31
Concepts of informational and statistical processing of archaeological data in the computer centre of the Institute of Archaeology and Ethnography in Novosibirsk

Anatoly P. Derevianko, Yury P. Khol'ushkin, Vasily T. Voronin, Dmitry V. Ekimov, Dmitry N. Goriatchev, Vladimir V. Schipunov, Helen V. Kopteva

The Institute of Archaeology and Ethnography of the Siberian Branch of the Russian Academy of Sciences, Novosibirsk, RUSSIA.

31.1 Introduction

The formation of the computer centre in the Institute of Archaeology and Ethnography of the Siberian Branch of Russian Academy of Sciences began in 1989 when the first Amstrad XT 8088 personal computer appeared. Within a year the Institute bought another four powerful IBM AT 80286 personal computers.

In the process of using personal computers it has become clear that research can be organised on a complex methodological and methodical basis. This resulted in the creation of an informatics group in the Institute and young mathematicians were invited to participate in its work. The creation of the group made it possible to work out the general concept of the computer centre, as well as to determine its immediate and long-term tasks which are discussed below.

31.2 Computer systems at the Institute

The technical base of the group consists of eight IBM-compatible PC 80486s linked by a local network. The local network is divided into three computer complexes (for research, publishing, and for users), each of which has its own configuration. Each complex is a micro-net of 3–6 computers with external devices for high-quality input and output of information.

From the point of view of a high-quality information service, the following relatively independent cycles of regular work within the Institute can be isolated:

1. Data gathering in field conditions.
2. Formation of databases during the processing of results of field study and literature sources.
3. Processing of archaeological information using standard or custom-designed software.
4. Remote message input and output (mail, fax etc.).
5. Mastering and tuning existing and new methods, techniques, and programming aids for the Institute.
6. Organisation of international, national and regional teleconferences on archaeological and ethnographical research problems.
7. Automatic computer translation of publications in foreign languages.
8. Preparation of reports, papers and monographs for publication, issue of periodicals, pre-prints etc.

Presently only some components of the system for automatic processing of field results are used in the Institute. The *Magellan Nav 5000* provides archaeologists with the opportunity to quickly measure, process and draw the geographic co-ordinates of certain points. The device has a large internal memory (500 named course points and 1500 fixed points). Laser theodolites and seismotomographs are also of significant value. Computer processing of results enables the full realisation of the latent potential of the laser theodolite and seismotomograph for piece plotting. This excludes paper from the work and raises the accuracy of archaeological study.

The development and introduction of a system for processing archaeological materials is an independent research process. Three main elements can be isolated:

1. Creation of a base of data and knowledge.
2. Making the materials for presentation of the Institute's activity.
3. Creation of a package for information and statistical data processing.

The first and second elements were dealt with using a variety of programming tools. The third element is primarily associated with a long-term project conducted by the authors of this paper. As a result of this work, new tasks of complex computer-based processing of archaeological data have been formulated. Preliminary formalisation of these tasks and their use in research concerning the history of Palaeolithic people (Khol'ushkin 1981) have demonstrated that such tasks are common, but adequate solutions have not yet been elaborated. It was felt that a special programming package should be developed based on the preliminary methodological results of the group's work (methods of grouping, clustering, determination etc.).

To ensure that all technical and programming solutions were effective and corresponded to international standards of scientific research, an appropriate operational and programming medium was required. The authors of this paper subscribe to the opinion that the era of Microsoft Windows for the IBM-compatible PC has arrived (Duncan 1991). This powerful system has proved to be the most appropriate operational medium for our local computer network. MS Windows allows a multi-component integrated system for computer data processing in archaeology and ethnology to be constructed on its foundations. The system includes:

- text processors
- spreadsheets
- business tools and graphical illustration
- powerful database management systems
- common visual medium for information and program storage
- full hardware-independent support of the Russian language
- tools for working in local networks
- data export/import.

The following packages were chosen: Microsoft Word for Windows as a text processor; PhotoFinish; the graphics package Corel Draw 4.0; Borland Paradox for Windows for databases; Borland C++ 4.0 for developing applications; MS Windows for Workgroups 3.1; OmniPage Professional v2.0.

