

CHAPTER 3

THE ARCHAEOLOGICAL RECORD OF BOHEMIA. AN ATTEMPT AT AN ANALYTICAL INFORMATION SYSTEM

MARTIN KUINA

Institute of Archaeology, Czech Republic

I. INTRODUCTION

Both of the major historical divisions of what is now the Czech Republic (Bohemia and Moravia, the latter in this instance being taken as including the Czech portions of Silesia) have a generally similar, and in most aspects independent, structuring of their archaeological institutions. This framework comprises an Institute of Archaeology of the Czech Academy of Sciences (in Prague and Brno respectively) as a research centre, universities teaching archaeology (in Prague & Pilsen/Plzeň and Brno & Opava respectively), similarly organised networks of museums and archaeological field units working on contract archaeology. The two parts of the Republic are also more or less independent in terms of their systems of information regarding the appearance of archaeological sites. This contribution stems from the present situation in Bohemia, although the situation in Moravia is not in a broad sense markedly different, due both to the common development of the discipline in both parts of the country and the gradual introduction of the same computer applications and fundamentals. The article provides information on the "national archaeological database" for Bohemia, created in the Institute of Archaeology in Prague, and on this basis considers several general problems facing databases of this type.

The Law on the Cultural Heritage (20/1987 Sb.) places on all those conducting archaeological excavations in Bohemia the obligation to submit information regarding the results of their investigations to the Institute of Archaeology of the Czech Academy of Sciences in Prague. This situation arose from the formal merger of the Institute of Archaeology with the State Archaeological Institute founded in Prague in 1919 (a branch later being established in Brno); one of the chief tasks of the latter was the registration of archaeological sites (Spurný, 1975). Even during its own existence the State Archaeological Institute gathered extensive documentation pertaining to fieldwork and finds, which were incorporated into the Institute of Archaeology's "finds report archive" (a slight but traditional misnomer). Today, the finds report archive contains tens of thousands of entries, and every year some 1,500 new reports are submitted (although regrettably this total represents, despite the clear legal formula, only around half of the actual number of field investigations conducted).

The finds report archive comprises not only finds reports, but also notifications, letters, museum registers and extracts from the literature. It is sorted by cadastre: even in the pre-war years card indexes organised by cadastre and archaeological period had been created. In 1963 the Institute of Archaeology also began to publish, annually or biennially, the journal *Výzkumy v Čechách* ("Archaeological field research in Bohemia"), which contains

brief details of excavations and finds from the relevant period. The volumes of this periodical are another source of information, as until the beginning of the 1990 submissions to it were made separately from those to the finds report archive, and thus their content and that of the archive do not entirely overlap submission. This situation has changed in the last decade, with the periodical now created as automatic output from the computerised database into which all the received information is input (Figure 3.1).

The creation of a computerised information system began in the early 1990, as soon as such technology became available in the workplace. A database records structure was created, along with the new printed form by means of which data are gathered. The content of the database became rooted primarily in information regarding new excavations conducted by Institute of Archaeology under the law cited. In addition all of the volumes of *Výzkumy v Čechách* have been transcribed in, as have several other published archive inventories (e.g. that of Charvátová, Spurný & Vendlová, 1992, covering the period 1919-1953, and of Justová 1968 for the years 1955-1964). This data has been further supplemented by that from several regional finds inventories (e.g. the finds inventory for East Bohemia created by staff of the museum in Hradec Králové, etc.). The database thus created is the *Archaeological Record of Bohemia* (ARB); at present it comprises some 60,000 records. The ARB only partially overlaps with the "finds report archive" because (as noted) it did originate by a breakdown of the latter; but from secondary information sources which themselves naturally are often based on the archive. A comparison of the ARB and the original documents in the finds report archive would no doubt be extremely useful, but it is beyond the means of the team currently administering the ARB (3-4 people).

