- Deals with any casual find of any period, i.e. all materials from Palaeolithic to Modern
Conflicts in the Data

• Much of the coinage data from same source types
  – Repetition of data?

• PAS Early Medieval Finds Advisor also reports to EMC
  – Consistency of recording between two datasets
  – Any repetition recorded and cited in each case
  – Unrecorded repetition should not occur

• For coinage only EMC data required
Obtaining the datasets

• Data format of prime importance
  – VASLE requirements
    • Easy editing and interrogation
    • Ability to be easily linked to other datasets for analysis
  – Online datasets
    • Searchable, but format not suitable for our needs
  – Data obtained
    • Permission gained prior to grant application for data in format required directly from source
    • MS Excel spreadsheets, imported into MS Access
The Data: raw materials

- Portable Antiquities Scheme
  - Multiple material, multi-period
  - Highly public oriented
  - Produced in two-tiered system
  - Primary identification by specialist unlikely
  - Data sorting and filtering difficult
  - Lack of standardisation, poor dating
  - Cleaning and enhancement necessary

- Early Medieval Corpus
  - Single material, single period
  - Produced by small team of specialists
  - Highly standardised, rigorous language
  - Material closely dated
  - Very little cleaning or enhancement required
The Data: cleaning & enhancement

• VASLE’s Aim 1 scheduled for c.10 months, 6 months for cleaning and enhancement of raw data

• What is data cleaning and enhancement?
  – Cleaning
    • deletion of all non-appropriate data
  – Enhancement:
    • ensuring that data is as accurate as possible
    • Standardisation of data for interrogation and analysis
The Data: cleaning

- Portable Antiquities Scheme
  - All ‘early medieval’ records, AD410-1066, c.5000 in total
  - Can delete those obviously outside study period
  - Others require checking for
    - identification and dating
    - only then deleted or enhanced

- Early Medieval Corpus
  - Filtering to requirements straightforward
  - Little cleaning required
The Data: enhancement

- Portable Antiquities Scheme
  - Difficulty owing to lack of standardised language, identification and dating
  - Large amounts of work to produce dataset for confident interrogation

- Early Medieval Corpus
  - Little work required
  - Highly standardised dataset
  - Easily filtered and searched
PAS enhancement case study: ‘spherical-headed’ Anglo-Saxon pins

• PAS records
  – Varied terminology
  – Varied dating
  – Class of Object often incomplete
  – Multi-find records

• VASLE database
  – All identified under same scheme, and placed under ‘class’
  – Dating standardised
  – Multi-find records separated
PAS enhancement case study: ‘spherical-headed’ Anglo-Saxon pins

- Over 350 pins
- Each checked separately against web images (where available)
- Each record placed within recognised typology
- Enhanced data entered onto database
- Two weeks (full-time)
The Data: integration

- Three datasets utilising similar data types
- Each produced and presented differently
- Physical integration
  - Link via parish name
- Difficulties in integration
  - Each dataset must be configured to same standards
  - Rigorous use of language throughout each database
  - Standardised terminology throughout
The value of VASLE’s experience

- Datasets utilising similar information can be extremely different
- Careful consideration regarding:
  - structure of background organisation, and
  - audience for the datasets to be used should be considered carefully
- Potential labour intensiveness of such an enterprise cannot be underestimated.
The value of VASLE’s experience

• The PAS is an ideal dataset for those working on particular areas or sites and in local history.

• The nature of the EMC makes it a more immediate and accessible tool for academic research. The PAS is extremely valuable but it requires more work prior to analysis.

• Regarding VASLE itself
  – our timetable for the initial data work may require some adjustment
  – modifications to our approach may be warranted in this.
Concluding remarks: are third party datasets worth the bother?

- Their existence is important- VASLE could not realistically collect this data itself
- Invaluable tool once cleaned and enhanced