

Procedures for Measuring Women's Influence: Data translation and manipulation and related problems

Authors' details

Penelope Allison

School of Archaeology and Ancient History, University of Leicester, University Road, Leicester LE1 7RH, United Kingdom

pma9@le.ac.uk

Patrick Faulkner

School of Social Science, Michie Building, The University of Queensland, St. Lucia, QLD 4072, Australia

p.faulkner@uq.edu.au

Andrew Fairbairn

School of Social Science, Michie Building, The University of Queensland, St. Lucia, QLD 4072, Australia

a.fairbairn@uq.edu.au

Steven Ellis

Department of Classics, University of Cincinnati, PO Box 210226, Cincinnati, OH 45221-0226, USA

steven.ellis@uc.edu

Keywords: GIS, intra-site analyses, Roman forts, Germany, digitising plans and artefact catalogues

This article will appeal to archaeologists interested in creating intra-site 'pseudo-GIS' environments for analysing artefact distribution.

Table of Contents

Summary

List of Figures

1 Introduction

2 The Methodology

2.1 The site plans

2.1.1 Capturing and preparing the site plans

2.1.2 Redrawing and preparing the plans for export into GIS

2.1.3 Setting up a system of coordinates

2.2 The artefact catalogues

2.2.1 Capturing data from the artefact catalogues

2.2.2. Correcting the data from the artefact catalogues

2.2.3 Tabulating the data from the artefact catalogues

2.3 Database querying

2.4 Plotting these data in ArcGIS

3 Discussion

3.1 Resolving the problems

Acknowledgements

Bibliography

Abstract

Preparing data from artefact catalogues of previously published German excavation reports, in the project 'Engendering Roman Spaces', required ongoing refinement of data translation and digital manipulation using a variety of software packages. This process included the use of Optical Character Recognition (OCR) software, spreadsheets, database and graphics programs, and the final presentation of the data in ArcGIS [LINK to glossary] as an interpretative tool. With each step, a number of challenges were encountered relating to the quality of the data and original cataloguing processes, and the limitations of the software packages being used.

Excavation reports of four Roman military sites – the forts of Vetera I, Ellingen, Oberstimm and Rottweil – are used in this article to highlight the range of problems encountered and solutions arrived to resolve them, a process requiring constant revision and refinement.

1. Introduction

This [section article](#) outlines the methodology employed for the 'Engendering Roman Spaces' project, described in Allison in this volume [LINK to PAMeasuringwomen.doc]. It highlights the processes by which 'legacy data' were transformed to address the research questions of this project. It also discusses some of the problems encountered in using these data, and the resolutions of such problems.

This project used GIS techniques to carry out small-scale inter- and intra-site analyses of artefact distribution within built environments. Moreover, it used GIS as an analytical tool to assess the data compiled in printed excavation reports. The difficulties with this approach were in the implementation of appropriate methodologies for importing and adapting existing printed data, and then with the creation of clear and cohesive displays for the analysis and interpretation of these data, so that the project's specific research questions could be addressed. These questions had not been taken into account when the excavations were originally recorded and published.

How did we get the data from the printed form to the final form in ArcGIS (Fig. 1 [LINK to AllisonetalFigure1.jpg])? The methodologies employed in this project involved the selection of suitable published sites, converting their printed plans and artefact catalogues into digital formats so that the artefacts could be sorted according to provenances and activities, and then mapping these artefacts, according to interpretative categories, onto the new digital site plans using GIS techniques (Fig. 2 [LINK to AllisonetalFigure2.jpg]).

As outlined by Allison in this volume, the sites were selected if their published excavation reports had the relevant characteristics for analysis [LINK to PAMeasuringwomen.doc section 3]. Each site had to have a relatively comprehensive artefact catalogue and usable site plans. Each catalogue entry needed to have an associated find-spot or provenance within the site, to locate each artefact or group of artefacts on a digital plan. Even within these criteria, each of the sites investigated – Vetera I, Ellingen, Oberstimm, Hesselbach and Rottweil Forts I-II – varied considerably in terms of the detail within each catalogue entry, the information in the site plans and, more importantly, the provenance information provided. The following

discussion examines the procedures used in each of these sites, in the order that they were processed.

The primary source data for this project were the printed artefact catalogues and site plans, the formats of which vary between the sites chosen. The first stage was to 'capture' these data and convert them into formats that could be used in a GIS environment, so that their distribution and density could be plotted.

2. The Methodology

2.1 The site plans

2.1.1 Capturing and preparing the site plans

The first step for the digitisation of the plans was to scan all the necessary printed plans as TIFF image files. The next step was to redraw them digitally to prepare them for importation into ArcGIS. However, each plan, or set of plans, had problems that needed to be overcome before they could be redrawn digitally as a single comprehensive map of the site with Cartesian coordinates.