These tools enable the complex processing of field archaeological documentation to be carried out. Joint efforts of archaeologists and draughtspeople are required to get accurate images from material of this kind. Lack of draughtspeople causes some difficulties in recording graphical information in field conditions with subsequent implications for processing in the laboratory. One aspect of this problem is associated with the need to present the results of processing the field documentation in a variety of formats and at different levels of detail.

Currently a lot of archaeological information of this sort has been accumulated as photographs, schemes, plans and drawings (made with pencil on plotting paper). Time after time the need to make some graphical manipulation arises: to remove 'noise', change the plotting area, as well as scaling and combining drawings, for instance. A colour scanner and laser printer are used to assist in such graphical processing of drawings.

This work is another next step towards the computerisation of archaeological field work, assisting in the creation of the computer database.

General principles of archaeological analysis which are independent of the individual character of objects also need to be worked out, along with the construction of some models of archaeological procedures for successive stages of archaeological investigation. It seems reasonable to use variants of the models and procedures described in the well-known book by D. L. Clarke (1968):

- a model of an archaeological procedure as a scheme of interaction of different stages and methods of investigation of archaeological objects
- a model of an archaeological object as some statistic unity of its elements and (or) properties where differences in stochastic degree of connections with the unity are to be chosen
- a model of an archaeological process as a specific dynamic system in which archaeological items in the form of artefacts, subcultures and cultural groups are transformed in time and space

The given sequence of models of procedures in archaeology reflects the fact that for an archaeological item, material and structural components are inherent (Derevianko, Felinger & Khol'ushkin, 1989).

Modern research technology assumes the application of complex methodological, informational, programming and technical tools, concentrated in research centres. These tools are as follows:

- semantic archaeological information structurally organised
- methods of archaeological data processing (personal computers, workstations and different peripheral devices) that allow an automatic means of primary data processing
- user networks (systems of connecting research centres with places of archaeological excavation)
- mathematical tools (applying software packages)
- methods of analysis, implying the division of archaeological objects into parts and the separation of the basic characteristics of these parts
- synthetic methods aimed at a reconstruction of the whole on the basis of characteristics of its components

31.3 A system of archaeological knowledge

The above-mentioned methods, models and tools represent a necessary methodological and technological basis for the elaboration of archaeological data and knowledge bases. The idea of constructing a system of archaeological knowledge has only recently become a realistic possibility. This is partly connected with the development of computer technology such as the 80486 PC, which, along with devices for video and sound input/output, are becoming increasingly available to a great number of users at the Institute. In addition,, there is the increasing sophistication of programming technologies – object-oriented design (OOD), dynamic data exchange (DDE), object linking and embedding (OLE) – which are becoming part of everyday practice. Last, but not least, is the fact that the Institute has bought appropriate programming tools such as those mentioned above, which provide modern technologies with a high flexibility based on a combination of the advantages of newly designed standards with already existing powerful instruments.

Object Oriented Programming (OOP) implies the creation of a graphical user interface in which all the objects on the screen have their own characteristics and methods. Object-oriented methods allow the characteristics of objects to be changed, to enable them to operate other objects and to pass commands to the operating system. For example, an object might be used to start a query, to view data, to run a program or to execute a series of calculations and present the results in another object's window. Object Oriented Design (OOD) is a basic tool for solving difficult problems. It allows rapid creation of new programs and re-use of already created program elements, and makes the resulting software more reliable and better structured, and hence more useable for work and education. The application of DDE and OLE enables the user to insert objects along with procedures to operate them, even if the programs run on different computers. For example, it is possible to insert a diagram into an article or a table; the diagram reflects the result of some statistical data processing, and an alteration to this statistical data will cause an automatic redrawing of the diagram.

The problem of the complete presentation of any archaeological information is concerned with the retrieval of extremely detailed information about any object along with a set of appropriate graphical and/or video presentations. One solution would enable various measurements to be done, as well as the scaling, projecting and grouping (pasting together, in particular) of different three-dimensional objects. The application of Object Oriented Database (OODB) tools mean that a database management system can be realised that will meet such requirements, but, given the use of video data, at the expense of an enormous amount of data storage which may cause problems. One of the ways of solving this problem is to develop a client-server structure with two varieties of remote data storage:

- hard disk drives, currently used for the storage of changeable information and for the insertion of the results of recent investigation)
- laser compact-disks or CD-ROM, for storing unchanging components of data, such as the graphical presentation of artefacts, pictures of rock drawings, or articles issued.

