THE TIBER VALLEY PROJECT. THE ROLE OF GIS AND DATABASES IN FIELD SURVEY DATA INTEGRATION AND ANALYSIS

1. INTRODUCTION

The Tiber valley, to the north of Rome, is one of the most intensively surveyed regions of the Mediterranean, yet even here the extent of archaeological knowledge is uneven and some areas have yet to be subjected to any systematic survey at all. Since 1997, the area has been the focus of the British School at Rome’s Tiber Valley Project under the direction of Helen Patterson. The aim of the project is to study the changing landscape of the middle Tiber valley from late prehistory to the medieval period. It examines the impact of the growth, success and transformation of the city of Rome on the history of settlement and economy in the river valley (for a fuller description of the project, see Patterson, Millett 1998; Patterson et al. 2000; and for preliminary results, Patterson 2004; Patterson et al. 2004). At the centre of the project is the restudy of the material from John Ward-Perkins’ South Etruria Survey and the integration of these data with material from other surveys and excavations, in particular the Forma Italiae survey series and the unpublished Farfa Survey. This paper briefly describes the role of information technologies in the Tiber Valley Project and, in particular, the use of databases and GIS to integrate and analyze field survey data; a fuller account is in preparation for the project monograph (Witcher in preparation).

Both the quantity of the data and the questions to be asked of them made the application of information technology a central component of the project from the outset. An initial investigation prior to the commencement of the current project assessed the quality of the archaeological data and the availability of digital geographical and environmental datasets (Harrison et al. 2004). Building on this work, the project has created a series of customized relational databases to manage and analyze the data and to realize their full interpretative potential. In addition, the implementation of a GIS environment to contextualize and analyze the data in greater detail has also been a central concern.

2. THE TIBER VALLEY DATABASES

The creation of a series of individual, but closely-integrated, purpose-built relational databases has been central to the Tiber Valley Project. The key requirement was to fully integrate survey data from a number of different sources, each with diverse levels of precision, accuracy and recorded detail.
The aim was to create a powerful tool that could quickly retrieve information in response to specific research questions but was also flexible enough to be able to respond to new questions as they emerged in the course of research.

The first stage of the project was to commence the systematic study of the material collected by the South Etruria Survey, stored at the British School at Rome, which remained partially unstudied and largely unpublished. The restudy was undertaken by twelve ceramic specialists and coordinated and aided by two Leverhulme-funded fellows, Helga Di Giuseppe and Rob Witcher. The database was explicitly developed as both a management and an analytical tool. The study of over 250 crates of material had to be organized and the quality of the data rigorously checked. For example, it was important that material without a known findspot was isolated and where possible its context identified. The Finds database now holds detailed information for c. 90,000 artefacts ranging from prehistoric to medieval. For the first time, the latent potential of this material can be accessed through structured queries and statistical analysis. For example, it is now possible to identify which classes of pottery are regularly found in association on different sites; this can be used to develop hypotheses about the chronologies of undated coarsewares.

As well as restudying the material collected, the South Etruria Survey archive has also been entered into a custom-built relational database. This is structured around findspots and is directly compatible with the Finds database through each findspot’s unique identifying number (Fig. 1). In reality, most, though not all, findspots are sites and the database is known as the Sites database. Currently, the database holds details of over 5500 findspots in the middle

Fig. 1 – Structure of the Tiber Valley Project databases.

1 The initial South Etruria Survey Pottery Database was designed and implemented by Sarah Poppy.
The Tiber Valley Project. The role of GIS and databases

Tiber valley ranging from small surface scatters to excavated villas and entire towns. The majority of these records (c. 3800) are findspots identified by the South Etruria Survey; the remainder derive from published surveys and excavations and new projects instigated as part of the Tiber Valley Project (Fig. 2). The critical requirement in the creation of the Sites database was a system that could fully integrate these diverse field survey data with all their differences in detail, precision and accuracy. It is therefore useful to review some of the main differences between the survey data incorporated into the Sites database.
The core of the archaeological dataset of the Tiber Valley Project is formed by the results of the South Etruria Survey, a pioneering field project directed by John Ward-Perkins from the 1950s to 1970s (see Potter 1979 for a summary). The survey originated as a response to the rapid post-war expansion of Rome and the introduction of deep and mechanised ploughing through the agricultural developments of the Riforma fondiaria; the result of these changes was destruction of the archaeological record on an unprecedented scale. The South Etruria Survey comprised a number of individual field surveys that shared a similar philosophy and strategies. Over the course of 25 years, these surveys covered a continuous block of land on the west bank of the Tiber from the outskirts of Rome, north to Civita Castellana. As well as fieldwalking, a number of important excavations and environmental studies were also undertaken.