The most important problem concerned the differences in scales and alignments in the printed plans for each site. This applied particularly to the excavations of Vetera I, for which there are a number of published plans but no single comprehensive and so usable one. Hanel published a plan of the labelled buildings of the last fortress, at a scale of 1:5000 (Hanel 1995, pl. 169). Plans of the remains from the earlier fortresses are found in Lehner's earlier publications, [also](#) at a scale of 1:5000 (e.g. Lehner 1930, pl. II). Hanel ~~also~~ published a separate overall plan of most of the 1629 excavation trenches at 1:2000 (Hanel 1995, plan 1). In addition, he published eleven more detailed plans showing the spatial relationships between particular buildings and excavation trenches in various parts of the fortress, at scales of 1:400 (Hanel 1995, plans 2-3, 5-12) and 1:350 (Hanel 1995, plan 4). These trench plans were published on sheets greater than A3 in size and needed to be scanned on a large-format scanner. Because of differences in scale and alignment, few of these plans could be easily registered alongside their neighbours. Therefore, to convert each plan to a common scale was quite achievable, but the unavoidable consequence was that gross inconsistencies of alignment and omissions of detail were revealed, each instance of which required considerable adjustment before digital redrawing could take place.

For the fort at Ellingen the process was simpler. One large A2 printed plan of the entire site, at a scale 1:250, provided the basis for the digitised map (Zanier 1992, plan 1). This printed plan includes the buildings and features within this fort, and their construction phases, which could be separated into 'layers' on the digital map.

For the fort at Oberstimm a number of the A4 printed plans, again at differing scales, needed to be compiled together to create the final digital map. These printed plans included overall site plans of the excavation areas at a scale of 1:1250 (Schönberger 1978, figs 2-3); overall plans of the building phases at a scale of 1:1000 (Schönberger 1978, figs 64-5, 67-8); details of particular excavation areas showing buildings and features at a scale of 1:200 (Schönberger 1978, figs 5-7, 13, 20, 22, 24-6, 30-1, 33, 42-4, 48-50, 54-6); and plans of individual buildings and their phases of construction, at a scale of 1:400 (Schönberger 1978, figs 23, 29, 32, 41, 46, 59).

Two main plans of the overall fort at Hesselbach were published at a scale of 1:500 (Baatz 1973, pls 2 and 4). These printed plans included all the buildings in the fort, with their construction phases, possible reconstructions and identifications.

The preparation for the digital map for Rottweil Forts I and II was more complex. It, again, required compiling information from a number of detailed plans at different scales, many of which were greater than A3 in size. Each plan was scanned, in sections, on an A4 scanner. The sections were then brought together and each plan re-drawn separately. The base plan showed the fortifications of the two forts, at a scale of 1:2000 (Franke 2003, plan 11). Summary plans of the excavation areas and reconstructions of the barracks buildings, also at a scale of 1:2000, provided the basic grid system on which the more detailed plans could be aligned (Franke 2003, figs 12-14). Details of each excavation area, and the features and phases excavated, were taken from plans at a scale of 1:200, with relevant grid coordinates (Franke 2003, figs 6, 20, 27-8, 30, 39-40, 45, 50, 52, 57, and 59, and plans 1-10). Reconstruction plans were also incorporated into the digital map (Franke 2003, figs 60 and 62).

2.1.2 Redrawing and preparing the plans for export into GIS

To redraw these printed and scanned plans digitally, vector graphics applications (such as Adobe Illustrator) rather than standard GIS software and techniques were used because these made the digitisation process quicker and required only basic skills in using GIS. In particular, it was easier to digitise 'non-regular' graphic features, such as the curved lines of the fortifications, buildings and excavation trenches. Vector graphics applications use Bézier 'splines' to produce continuous, mathematically precise, curves and lines using 'end points' and automatically generated 'control points', whereas GIS uses a continuous string of XY coordinate pairs (called vertices) (see Allison *et al.* 2004, fig. 6.3). An arc, such as the fortifications of a Roman fort, is never perfectly smooth when drawn with GIS software. More importantly, while straight lines are simple to draw in GIS, closed shapes, particularly the type of irregular 'polygons', or multi-sided features, that represent areas common in archaeological plans, are not. A further advantage of maps drawn with vector graphic applications is that these high-resolution vector graphics files are of a higher quality for both print and web-based publications.