A development of the united information network of the Institute (with its branches in Krasnoyarsk, Omsk, Irkutsk, Tomsk, Blagoveshensk and in other cities of Siberia), will enable the streaming of the whole body of information according to various criteria and regional features and, consequently, a reduction in server loads. Other structural elements of the client-server model are also advantageous, such as multitasking, transfer of the minimum necessary information across the network, and support of data integrity and data protection

The database management system Borland Paradox for Windows (BPW) has been chosen as a powerful and convenient tool for developing this information system for archaeological research. It has a basic set of objects with preassigned characteristics and offers the ability to change these characteristics, as well as to find relations between objects. In interactive mode BPW allows the creation and modification of table structures (types of fields, default values, rules of correct input, and so on), queries (relations between tables, visible fields, filter conditions etc.), and screen and printed forms and reports.

The system, which is presently being created, consists of three parts:

- the kernel, consisting of the set of databases (sites, artefacts, institutes, archaeologists, publishing firms and publications) and the set of classifications for databases and the geographical base maps
- the user interface, based on a multimedia design
- the programmer interface, which provides the facility to develop necessary processing or statistical procedures and to add them into the user interface

The user interface is primarily a means of generating hypermedia queries, to enable a user to get additional information about some object, system or term along with alternative graphical presentations, video and/or audio records.

Two levels of work with objects are provided: internal and external. Operations with objects on the internal level carried are out by limited queries. These allow a user to construct extracts on the basis of some conditions of geographical, chronological and/or classification character, and provide a variety of different procedures for outputting the results. On the external level, the objects are processed by an external procedure and results can be represented both by the external procedure and by the system's own tools. For example, a refitting procedure for flakes makes it possible to reconstruct a core after the required objects have been retrieved from the database. Editing work can be carried out using interpolated 3D frames (which are sometimes more demonstrative since one can see the back of an object), and then the whole picture is viewed. The internal level tools are mainly used for this, together with external procedures for the interpolation of 3D objects and for working with 3D frames which are loaded as required.

This integrated archaeological database creates the necessary conditions for developing the informational and statistical methods for data processing. It seems reasonable to consider a database as a kind of information space with traits, features and other characteristics of archaeological objects being its elements. For this space a metric is suggested. It characterises a semantic similarity of concepts, terms, and categories used in archaeology. Using the metrics, computing models and algorithms for isolating the points where the similar objects cluster are constructed. These algorithms are constructed on the basis of similarity of isolated or group traits and features of the objects. Along with standard methods of cluster analysis, the distances between objects are calculated on the basis of the algorithms and schemes. These distances characterise their informational similarity and statistical indices reflecting the degree of structurisation, vagueness and interrelationship of traits and features of sets of objects or their subsets.

Thus, the general scheme of computer technology for research is apparent. It includes collecting and preliminary processing of the field database, preliminary processing and input of this information into databases in laboratory conditions, transformation of these data to more convenient form for subsequent use, formation and realisation of research' queries, informational and statistical data processing, and preparing data for publication.

Elaboration of data bases of this kind can, with good reason, be regarded as a real prelude to creation of a unique 'machine-readable archaeological encyclopaedia'. As a store of specialist archaeological knowledge, this encyclopaedia can be used as the basis for developing expert systems, which may be considered as formal mechanisms which conform to the rules irrespective of the object of computer analysis.

References

CLARKE, D. L. 1968. *Analytical archaeology*, Methuen, London.

DEREVIANKO, A., FELINGER, A., KHOL'USHKIN, Y. 1989. *Metody informatiki v archeologii kamennogo veka* (Methods of informatics of the Archaeology of the Stone Age), Nauka, Novosibisk.

DUNCAN, R. 1990. 'Redefining the Programming Paradigm: the Move Toward OOPLs', *PC Magazine*, November 13.

KHOL'USHKIN, Y. 1981. *Problemy korrelatsiy pozdnepaleoliticheskih industriy Sibiry I Sredney Azii* (Problems of correlatiction of the Late Paleolithic Industries of Siberia and Central Asia), Nauka, Novosibisk.