Given that the ARB was born out of the merging of data series of diverse origins, the quality of the individual records within it varies markedly, and one of the results of the approach to creating it adopted has been a high number of duplicate entries. These problems are being addressed through the gradual revision of the database which will also seek to remove imprecisions in the identification of finds and, in particular, to complement the localisation of finds wherever possible (new finds are always localised by coordinates, but for the majority of the older data spatial locations are given only verbally, and often imprecisely). Great emphasis generally is placed on site localisation: one of the staff of the Department of Spatial Archaeology (ARB team) concentrates solely on the control and registration of spatial data.

ARB data are distributed freely. In the immediate future, as part of an ongoing project on archaeological prediction, it is also intended to make ARB data accessible through the Internet, with the automatic explanation of the most important parts (the keywords) in English. The availability of spatial information through the Internet will, of course, be subject to further discussion – the reason for this being the frequent illegal use of metal detectors, which have recently started to become a major threat to many archaeological sites in Bohemia.

2. SITES OR "ARCHAEOLOGICAL EVENTS"?

The basis of archaeological databases covering large territories comprises as a rule archaeological "sites". "Sites" are understood as spatial, sometimes spatial and chrono-

logical, wholes of finds, which can be clearly delimited in space and divided from other similar wholes. The use of this term is, in practice, often problematic. The "site" in this sense can only be delimited in cases where finds are not particularly dense and/or where the primary elements are conspicuous features visible on the surface ("monuments" such as, for example, earthworks, tumuli or buildings). As soon as the information in the area becomes denser, and its quantity and arrangement is not clear on the surface, the distinction of "sites" becomes unrealistic in practice (Figures 3.2, 3.3, 3.4 and 3.5). These situations arise wherever, for example, modern prospection methods are applied, leading to increases by orders of magnitude of the quantity of information available. In such cases, the landscape becomes a continuum of archaeological information, the internal structure of which is of course seldom clear (Kuna, 1991; 1997; 1998; 2000; Gojda, 2000). This assumption is not only empirically proven by many examples, but can also be arrived at theoretically. In settled areas the landscape was a continually used space, a set of activity areas of diverse functions (Neustupný, 1986; 1998). The term "site" (a discrete whole in an empty landscape) is thus not only methodologically but also theoretically disputable. The units that are perceived as sites are entirely arbitrary heuristic devices, sometimes the result of secondary processes (of archaeological transformation).

In terms of method, the defined "site" represents a certain form of synthesis of component field observations, empirical facts (Neustupný, 1998). This synthesis implicitly includes, above all, the building of a delimited finds unit, from which database records are created, and the establishment of its spatial relationships (topology). In the process of synthesis, primary spatial and other units become derived secondary units, with a necessary loss of some primary information. In and of itself this need not matter, but the problem arises that the new, secondary units are far from being able to be called definitive. Every new find, indeed, brings the danger that it changes the topology of the existing units, and that it will be necessary to re-define not only their sizes and shapes but also the mutual relationships between them (it will, for instance, be necessary to classify some new finds into an existing "site", others may require that two "sites" are joined into one, etc.). In archaeologically rich areas this approach means the risk of the endless re-writing and correction of existing database records, an unpleasant aspect that leads inevitably to errors.

An alternative to this "synthetic" approach is that which is adopted by the ARB, which can be better termed "analytic" (for this term, again see Neustupný, 1998), as individual archaeological observations ("events") are not reconciled with "sites", but are registered separately and are furthermore divided into smaller logical units (components). The ARB is thus not a database of "sites" but a database of "archaeological events" (individual observations) or components. The database merely comprises an accumulation of records of events, without the need to change anything in the earlier records as research proceeds. If several events take place at one and the same location, their pictures may overlap in many ways while the topological relationships between them are examined only when data are drawn from the database for particular professional reasons. This approach retains the maximum primary information, and does not distort the original data regarding the state of research and the subjective opinions of the author during database entry.