The detailed recording system instigated by Ward-Perkins in 1954 (Fig. 3) on his initial surveys along the Via Veientana and Via Clodia was adopted by
all the subsequent surveys. The South Etruria surveyors noted basic details of the site on a printed record card, together with an accurate, but imprecise, six-figure IGM grid reference; crucially, these coordinates were also marked on each sherd collected. The record cards provide a wealth of information on recovery conditions, the date of the survey and the material observed and collected; much of this information was never included in the survey publications which appeared in the «Papers of the British School at Rome». Ward-Perkins’ systematic approach has been fundamental to the current project, in particular linking material to its relevant findspots and exploiting the spatial component of the data.

The South Etruria Survey never followed a strict research design or methodology and the individual surveys built cumulatively on the collective experience of the many people involved, refining and adapting the methodology over two decades. To this extent, there is a notable difference between the amount of information recorded by the initial surveys and that of the later surveys. The Sites database therefore required sufficient flexibility to accommodate these differing methodologies and results but, at the same time, economy of design through normalization of the data. Further, the long duration of the South Etruria Survey and the large number of people involved meant that many inconsistencies had slipped into the data. The Sites database therefore required strict data entry constraints and checks designed to screen the data and ensure the highest possible quality dataset. For example, the terminology used to describe artefacts and sites changed enormously during the South Etruria Survey’s duration. Data entry therefore required some normalization of terminology through thesauri or predetermined vocabularies of pottery types, site interpretation, etc. This is essential for consistency and comparability and to ensure that all relevant data are retrieved through database queries.

2.2 Forma Italicae surveys

Much of the area covered by the South Etruria Survey has since been subject to resurvey, particularly as part of the Forma Italicae series (formerly the Carta Archeologica d’Italia). A number of these surveys, which have a particular concentration in the Roman suburbium, have been entered into the Sites database, including Cures Sabini (MUZZIOLI 1980), Nomentum (PALA 1976), Sutrium (MORSELLI 1980) and Vicus Matrini (ANDREUSSI 1977). These records and their spatial location, derived from 1:25.000 IGM maps, were screened through a series of checks and cross-checks in order to identify any findspots which had already been located by earlier surveys and to avoid duplication within the database. One important consideration when using these surveys is the date at which they were completed (MATTINGLY, WITCHER 2004, 179). In particular, the most recent surveys have identified far greater numbers of surface scatters than earlier surveys – this cannot be explained by localized differences in the
settlement density alone and indicates more intensive collection and surveying techniques. As with the South Etruria Survey results, the *Forma Italiae* data also underwent a process of standardization of terminology.

2.3 *Regione Lazio* surveys

Several surveys, commissioned by the *Regione Lazio*, have also been incorporated into the project’s Sites database. The survey in the Comune of Capena (1995) intensively resurveyed a 30 km² area previously studied by Barri Jones (1962, 1963). This later, more intensive survey again collected more material and subsequently identified a greater number of sites than Jones. Also entered into the database were the results of the survey of the Comune of Monterotondo, which lies on the opposite, eastern bank of the Tiber. This survey, covering an area of c. 40 km², has served to enhance the dataset in an area where comparatively little survey has been undertaken. Like the Capena survey, the Monterotondo survey used 1:10,000 mapping and included details such as scatter size and field conditions which were only unsystematically recorded by many earlier surveys. All of this information had to be preserved and made accessible within the database.

2.4 *Tiber Valley Project. New surveys*

The Sites database design also had to accommodate the results of more recent and methodologically developed surveys. In particular, this includes surveys which have adopted an off-site approach such as the Farfa survey directed by John Moreland in the mid-1980s. This intensive survey covered an area of 35 km² and systematically collected all surface material field-by-field and used a quite different methodology to that shared by the South Etruria Survey and the *Forma Italiae* (Moreland 1987). A number of new survey projects have been instigated as part of the current Tiber Valley Project – some are designed to revisit and reassess landscapes already subject to survey (e.g. Corese survey: Di Giuseppe et al. 2002; Nepi survey: Di Gennaro et al. 2002); others are investigating poorly-known areas (e.g. Galantina Survey: Gabrielli et al. 2003). All, however, use advanced methodologies which lead to great variation in the overall quality and quantity of data in the Sites database (e.g. the accuracy and precision of spatial recording has increased dramatically with the advent of GPS).