As part of the digital redrawing procedure the various features within the fort plans – the building or structural features, the building phases, the excavation areas – were each assigned their own 'layer', a system which is generic to all graphics programmes. The layer system also allows these individual features to be converted into 'shape' files, one of the formats used in GIS to create a comprehensive geo-referenced [LINK to glossary] map where different aspects (e.g. building phases or reconstructions) can be viewed separately or together, as required. For Vetera I, each building was assigned a 'layer'. The most important 'layer', for plotting artefact distribution, however, comprised the 1629 trenches. Artefacts in the catalogues for Vetera I were provenanced to a trench or to a feature within a trench (e.g. a pit). Each of these trenches had to be drawn as a closed polygon, and then assigned a unique numerical identifier, or 'common key', within its boundary. For Vetera I this identifier consisted, in most cases, of the trench number and was assigned to the mid-point of the trench. Where there was more specific information a numeric identifier was assigned to that specific feature (e.g. Pit 6 within Trench 311). For the other sites, the 'layers' consisted variously of excavated areas, buildings, specific features, and

construction phases, either of particular buildings or of the overall fort. This was because the artefacts in the published catalogues of these sites were provenanced to these features or aspects of the site rather than to specified excavation trenches. The next step in the procedure was to import the various 'layers' that made up the digitised site plan, into ArcGIS. This was done by using an Adobe Illustrator 'plug-in' [LINK to glossary] supplied by Avenza (MaPublisher), which converted the layers in the graphics application into 'shape' files.

2.1.3 Setting up a system of coordinates

Because the original plans, and therefore the digital plans, were not truly geo-referenced, the next step in the procedure was to create an XY coordinate system for each site that was internally consistent for that site, but not necessarily connected to a global coordinate system. To be able to plot data from the artefact catalogues onto the GIS map both sets of data needed the 'common key' that referred to a unique position on the map. This 'common key' was compiled from the published provenance registers for each site and from provenance information in the artefact catalogues. It needed to be exactly the same in the GIS map as in the digitised artefact catalogue.

As is evident from the above discussion, the level of detail provided for each provenance, in the site publications, varies considerably within and between the sites. At times the provenance information refers only in general terms to an excavation trench or building/area within the site, including an area or trench that may cover more than one building. At others information is provided on more specific locations within these areas, such as a particular room. And sometimes finds may be provenanced to specific and identifiable features, such as clearly defined and stratified features (e.g. pits). These variations in levels of detail had to be taken into account when setting up the coordinate systems for each site. And because of this variation, XY coordinate systems, rather than polygons representing specific areas of the fort (e.g. buildings or rooms), were used.

The published plan of the excavation trenches at Vetera I had a general set of coordinates that were listed in the trench provenance register (Hanel 1995, 341-52 and plan 1). However, the coordinate grid in this plan was not precise enough for the objectives of this study, especially for the more specific provenance information provided in the artefact catalogue (e.g. artefact provenances to pottery kilns or pits) and in the more detailed trench plans (i.e. Hanel 1995, plans 2-12). Therefore, a new set of XY coordinates was generated from the GIS map using a GIS script supplied by ESRI [LINK to glossary] (addxycoo.ave). The application of this script generated an XY coordinate at the mid-point of each feature in the plan (e.g. the trench). However, on this occasion, this first required that each trench, or feature, was drawn as a closed shape, or polygon. These closed polygons ensured that the non-geo-referenced, but fixed, system of coordinates could be securely attached to these locations within the digitised site plan. As each uniquely identified location now had its own XY coordinates, an attribute table could be generated in ArcGIS using the unique location identifiers and the XY coordinates encoded in the shape files. This attribute table, with fields for the X and Y coordinates for all the numeric identifiers, could then be exported into the artefact catalogue Excel spreadsheet as a separate table of provenance data. The XY coordinates of each unique identifier could then be linked to the artefact data by querying the database.

The printed site plan for the fort at Ellingen had a more precise grid coordinate system than the plans for Vetera I (Zanier 1992, plan 1). These coordinates are included in the provenance register (Zanier 1992, 320-32) and could be incorporated in the GIS map. While the plan for Ellingen includes a site grid, this was an open-plan excavation. So, no trenches are identifiable in this plan, for which the XY coordinate of the mid-point could be generated using the script, as was the process for Vetera I. A new system for registering spatial information was required; one that would recognise and incorporate the spatial information from the provenance register. The coordinates on the printed site plan needed to correspond to the unregistered and 'floating' XY coordinates that are built into the ArcGIS program. The simplest method was to drag and re-scale the digitised site plan so that the 0-0 XY coordinate point on the plan of Ellingen was now located directly over the 0-0 point in the ArcGIS program. Once the 0-0 point was in place, the site could be re-scaled so that each 10 metre unit on the grid corresponded to the measurement of 10 metres from the ArcGIS 0-0 point, and so on. All of the features digitised onto the plan could now be spatially registered to the 0-0 point. This procedure also necessitated the editing of the projection in ArcGIS so that the units of measurement were now simply registered in metres, rather than, for example, in feet or degrees. The process for producing a coordinate system for Ellingen's artefact catalogue and digitised site map essentially meant that ArcGIS converted the numeric grid of the original plan to produce a set of pseudo-latitudinal and -longitudinal readings. These readings correlated with the coordinates in the published provenance register (Zanier 1992, 320-32) and in each entry in the artefact catalogue. A summary provenance table was generated by importing the printed provenance register into a worksheet in the spreadsheet for the artefact catalogue.