3. THE DESCRIPTION OF ARCHAEOLOGICAL EVENTS

The ARB is a relational database system made up of a series of separate tables. The basic unit is the "archaeological event" or spatially and chronologically related archaeological field observations (e.g. surface artefact collection in a particular polygon, monitored stretch of long distance trench, one season of a long-term open-area excavation, etc.). Within the ARB the event is classified by a whole range of information: the name of the cadastre and administrative district, the name of the principle investigator and institution, the date of investigation, the stimulus leading to investigation, the type of archaeological evidence, the types of analysis conducted, and of course the spatial delimitation of the investigated area as a point, line or polygon (as a norm, such delimitation uses co-ordinates drawn from 1:10,000 scale maps). From the professional standpoint, however, the descriptions of the individual parts of the event, the "components" are even more significant. Components are understood to be the separate chronological, spatial and functional wholes identified during a single archaeological event (on this term, albeit in a somewhat different sense, see Neustupný, 1998). Within the framework of the given system, components are thus the smallest, further indivisible, "analytic units". Their descriptions consists of the following data (Figure 3.1):

- *spatial delimitation* (which may of course be identical to the delimitation of the entire event, if it occurs in a small area, or if more precisely located components are not known);
- *dating* with the aid of a hierarchical system of chronological terms, the mutual relationships between which are precisely set out, and enable varying breadths of data selection (e.g. the term "Únětice culture" is contained within "Early Bronze Age", itself falling within "Early or Middle Bronze Age", "Bronze Age", "Bronze Age or Hallstatt" and "Agricultural Prehistory");
- *functional identification*, e.g. settlement, funeral, cult, etc.; finds from secondary positions (e.g. intrusion or residual) are registered as a separate type;
- *precision and "shape"* of the spatial data: three levels of localisation accuracy are recognised (with the margin of error at less than 3 m, 30 m and 300 m), along with four "shapes" (point, line, polygon and area, i.e. a space "somewhere around a given point");
- a *brief inventory of types and approximate quantification of the features and objects* that make up the given component;
- *metadata*, e.g. who assessed the chronology of the given finds.

The opportunities afforded by the descriptions of individual events and components vary markedly depending on the nature of the archaeological observation. For example, in suitable areas aerial photography can provide a good impression of the density and overall extent of the sub-surface features, but it will be difficult to classify these chronologically (the "components" in this case are spatially distinct, but chronologically must be described simply as "prehistoric"). For older finds from sand extraction works it is possible to recognise individual chronological components but their adequately precise localisation is not possible (the various components are most likely to be localised at a single point

in the centre of the extraction area, but with a precision of "3", meaning a margin of error of up to 300 m). Other types of archaeological evidence, too, have specific characters in terms of their spatial and chronological ordering: various types of surface artefact collection, small trenches, the monitoring of long-distance trenches, surface surveys in wooded areas, etc. These differences, too, speak for the need for an "analytic" approach to database constructions, as between the various observations a common denominator can often not be found.

4. APPLICATION OF GIS

The Department of Spatial Archaeology of the Institute of Archaeology in Prague (within whose competence the "finds report archive" falls) has at its disposal the basic technical equipment necessary for the creation and use of GIS. The application of GIS is, however, presently limited to projects that are regional in scale (Kuna, 1996; 1998), rather than across the whole state. The creation of a GIS for a larger territory is for the time being restrained by the high cost of digital maps, as well as by the technical limitations of the hardware available. Moreover, it is also restrained by the character of the data (which would need to be revised before being brought across into a GIS), and the overall aims of the Institute of Archaeology, for which management of the database is a secondary task and not the major constituent of its work. In the field, too, the priorities for GIS technology lie rather elsewhere, e.g. in the computer analysis of landscapes or the elaboration of archaeological prediction methods.

The structure of the ARB enables the straightforward transfer of data into GIS. Events and components within the ARB are localised by one or several pairs of co-ordinates (in the version for dBase a maximum of 6 pairs, while the prepared version for MS Access has no limit), which characterise them as points, lines or polygons. Spatial data are reduced on entry to distances in millimetres for the edge of the 1:10,000 map sheet. These values can easily be read by anyone from the maps: during transfer into the database these data are checked and automatically recalculated into geographical grid co-ordinates.