2.5 *Data integration*

All these variations in the quality and quantity of data, whether caused by survey intensity or problems of surface visibility, required a highly flexible database. The great majority of findspots are relatively simple scatters of material with comparatively few details; however, the database also had to
accommodate more complex sites, perhaps subject to excavation or detailed grid-ded surface collection. The resulting structure is a compromise between detail and usability. Hence it is flexible enough to accommodate findspots with diverse histories and material records, but not so large and complex that it is impossibly cumbersome to retrieve data. Wherever possible, information has been broken down into structured pieces of data and entered into carefully constrained database fields. There is also space for free-text description – the detail of excavation, for example, is summarized and included here, rather than modelled in detail.

As noted above, the terminology used has been carefully screened and a uniform vocabulary imposed. However, it was also considered important to preserve the terms used in the original publication or archive. Hence a free-text field allowed the original detail to be entered and a parallel field required the operator to identify a suitable equivalent term from a thesaurus. Preservation of the original terminology has facilitated the construction of “biographies” of findspots and archaeological knowledge of them across the twentieth century, illustrating changes both in the archaeological record itself and its interpretation. Finally, the full integration of the Sites and Finds databases (Fig. 1) has permitted the results of the material restudy to be re-contextualized with their findspots and a final consistent date and site interpretation applied to each.

3. The Geographical Information System

The project uses ESRI GIS software; work was initially undertaken using ArcView 3.2, but has migrated through different versions to the current ArcGIS platform. Other software packages used to process and analyze data include ArcINFO, ERDAS Imagine and Idrisi. This off-the-shelf software solution was selected for several reasons. ESRI programs use proprietary formats (such as shapefiles) which are industry-standard, especially in the public sector. Digital data acquired from the Regione Lazio and Istituto Geografico Militare (IGM) were distributed in this format. These file types are also read by many other software packages, making it easy to pass data from one program to another and to distribute and share data amongst project members. The program can also read a wide range of other proprietary formats, further easing the movement of data between different programs (e.g. AutoCad). A further critical advantage is the ease of communication between ESRI GIS and database packages (in this case, Microsoft Access). Through an SQL connection, any amount of data stored in an external relational database can be imported, queried and plotted in geographical space. Once created, this link remains active, allowing any changes made within the database to be reflected in the GIS. This was of great advantage during the course of data entry, permitting the continuous evolution of the data to be visualized and in turn helping to inform the interpretative process.
At its most basic, the GIS may be used to create conventional distribution maps, illustrating data by type, period or settlement. However, the flexibility of the system means that any number of maps can now be rapidly and confidently created and recreated in response to specific and changing user needs (e.g. Patterson et al. 2005). But, as with the project databases, the development of a GIS has not only helped to store, manage and display the data, but also helped to realize its analytical potential, for example through predictive modelling (Kay, Witcher forthcoming), multicriteria analysis (Goodchild forthcoming), cost and visibility (Belcher et al. 1999) or location analysis (Rajala et al. 1999).

4. GEOGRAPHICAL AND ENVIRONMENTAL DATA

Obviously, GIS technology was not available during the original South Etruria Survey; however, many of the questions posed at the time could have been profitably addressed in this way, for example, the topographical location of sites or the exploitation of certain environmental niches. Subsequently, GIS has become a central technology in the analysis and presentation of regional survey data (see papers in Attema et al. 2002 for Italian examples; also Lock 2000; 2003; Wheatley, Gillings 2002). One reason is the strong spatial component inherent in the study of settlement patterns (Wheatley 2004, 4). The application of GIS technology was therefore considered a central part of the research strategy of the current Tiber Valley Project.

In order to analyze, view and interpret the survey data spatially, a range of digital cartography has been obtained or created (Fig. 4). A fuller discussion of the datasets acquired by the projects can be found in Witcher and Kay (forthcoming), but it is useful here to discuss a few of the geographical datasets and, in particular, how they have been combined with the information from the databases.

Digital maps of rivers and lakes derived from 1:10,000 maps were acquired from the Regione Lazio. Whilst these are modern rivers which may have changed course or been canalised since antiquity, they are crucial to understanding the modern landscape from which archaeological data have been collected. Modern land-use data was also obtained from the Regione. As well as providing useful information for the planning of new fieldwork, current land use could be compared with that recorded during the South Etruria Survey in order to assess the changes to the landscape and their effect on the surface archaeological record over the past fifty years.

Another important layer of information is geology. Data were obtained partially through the digitization of the 1:500,000 Servizio Geologico d’Italia and also through the acquisition of a 1:50,000 coverage from the Regione Lazio. These have been particularly useful for the study of basalt and clay sourcing
in the Tiber valley (Black et al. forthcoming) and for assessment of potential agricultural production in the middle river valley (Goodchild forthcoming).