In the provenance register each location entry included a provenance code or unique numeric identifier; a location description according to the building, part of a building or feature; and, generally, a set of grid coordinates corresponding to those in the printed site plan. Those entries in the provenance register without coordinates were either generalised areas or referred to stray finds from the fort or *vicus*. The latter were often from earlier excavations and chance finds. Some numeric location identifiers in the provenance register were assigned a range of coordinates; and sometimes two numeric identifiers had the same set of XY coordinates, often because these identifiers referred to the same location but to different building phases. For those locations without coordinates, appropriate coordinates had to be compiled manually from the description of the location, or from its mid-point (e.g. the mid-point of the east part of Building C). A mid-point also had to be used where the numeric identifier had a range of coordinates in the provenance register. Stray finds, without coordinates but with a provenance code and a general location (e.g. the *vicus*, the north gate area, etc.), were assigned appropriate coordinates. Artefact inventory numbers were based on these provenance codes but not every catalogue entry had such an inventory number. Such cases were provided with a set of XY coordinates that plotted them together, as stray finds, outside the fort precinct in the bottom right corner of the final GIS map. Thus, further unique codes, each with an XY coordinate, were added to the provenance data table in the artefact catalogue spreadsheet, so that each actual location had a unique identifier code and each artefact in the catalogue could be plotted.

No coordinate system was available for the printed plans of the fort at Oberstimm. The artefacts from Schönberger's excavations are provenanced to excavation areas or to specific features within these areas. Artefacts from the earlier excavations (e.g. Witz 1911) were provenanced to buildings. Thus, the location descriptions in Schönberger's provenance register (Schönberger 1978, 314-18) and in the artefact catalogues for this site are published in varying levels of detail, relating either to entire open excavation areas, to buildings, to particular features within these areas, or to stratified contexts. The process used for adding a coordinate system to these data was a combination of those used for Vetera I and for Ellingen. As for Ellingen, a provenance data table was compiled from the provenance register. As for Vetera I, each excavated area or feature was given a unique identifier code, but these had to be compiled manually. For example, posthole 20 in excavation area 1a was given the code 1:010200, and this unique identifier code was manually added to the relevant entry in the provenance data table.

These codes were then added to the attributes table in the GIS map, by creating a new field for artefact provenances within this table and then manually inputting the relevant code for each required location (e.g. for a pit or for general location within the trench) into the new field. The XY coordinates thus generated for each identifier code were then exported back into the provenance data table.

For the fort at Hesselbach, the published plans and the provenance register (Batz 1973, 113-14) were in a similar format to those for Oberstimm, and most artefacts were provenanced to a building or part thereof. However, there were only forty-five identifiable locations in the provenance register for Hesselbach so the easiest way to bring this information into the provenance data table was to take manual readings of the locations from the GIS map and paste these readings directly into the provenance table.

This manual procedure was also used for Forts I and II at Rottweil, which had only fifty-five identifiable excavation areas and features, by which the artefact catalogue was organised in the publication, with the artefacts usually provenanced to specific features within an excavation area. Some of these provenances were precise, such as pits and wall trenches, some were particular buildings and building phases, but some comprised a general excavation area. The excavation areas usually included not more than two buildings or comparable feature (e.g. a street).

By this process, which varied slightly for each site, the XY coordinates for each provenance code could be generated where required (e.g. where codes were not originally assigned in the published documents) for each individual artefact from each site (Fig. 3 [[LINK to AllisonetalFigure3.jpg](#)]). The XY coordinates themselves were not added directly to each individual artefact or digitised catalogue entry, however. Instead, these coordinates were included in a separate table of the summary data for each provenance – the provenance data table. Although each site required different methods to generate XY coordinates and provenance codes, once these were defined, the procedures were the same for each of the sites. That is, the provenance data table could be linked to the artefact catalogue table through the unique identifier codes allocated to each provenance location within the plans, and so provide a set of XY coordinates for each artefact, using a select query in Access. However, the artefact catalogue tables first needed to be created.

2.2 The artefact catalogues

2.2.1 Capturing data from the artefact catalogues

The procedures for converting the printed artefact catalogues into a format that is usable in GIS involved the following main steps: capturing the data; correcting them; tabulating them in a spreadsheet; importing them into a database; and querying them.