Some idea can be obtained of the density of archaeological information in several parts of Bohemia from the maps in Figures 3.2, 3.3 and 3.5 (prepared with the aid of ArcView and Idrisi GIS) – in an area of some 6.2 × 4.8 km on the north-eastern periphery of Prague there are in all 1,382 archaeological records (events) with a total of 1,385 components of prehistoric age. Of these, 43 records (90 components) represent data from old finds (predominantly stray finds in the numerous sandpits and brickworks pre-1980), 20 records (147 components, in given cases individual features) are associated with excavations by the District Museum of Brandýs and Labem in the 1980, mainly along long-distance mains trenches, 167 records (370 components) are polygons of surface artefact collection in the 1980, and 952 records (with 778 prehistoric components) are 100 × 100 m grid squares of surface artefact collection conducted in the 1990. The density of archaeological data in this area is above average, but it is by no means unique in Bohemia. At the present time, every year sees further, even large-scale, interventions in this territory.

The archaeological picture in the given area is so complex that it cannot be evaluated without the use of information technology, specifically database systems and GIS. At the same time, the author believes that it is clear that no evaluation (e.g. the establishment of particular numbers of activity areas relating to particular periods, etc.) can be definitive, as increasing numbers of finds not only complement the overall picture but can also change its internal structure. The ARB's methods thus incorporate the principle that that data obtained must be preserved in primary form, in order that it does not suffer premature limitation of its informative value, and that data may be repeatedly used for up-to-date syntheses of the state of knowledge.

GIS technology can be usefully applied not only in the visualisation of data and the identification of secondary information, but also in the actual synthesis of spatial information at any moment it is needed. Such a synthesis may take various forms, and an illustrative example approach is presented here:

- *thematic selection* of the components is conducted, e.g. to a chronological key;
- *categorisation of the selected data* (if this had not already been performed). The significance of components can, for example, be categorised so that the lower values indicate components in the form of isolated artefacts, occurring as intrusions into features of a different period, as components from surface artefact collection which are isolated within the given period, as components from relatively large spatial polygons (imprecise localisation), etc., while higher values are demonstrable and precisely localised evidence for activities of a certain kind;
- transfer of the database extract to (point or polygon) *GIS vector layers*. Spatial data are read by the GIS directly from the database, vectorisation of individual records with the aid of a digitiser in this approach not being required;
- *data integration* from point and polygon data in the framework of vector or raster maps of components of the chosen period (Figure 3.4, upper);
- *modification of map layers* according to elected principles (e.g. the removal of cells with low values; removal of spatially isolated cells from raster maps; interpolation of values; creation of buffers around cells, etc.). Relatively complex operations are also possible between multiple raster maps (e.g. principal component analysis, identifying which areas are typical for a given component type and which are not; cf. Figure 3.4, lower);
- *classification of spatial units* on the newly obtained map, and their transferral into the form of vector layers, or back into the database with units of the "next generation".

This approach simulates the synthesis that would otherwise be conducted by the archaeologist during fieldwork or in the preparation of records for the database. Unlike such syntheses, this approach has the advantages that (1) it is based on describable and quantifiable assumptions, and (2) that it first takes place at the moment that the archaeologist requires information, i.e. on the basis of up-to-date data. The author believes that the described approach can to a great extent be automated, in order that the system of spatial synthesis is conducted independently on the basis of the given parameters.

5. A GLIMPSE OF THE FUTURE

The "analytical" approach contained within the ARB reflects several facts. Firstly, the database came into being under the care of a small team, who are unable by themselves to undertake the complete revision (synthesis) of archaeological data for a large territory. On the other hand, the gathering of primary data by additive means is possible, as in this the information obtained from many colleagues is controlled only in its fundamental points and entered into the database. Secondly, the emphasis on the gathering of "analytical data" stems from the presumption that archaeological finds form highly complex structures that cannot be forced into models of "sites", and which, moreover, gradually change with the advance of archaeological investigation. An archaeological database should thus, in the author's opinion, preserve the maximum primary information, the synthesis of which should take place at the moment when it is needed for a specific purpose.