All of the above are vector datasets; in terms of raster datasets or imagery, one of the most useful resources is the 1:10,000 Carta Tecnica Regionale (CTR). Whilst the South Etruria Survey and the Forma Italiae made use of the IGM 1:25,000 series, the additional detail provided by larger scale mapping has been vital for (re-)locating sites. Recent surveys have taken advantage of this larger scale mapping to plot more accurately individual findspots and transects. Another raster dataset comprises an archive of the 1944 RAF aerial photographs, originally acquired by John Ward-Perkins after the Second World War and currently curated by the Aerofototeca. Whilst their altitude limits their resolution, they provide a useful source of generic information for land use prior to modern agricultural developments of the 1950s and 60s.

The final and perhaps most important geographical dataset is a digital elevation model (DEM). As illustrated in Fig. 5, it offers a three-dimensional representation of the landscape at a resolution of 30 m, and from this a whole series of other landscape attributes can be derived, such as elevation, slope, aspect, topographical form and watercourses. The DEM was created from 34 tiles of digital contours and spotheights derived from the 1:25,000 IGM maps. This scale was chosen for two reasons: first, the bulk of the project’s data was collected and recorded at this scale and, second, this was the largest scale digital data readily available when the project commenced in 1997.
During the process of creating the DEM several problems were encountered, for example broken contours, especially between mapsheets. Whilst the majority of these problems could be rectified through some processing of the data, it was clear in some areas, particularly the dissected valleys around Nepi and Civita Castellana, that the 1:25,000 coverage lacked sufficient detail and additional higher-resolution data were needed to be merged into the DEM to model the extreme topography more effectively. A pilot project was intended to derive a high resolution DEM from SPOT Panchromatic satellite images (Belcher et al. 1999, 96-98); however, the heavy shadows caused by the highly complex and thickly wooded topography of the area made this impossible. Subsequently, a recent joint application with the University of Cambridge to the National Environmental Research Council (NERC) Airborne Research and Survey Facility has enlisted the help of a low-level airborne survey system during a deployment in the Mediterranean in the spring of 2005².

² The Lower Tiber Catchment LiDAR Research Project, that forms part of the wider Tiber Valley Project, is directed by Simon Stoddart (Cambridge University) and conducted by Stephen Kay (British School at Rome). The authors wish to acknowledge the generous support and help in the establishment of a ground GPS base station offered by Nicola D’Agostino of the Istituto Nazionale di Geofisica e Vulcanologia, Roma.
A number of different data types were collected during the flight, including aerial photography at a scale of 1:6500, 2 metre interval CASI vegetational spatial data and, in particular, LIDAR (Light Detection And Ranging) data. LIDAR collects 33,000 laser observations per second and, when operating from an altitude of 1000 metres, the resulting elevation data has an absolute RMS accuracy of better than ±15cms. LIDAR data provide the basis for high-definition terrain models. A case study area of 10×10km was defined, covering the ancient sites of Nepi, Falerii Novi and Falerii Veteres in the ager Faliscus, centring on the complex topography of the Treia Valley where the data quality of the 1:25,000 digital coverage was poorest. The data are currently being processed. Whilst it is clear that this solution is prohibitively expensive to apply to the entire Tiber Valley Project study area, its judicious use can help to mitigate some of the more severe problems with existing digital data.

5. CASE STUDIES AND CONCLUSIONS

The contribution of information technology to the work of the Tiber Valley Project is implicit within all of the project publications cited in this paper, whether in the form of identifying patterns in the data or simple mapping. In this concluding section, we illustrate some of the more explicit IT applications under development and other related issues.

As noted above, a major problem in assessing the significance of the archaeological evidence is the uneven quality of the data due to changing methodologies or even the absence of fieldwork or the destruction of the archaeological record. One approach is to model the biases in the dataset using indicators such as proximity to roads or rivers (RAJALA et al. 1999). Another approach is predictive modelling. This technique has been widely used in the US and the Netherlands for Cultural Resource Management purposes, that is, predicting the location of unknown sites in order that they can be recorded and/or protected from destruction (DUCKE forthcoming).

A preliminary study of the location of Early Imperial villas in the Tiber valley indicates moderate predictive potential using a small number of environmental variables as illustrated in Tav. V (KAY, WITCHER forthcoming). The results were interpolated across the wider research area and used to identify the probability of sites existing in unsurveyed or poorly surveyed areas; this process has helped to improve understanding of the overall distribution of settlement. An interesting aspect of the model, which was developed using Early Imperial villa sites, was its ability to predict the location of discontinuous twentieth century housing development (Tav. VI). This does not come as a complete surprise (e.g. the correspondence of Roman villas and recent house building in the Sabina was noted during the Corese survey, Di GIUSEPPE et al. 2002) and helps to emphasize the importance of understanding the environ-
mental context of settlement distribution and possible settlement location preferences.