The first step was to scan the artefact catalogues into TIFF image files, as multiple-page images at a minimum resolution of 300 dpi. These TIFF files were then converted into digital text using ABBYY Fine Reader Optical Character Recognition (OCR) software, with the OCR output being saved as Microsoft Word files. The Fine Reader software is able to recognise and convert characters in most languages, including German, into electronic form and to check the spelling of the resulting text. In general, the OCR software produces reasonably reliable data from TIFF image files scanned at 300 dpi, but the software's reliability decreases significantly if the scans are produced at a lower resolution. Transcription errors still occurred, but the output could be easily checked against the original catalogue because errors or unrecognisable text are highlighted by the OCR program. The production of multiple-page scans and the capacity of Fine Reader to process these multiple-page TIFF files rendered this part of the process was relatively speedy and effective.

2.2.2. Correcting the data from the artefact catalogues

While ABBYY Fine Reader produced editable digital text files of the original printed catalogues for each of the sites analysed, the formats of the originals meant that these text files could not be imported directly into the relational databases needed for use in GIS. Each entry in the printed artefact catalogues generally consists of a continuous block of text. In most cases, particularly for Vetera I and Ellingen and most of Oberstimm, each entry contains a single artefact. However, many catalogue entries, across all sites, include a number of artefacts of a particular type, from different provenances, followed by a description of that type. At all sites the entries are arranged into groups according to typological classification, with that classification as the heading for the group. To be useful to the aims of the project, the information contained in the headings and in each catalogue entry had to be re-organised and re-formatted into a series of fields.

The process of re-formatting the catalogue entries involved the search and replace tools in Microsoft Word and Microsoft Excel. Basic formatting of the OCR text files was carried out in Word, with the formatting of the data matrix in Excel. The first step was to split the OCR text output into paragraphs, one for each catalogue entry. Each paragraph, or row, was then split into a series of cells by inserting tab-stops in the Word documents at the point where cell boundaries were required. For example, each of an artefact's catalogue number, description, measurements, provenance, and illustration references needed to be assigned to a separate cell.

This process worked well when reformatting the catalogues for Vetera I, Ellingen and Oberstimm, as these catalogues had abbreviations at the start of each catalogue entry sub-section, indicating the type of information it contained (e.g. 'Dm' for dimensions'). These abbreviations could be automatically replaced with tab-stops using the 'find and replace' tool in Word. Once rows were arranged and cell

boundaries defined, the text was simply copied and pasted into a new Excel worksheet. Excel automatically organised the data according to these cells and rows, using the paragraph mark to signify a row-ending (i.e. a single catalogue entry) and tab-stops to signify cell boundaries within a row (i.e. a sub-section of a catalogue entry).

Because the quality and detail of the published artefact catalogues varied considerably, both within and between these sites, this process had to be constantly adjusted. This adjustment was particularly significant for the catalogue entries that do not conform neatly to the general catalogue format of one artefact per entry. This applied, for example, to certain entries for Ellingen where the individual artefacts were not considered important (e.g. one entry with several nails from various provenances) and to the entire Hesselbach catalogue, where a single entry is used to describe a particular type of artefact and includes a number of artefacts from different provenances. In such cases these data needed to be extracted and separated, and then double checked against the original printed catalogue, to make sure each entry was in the correct single row and not split over several rows, and that each row did not contain several entries. This process was much more time consuming for Hesselbach and Rottweil than for the first three sites. However, it provided adequate opportunity to check for misspellings and other transcription errors generated during the OCR process. Only after this double checking could these data be exported into one Excel worksheet as the original file and back-up, and then copied into another Excel file for manual reformatting by ‘dragging and dropping’, or ‘cutting and pasting’ the data. For Hesselbach and Rottweil, in particular, this process helped reduce the number of errors that could occur through more automated data translation processes and it ensured the assignment of information to the appropriate cell in the resulting spreadsheet.

2.2.3 Tabulating the data from the artefact catalogues

For the information in the Excel spreadsheets to be queried in a database, the cells for each entry needed to be arranged into fields that each contained only one specific type of information, as relevant to the aims of this project (e.g. a catalogue number, artefact description or provenance). Because of the correcting procedure, some of this organisation occurred automatically, but some information again had to be manually ‘dragged and dropped’ into the correct fields.

Once the original data were formatted correctly, new fields were created to add in further information, specifically chronological information and the interpretative categories, such as activity, gender, and status categories, used in this project. Ascribing these interpretative categories required considerable decision-making by Penelope Allison, often including on-going adjustment of these decisions and these categories (see Allison in this volume [[LINK to PAMeasuringwomen.doc section 4.1](#)]).