The technology that archaeology now has at its disposal, however, enables the use of approaches that were unrealistic even in the recent past. Today, the key problem with the spatial and temporal delimitation of database units, whether "sites", "components" or anything else, can be side-stepped by database units based on archaeological wholes the spatial, functional or chronological integrity of which is more or less evident, i.e. individual features/complexes. Hitherto, such an approach was difficult, as it has not been technically feasible for every feature (house, storage pit, grave, tumulus) to obtain geographical co-ordinates, and to work with these in the framework of maps of large territories. At the present time, with the majority of archaeological excavations conducted as a rule with the aid of EDM or GPS, the creation of a database of features with their own co-ordinates is entirely commonplace, and GIS technology enables ease of working with such data in contiguous areas at various scales. It seems, then, that databases of such units are essentially easily attainable, and simpler in type than databases of any imported, larger spatial and temporal wholes. For every feature in such an independent record, classification by basic type (function) and chronological position is possible in most cases (and errors are easily rectified, as they have no effect on the other database units). The majority if such data exist anyway: in finds reports, publications, etc. Further archaeological features, such as tumuli, scatters, earthworks, are today systematically measured, as it is their precise position in geographic co-ordinates (ascertained by GPS) that is important to their repeated identification and protection in threatened cultural landscapes. It does not appear, then, that obtaining such data and bringing them into a summary database presents any great technical problem or time requirement. This approach would be straightforward, too, in that the majority of features in these categories can (given the scale used) for the time being be written to a single point; a more complex approach would be needed only for larger, e.g. linear, features (earthworks, etc.). A database of individual features could, of course, also include cases where the number and types of features cannot be ascertained, where features cannot be precisely localised, etc., for which it would be possible to use several of the approaches outlined above (including principles of the fuzzy logic). Thus, the future seems to lie in databases of this type, working with large quantities of data, albeit relatively simple data, internally integrated and enabling quantitative spatial syntheses using various mathematical methods.

6. REFERENCES

- Charvátová, K.; Spurný, V. & Venclová, N. (1992): *Nálezové zprávy StAÚ 1919-1952*.
- Gojda, M. (2000): *Archeologie krajiny. Vývoj archetypu kulturní krajiny*. Praha (Academia).
- Justová, J. (1968): "Nálezové zprávy Archeologického ústavu ČSAV 1955-1964". *Archeologické studijní materiály* 6. Praha. Archeologický ústav ČSAV.
- Kuna, M. (1991): "The structuring of prehistoric landscape". *Antiquity* 65, 332-347.
- Kuna, M. (1996): "GIS v archeologickém výzkumu regionu: vývoj pravěké sídelní oblasti středních Čech". *Archeologické Rozhledy* 48, 580-604.
- Kuna, M. (1997): "Geografický informační systém a výzkum pravěké sídelní struktury". In Macháček, J. (Ed.): *Počítačová Podpora v Archeologii*, 173-194. Brno. Ústav archeologie a muzeologie.
- Kuna, M. (1998): "Keramika, povrchový sběr a kontinuita pravěké krajiny" ("Ceramics, surface survey, and the continuity of prehistoric landscapes"). *Archeologické Rozhledy* 50, 192-223.
- Kuna, M. (2000): "Surface artefact studies in the Czech Republic". In Bintliff, J.; Kuna, M. & Venclová, N. (Eds.): *The Future of Surface Artefact Survey*, 29-44. Sheffield. Sheffield University Press.
- Neustupný, E. (1986): "Sídelní areály pravěkých zemědělců" ("Settlement areas of prehistoric farmers"). *Památky Archeologické* 38, 525-549.
- Neustupný, E. (1998) (Ed.): *Space in Prehistoric Bohemia*. Prague. Institute of Archaeology.
- Neustupný, E. & Venclová, N. (1996): "Využití prostoru v laténu: region Loděnice". *Archeologické Rozhledy* 48, 615-642, 713-724.
- Spurný, V. (1975): "Nálezový archiv Archeologického ústavu v Praze". In *Vývoj archeologie v Čechách a na Moravě II, Archeologické Studijní Materiály* 10/2, 150-158. Praha. Archeologický ústav ČSAV.