Another ongoing study uses a range of environmental variables such as geology and topography to model agricultural production capacity in the middle Tiber valley. The writings of Roman agronomists such as Columella are used to effect a multicriteria reclassification of the landscape and its agrarian potential. Comparison with known settlement patterns then allows assessment of a range of considerations including the veracity of the ancient sources, the scale of rural population and the level of urban dependency on local agricultural supply (GOODCHILD forthcoming).

The use of GIS to assess visibility and viewsheds has been of particular importance in archaeological applications of GIS, in particular because it moves away from, what are often perceived to be, environmentally deterministic interpretations of past human activity (WITCHER 1999). Some preliminary work combining visibility with cost surface analysis of Archaic funerary practices around the city of Nepi suggested a strong contrast between the visual and physical accessibility of the rock-cut tombs (RIVA, STODDART 1996; BELCHER et al. 1999). Other patterns of landscape visibility can be detected in the pre-Roman funerary evidence of the area. For example, ten tumuli of Orientalizing date are known around the major urban site of Veii (Fig. 6). A cumulative viewshed indicates that circa 30% of the local area is visible from the urban plateau, but that nine of the ten tumuli are within view and most are in positions visible from a range of locations (Tav. VII). Similar patterns in the placement of tumuli have also been noted at the Banditaccia necropolis at Cerveteri (CECCARELLI 2001). This suggests that the siting of these monuments were carefully and deliberately structured.

The applications and possibilities offered by the use of Internet Mapping have been widely discussed, particularly in this journal, and are being applied
within the context of the Tiber Valley Project. For example, the preliminary results of geophysical investigation and excavation conducted at Forum Novum-Vescovio by The British School at Rome, British Museum and Birmingham University are accessible via the World Wide Web. Through the application of ArcIMS software (ESRI) hosted by Birmingham University, researchers can explore the results of resistivity, magnetometer and GPR surveys of the Roman and late Antique town (http://www.vista.bham.ac.uk/ims/maps.htm). Another study that forms part of the broader Tiber Valley Project, also applying geophysical techniques to urban sites, is the Roman Towns Project directed by Simon Keay and Martin Millett. Again, the results of magnetometer survey on a range of different urban centres will shortly be available via the World Wide Web for downloading, viewing and analysis (http://ads.ahds.ac.uk/catalogue/resources.html?tibervalley_var_2004).

An important consideration in the creation of digital data is their dissemination and long-term preservation. In line with the requirements of Leverhulme and British Academy-funding, the project will deposit an archive of its digital data with the Archaeology Data Service (ADS) at York, UK. This body is responsible for the curation and dissemination of (publicly-funded) digital archaeological data. The Roman Towns geophysics archive represents the first stage of this process. In due course, the wider project databases and GIS will be deposited with the ADS and become freely available to students and researchers to access and study.

STEPHEN J. KAY
The British School at Rome

ROBERT E. WITCHER
Department of Archaeology
University of Durham

REFERENCES

ADS = Archaeology Data Service (http://ads.ahds.ac.uk/).


BLACK S., BROWNING J., LAURENCE R. (forthcoming), From quarry to road: The supply of basalt for road paving in the Tiber valley, in COARELLI, PATTERSON (forthcoming).


DI GENNARO F., CERASUOLO O., COLONNA C., RAJALA U., STODDART S., WHITEHEAD N. 2002, Recent research on the city and territory of Nepi (VT), «Papers of the British School at Rome», 70, 29-77.


Ducke B. (forthcoming), Regional scale archaeological predictive modelling in north-eastern Germany, in 32nd Annual Conference of Computer Applications in Archaeology (CAA) “Beyond the Artifact” (Prato, Italy 2004).


GOODCHILD H. (forthcoming), The productive potential of the middle Tiber valley, in COARELLI, PATTERSON (forthcoming).


The Tiber Valley Project. The role of GIS and databases


ABSTRACT

Information Technology is at the heart of the Tiber Valley Project, from the integration, storage and analysis of data, through project management to the visualization and dissemination of results. Here, some of the ongoing applications of this technology, both implicit and explicit, have been presented. Detailed results will be published as the project continues, with a synthetic volume currently in preparation (PATTERSON et al. in preparation).