When all these fields and their data were complete, including the addition of interpretative fields, the final addition to the artefact data table was the insertion of further fields with SQL-consistent abbreviations of the descriptive field entries to be used in the GIS analyses – i.e. the fields with dates or building phases, and those with activity and gender categories. Full-text descriptive fields could not be used for

analysis. ArcGIS requires field names that are SQL (Structured Query Language) consistent. This is the standard language used with relational databases where each entry must have a maximum of eight characters and contain only alphanumeric characters or underscores (e.g. not question marks). These SQL abbreviations were created and inserted into the relevant fields (Table 1). These then became the codes that were used as the fields in the summary artefact queries.

Table 1: Sample of SQL abbreviations

Artefact use category	SQL abbrev.	Gender category	SQL abbrev.
dress	D	child	Ch
dress-combat equipment	DE	child?	ZCh
dress?	ZD	female	Fe
dress?/cloth working?	ZD_C	female?	ZFe
dress?/cloth working?/toilet?	ZD_C_T	female?/child?	ZFe_Ch
dress?/cloth working?/writing?	ZD_C_W	male	Ma
dress?/combat equipment?	ZD_E	male?	ZMa
dress?/combat equipment?/horse equipment?	ZD_E_H	male?/female?	ZMa_Fe
dress?/horse equipment?	ZD_H	N/A	U

The end product of this process was a digital version of each original printed artefact catalogue, largely translated into English. This digital version was in the form of a spreadsheet, or data table, which contained the descriptive information from the printed catalogues, plus interpretative categories (e.g. dress, combat equipment, male, female, etc.), additional chronological information (e.g. date ranges for specific brooches), and SQL abbreviations. Each artefact was presented in a single row of data cells in the spreadsheet, arranged into fields, each of which contained a specific type of information.

However, although these tables of artefact data, with their respective provenance codes, and the provenance data tables were now fully formatted, they were still not ready for plotting the artefact distribution in ArcGIS.

2.3 Database querying

The artefact data tables organised data by catalogue entry and the provenance data tables organised data by numeric location identifier, or provenance code. ArcGIS, with its limited database capacity, is unable to generate the quantitative summaries of these data (i.e. the number of each artefact category from each provenance) needed to plot their distribution and density. The next step was, therefore, to import the artefact and provenance data tables into a relational database programme which could generate outputs using its query function, to produce summary quantitative data that could be used in GIS. Because of its widespread availability, Microsoft Access was used. The artefact and provenance data tables were linked within the Access database by creating a relationship based on the provenance code fields as the unique identifier in both tables that enables the artefact data to be queried relative to their locations. For example, at Vetera I, the link was created on the trench and feature numbers in the artefact catalogue and provenance data tables (Fig. 4 [LINK to AllisonetalFigure4.jpg]).

To generate a quantitative summary of the number of artefacts of each category (e.g. female, male, child, etc.) within a given interpretative type (e.g. gender) found in each provenance within the site, Access cross-tab queries were used (Fig. 5 [LINK to AllisonetalFigure5.jpg]). The cross-tab query process involved:

- selecting as the first column heading the field with the provenance codes;
- selecting for further column headings all the SQL abbreviations within the field to be summarised in the query, such as activity or gender (e.g. all those that are categorised as possibly dress-related artefacts);
- selecting the ‘number of artefacts’ field for each category within the specified interpretative type, which is automatically summed by provenance; and also
- setting other parameters, by adding in further columns or limiting them, to refine the query further, e.g. selecting a chronological field to indicate in which specific building phases or periods particular activity or gender categories occur, or limiting the gender categories to only female- and child-related artefacts.

The resulting output, the quantitative summary, indicated the number of artefacts of each category within a selected interpretative type at each location (Fig. 6 [LINK to AllisonetalFigure6.jpg]).

For these data to be plotted using GIS, the XY coordinates assigned to each artefact are also needed. For this, select queries were created that linked the provenance data table to this quantitative summary of artefacts. The provenance codes, or unique identifiers, and the X and Y coordinates in the provenance data table were selected and linked to the relevant provenance codes in the artefact quantitative summary, and thereby to the field categories (or column headings) that we wanted to export for mapping in ArcGIS. This output basically replicated the summary quantitative data produced by the cross-tab query between the artefact data and the provenance data (see Fig. 6 [LINK to AllisonetalFigure6.jpg]), but also included the XY coordinates for each location (Fig. 7 [LINK to AllisonetalFigure7.jpg]).

2.4 Plotting these data in ArcGIS

To create distribution plots of these data in ArcGIS, the first step was to create maps of each site using ArcMAP [LINK to glossary], a component of ArcGIS. The ‘shape’ files of the buildings, excavation trenches and recorded site features, created during the digitisation of the published plans, were included as layers in each map. Then the queries generated in Access, which summarise the quantity of artefacts of each particular category in each location, were exported from Access into DBF IV (.dbf) files. The tabular summary data in these quantitative summary queries, and the XY coordinates they contained, were used to plot the data. This was done by simply adding the query data, stored as dbf files with relevant pathways (i.e. folders and directories), to the relevant GIS map as further layers (Fig. 8 [LINK to AllisonetalFigure8.jpg]). If the pathway of a dbf file was altered after these plots were created, for example if the DBF IV query file was moved from one folder to another, then the data would no longer be readable in GIS, and the pathway, or ‘source’, had to be re-established. Distribution plots of the quantities of artefacts of the various categories within each query were produced. These showed, for example, which

locations or buildings particular gender groups might have frequented or where particular activities may have taken place.

The distribution plots were designed using the 'Symbology' menu in the 'Properties' dialogue box of ArcGIS (Fig. 9 [LINK to AllisonetalFigure9.jpg]), which contains a number of formatting options for plotting such queries in GIS. Initially, several different plot types were tried, but pie charts were found to be the most effective style for visual display and for the analyses of artefact distribution and density. In particular, the diameters of all the pie charts could be scaled according to the number of artefacts in each location and then all plotted at the same scale. Thus, the pie charts allowed the distribution and the relative quantity, or density, of artefacts of several different categories within a particular query to be visualised simultaneously (Fig. 10 [LINK to AllisonetalFigure10.jpg]). This was a useful means of displaying and analysing general distribution patterns, the details of which could then be investigated by, for example, generating plots of individual categories (e.g. of only women and children, or of only a certain building phase).

The more detailed distribution of artefacts within specific buildings, or rooms, can also be investigated by zooming into desired parts of the GIS maps. However, because of the system of mid-points, which was required for many provenances, data plotted to a single point sometimes represents a summary of the artefacts found in a potentially large area of the site (e.g. an entire trench, or more than one building). This was taken into account when analysing these distributions.

3. Discussion

3.1 Resolving the problems

As is evident from the above discussion, the main problem with this whole process was that it was quite time intensive. A further, technical, problem concerned the transfer of the digitised artefact data into GIS in a quick and easy manner. Several options were tested during the initial stages of the project, but there appears to be two main procedures for adding data to a GIS project. The first is by a direct connection to the quantitative summary data with associated XY coordinates in Access, using the OLE DB provider for ODBC driver in the ArcCatalog component of GIS itself, and the second is by exporting the Access select queries containing the XY coordinates and quantitative data as separate Dbase IV format tables and then importing the latter into GIS.

Direct connection to the Access data through ArcCatalog is potentially a useful means by which to add tabular data to ArcGIS, but, in the early stages of the project, we discovered that much of the data could not be plotted, especially in the pie-charts which were the most useful chart style in ArcGIS for this type of data. This was because Access generated null values in the queries' data rows where no occurrences of a specified category were found. ArcGIS 8.3, used for these earlier stages, ignored any row, and thereby any provenance location that contained null values in its pie-chart plots. This effectively excluded most of the data in many of the queries. One of the potential solutions to this problem was to use the Nz function in Access to replace the null values with zeros. This is not a standard SQL command, however, and so it could not be read by ArcGIS.

Thus, the insertion of null values into Access queries, and these values' lack of SQL consistency, caused problems when trying to plot these artefact data directly in ArcGIS. The solution to both these problems was either to use Dbase IV files, exported from Access and imported separately to ArcGIS with the null values replaced by zeros manually, or to use the ArcCatalog component of ArcGIS to create a direct link to Access, which performed the operation automatically. While exporting the DBase IV files from Access and then adding them as layers in ArcGIS meant the production of a large number Dbase IV files, it provided a robust and relatively simple solution to this technical problem.

This null-value translation problem was encountered only during the analysis of Vetera I (See Allison *et al*. 2004, section 7.2.3). We used ArcGIS 9 for the other sites and found that this newer version could resolve some of these difficulties. In effect, rows containing null values were no longer ignored by this version, and the manual insertion of zeros was no longer required.

However, another complication in a project that uses interpretative data is that these data often needed to be re-evaluated and re-interpreted. A need for the re-interpretation and re-classification of the data, and therefore the need for changes to the main catalogue, was often only evident after the data had been plotted and analysed in GIS. For example, subsequent studies (e.g. Böhme-Schönberger 2002) have required changes to some of the categories (e.g. the gender categorisation of some brooches), and hence changes to the artefact distribution plots in earlier publications, as indicated in Allison in this volume (e.g. [LINK to Allison section 5.1](#)). Also, after a series of queries on the data from Oberstimm had been run and plotted, it was apparent that the building phases, to which the artefacts were assigned, needed to be refined and the sub-phases included. In these two cases all the queries relating to the gender category or to the phasing of artefacts had to be re-constructed and re-exported into Dbase for importing into GIS. GIS often did not directly recognise the additional or changed information, in these new queries, so the data had to be re-formatted and re-imported. This process could have been simplified by changing the data in the Dbase files but it was important to maintain an original up-to-date and complete set of data that could be relied upon in any future analyses. This was held in the formatted artefact spreadsheet, in Excel, which is an easy form in which to adjust any data, and which contains all the data from the original printed catalogues, not just those being queried. This is important if the data tables are to be useful in future analyses which seek to answer other research questions.

In summary, although null values are an issue, the direct linking of Access to ArcGIS is not advisable. The process of data re-evaluation means ongoing correction to the data and subsequent changes to the SQL codes used in the catalogue. The ongoing processes of renewed large-scale querying and re-analysis appear to be the only valid methodology for dealing with these types of interpretative data.

Acknowledgements

As in Allison in this volume, we are grateful to the excavators and cataloguers of these forts and to the Australian Research Council for funding this research. We are also especially grateful to Chris Black (formerly of the Centre for Cross-Cultural Studies, Australian National University) for his advice in digitising the plans and to

Karl Nissan (Department of Geography, Australian National University) for his advice and support in using ArcGIS.

Bibliography

- Allison, P.M., Blackall, C., Ellis, S. and Fairbairn, A., 2004 'Extracting the social relevance of artefact distribution within Roman military forts', *Internet Archaeology* **17**. http://intarch.ac.uk/journal/issue17/allison_index.html.
- Baatz, D., 1973 *Kastell Hesselbach und andere Forschungen am Odenwaldlimes*. Berlin.
- Böhme-Schönberger, A., 2002 'Die Distelfibel und die Germanen', in K. Kuzmová, K. Pieta, J. Rajtár (eds), *Zwischen Rom und der Barbaricum, Festschrift für Titus Kolnik zum 70. Geburtstag*. Nitra, 111-16.
- Franke, R., 2003 *Arae Flaviae V: Die Kastell I und II von Arae Flaviae/Rottweil und die römische Okkupation des oberen Neckargebietes*. Stuttgart.
- Hanel, N., 1995 *Vetera I: Die Funde aus den römischen Lagern auf dem Fürstenberg bei Xanten*. Cologne and Bonn.
- Lehner, H., 1930 *Vetera: Die Ergebnisse der Ausgrabungen des Bonner Provinzialmuseums bis 1929*. Frankfurt.
- Schönberger, H., 1978 *Kastell Oberstimm, die Grabungen von 1968 bis 1971*. Berlin.
- Witz, D.H., 1911 'Ausgrabungen bei Oberstimm 1909 Sammelblatt des Historischen', *Vereins Ingolstadt* **33**, 3-47, leaves 1-6.
- Zanier, W., 1992 *Das Römische Kastell Ellingen*. Mainz am Rhein.

Formatted: French (France)

Formatted: French (France)

Formatted: French (France)

Formatted: English (U.K.)

List of Figures

- Figure 1: Sample of published artefact catalogue of brooches (Zanier 1995/1992, 177) and final GIS plot for dress-related artefacts, plotted according to gender, from Ellingen [LINK to AllisonetalFigure1.jpg]
- Figure 2: Methodology and software workflow (drawing A. Fairbairn and C. Blackall) [LINK to AllisonetalFigure2.jpg]
- Figure 3: Provenance and artefact data tables for Vetera I, showing unique identifier codes (i.e. trench numbers) in both tables and XY coordinates added from GIS map [LINK to AllisonetalFigure3.jpg]
- Figure 4: Example of relationships between tables in Access, linked by provenance codes in the 'trench' field for Vetera 1 [LINK to AllisonetalFigure4.jpg]
- Figure 5: Example of the cross-tab query procedure for possible dress items at Ellingen, with linking provenance codes in the 'code' field [LINK to AllisonetalFigure5.jpg]
- Figure 6: Example of the cross-tab query output for possible dress items at Ellingen [LINK to AllisonetalFigure6.jpg]
- Figure 7: Example of access selection queries output for possible dress items at Ellingen [LINK to AllisonetalFigure7.jpg]
- Figure 8: Adding query data to ArcMap of possible dress items at Ellingen [LINK to AllisonetalFigure8.jpg]
- Figure 9: Symbology function in ArcMap of possible dress items at Ellingen [LINK to AllisonetalFigure9.jpg]
- Figure 10: Formatted plot in ArcMap for possible dress items at Ellingen. [LINK to AllisonetalFigure10.jpg]