Introduction

Computer-based databases for archaeology have continuously been developed since the earliest applications in the 1980s (Ozawa 1985; Richards and Ryan 1985). It is noteworthy that the CAA 1991 were characterised by the papers associated with SQL-compliant...
relational database management systems (RDBMS) (Cheetham 1992; Ryan 1992) and the bibliographical database for archaeology (Heyworth 1992). Since the mid-1990s, geographical information systems (GIS) have increasingly been incorporated into archaeological database projects (Lock 2000; Yokoyama and Chiba 1997). Using these systems, database technologies further developed and diversified during the first decade of the twenty-first century. The implementation includes: 1) clearing houses to aggregate different online data sources and database systems (Usui et al. 2000), 2) autonomous decentralised Internet GIS (Miura and Ozawa 2000), and 4) object-oriented database schema (Conolly and Lake 1996,55; Lock 2003,89–90) using XML (eXtensible Markup Language) (Jordal et al. 2010), UML (Unified Modelling Language) (Usui et al. 2006), and standardised geospatial information (ISO 191xx series) (Fujimoto 2008).

Today, databases are widely employed in the field of archaeology right from local fieldwork and laboratory work to supra-regional cultural resource management. With regard to Palaeolithic archaeology, which is taken as a case study of this paper, the large-scale site databases of Africa (Märker et al. 2009), Europe (van Andel and Davies 2003; D’Errico et al. 2011), the Far East (Gillam et al. 2005), South Korea (Choi et al. 2006), and Japan (Japan Palaeolithic Research Association 2010) have already been published. Such an archaeological database has usually been edited in the closed, offline environment. However, this practice is less effective for the recent ‘inter-institutional’ projects where researchers from different institutions collaborate to achieve new scientific outcomes within a finite period of time. For instance, the method wherein researchers edit their own versions of a database in the local environment and merge them afterwards may cause a discrepancy in the data field, classification, description, and input rule. It is also too risky to modify the structure of a database once a template is distributed and employed individually. In light of these problems associated with the use of a closed database, a networked system, in which a number of users can simultaneously access, edit, and share a common master database is in high demand.

Fortunately, recent rapid progress in the computing environment, exemplified by multi-core processors, terabytes of storage, gigabytes of network communication, and sophisticated API (Application Program Interface) technologies, has enabled such network computing. This paper presents the authors’ Neander DB project as a practical case study of a network-based and scalable archaeological database and then discusses the significance of network computing for archaeology.

**Neander DB and Network-Based Database Editing**

**Research organisation**

In 2010, a five-year multidisciplinary research project, the ‘Replacement of Neanderthals by Modern Humans’, was launched in Japan. The objective of this project is to explain the distinction of Neanderthals (Homo neanderthalensis) and the diffusion of anatomically modern humans (Homo sapiens, hereafter called AMHs) in terms of their differential learning abilities (Akazawa 2010). It comprises six research groups specialising in archaeology, cultural anthropology, evolutionary modelling, palaeoenvironment, comparative anatomy, and neuroscience (Akazawa 2010; Nishiaki 2011; Terashima 2011). The authors, as part of the archaeology group, are working on creating a database of the palaeoanthropological sites where either human fossils or lithic artefacts of Neanderthals (Mousterian lithic traditions) and/or AMHs (such as Aurignacian, Gravettian, and Solutrean traditions) have been unearthed.
The goal of this sub-project is to visualise the spatio-temporal process of the replacement of Neanderthals by AMHs in a higher resolution than in previous projects (for example: van Andel and Davies 2003; Banks et al. 2008).

The database project covers a long temporal range from 200 to 20kya and a wide geographical zone from Africa to Eurasia where either the replacement event itself or the evolution of AMHs took place. The principal data source is excavation reports written not only in English but also in French, German, Spanish, Russian, and other languages. In order to extract the necessary information from these diverse data sources, the target area is divided into: 1) Africa and the Middle East, 2) Western Europe, and 3) Eastern Europe and Northern Eurasia, and a pair of archaeologists with professional knowledge and language skills are in charge of each region. The primary author is acting as the database manager. In total, seven researchers based at six different institutions are working in collaboration.

**Structure of the main database**

The database team intends to minimise the time spent for data collection so that they can start with the spatio-temporal visualisation and analyses of the ‘replacement’ events as early as possible. To this end, the team employs a client-server network database system – *Neander DB* – that allows inter-institutional collaborators to access, edit, and share one master database through the Internet in order to save time and avoid redundant errors.

The database system is simple and classical.
A database server is placed at the University Museum, University of Tokyo (Fig. 1). It is a Mac mini, operated by Mac OS X Server 10.7, with a 2.66 GHz Intel Core 2 Duo processor, 8 GB RAM, and a 1 TB disk drive. The database is controlled by FileMaker Server 11 and is password and firewall protected. It is open to access by authorised clients using FileMaker Pro 11 for Windows and Mac OS computers and FileMaker Go for iPhone and iPad. The users can access the master database at any places where the Internet is available (office, home, library, and conference hall for example). The Neander DB is not a so-called ‘cloud database’ because the data is stored in the team’s own server.

The database scheme was designed to fit the typical workflow of archaeologists to extract from an excavation report: 1) bibliographical information (title, authors, year of publication), 2) fundamental information of the site (such as toponym, latitude, and longitude), 3) detailed information on each cultural horizon, and if any 4) information on radiometric dating (Figs 2 and 3). The graphical user interface (GUI) was also designed to follow this flow and to assist non-expert users in operating intuitively with text autocomplete, pull-down menu options, and on-click scripts (Figs 3–9). These devices also help reduce errors and redundancies.

From the viewpoint of RDBMS, the database contains two different entities of information – attributes of the archaeological site and bibliographical reference. In terms of a database entity, an archaeological site can be divided into 1) one record of the fundamental information (Fig. 4), (2) one or more \([1\ldots n]\) record(s) of descriptive information on each cultural horizon or layer (Fig. 5), and 3) (if any) one or more \([0\ldots n]\) record(s) of descriptive information on radiometric dating (Fig. 6), associated with a layer. Therefore, by means of a unique identifier, records of a layer are related to the record of the site in a many-to-one cardinality, and then, records of radiometric dating are related to the record of the layer in a many-to-one cardinality. The relationship between site and bibliographical reference (Fig. 7) is not always in a one-to-one cardinality: in some cases, the excavations at a site may be reported in multiple books or articles; in others one report may contain information on more than one site. In other words, a table of site is connected to that of bibliography in a many-to-many cardinality, which could not be managed by RDBMS. Therefore, a table of citation (Fig. 8) is inserted to explicitly define the one-to-one relationship between the record of a site and that of a reference.

**Fundamental information of a site**

Fundamental information of a site consists of toponym, geocoordinate, location (cave, rock shelter, or open site), and a unique identifier (Fig. 4). Of these attributes, geocoordinate, that is, longitude and latitude, is most important.
for GIS-aided research: without the relevant coordinate it is impossible to plot point features of a site onto a GIS-based map, which is the only way to visualise the spatio-temporal process of an archaeological event. A highly accurate location allows a spatial analysis in higher resolution. Thus, it is desirable to record geocoordinates as accurately as possible – ideally as one arc second (approximately corresponding to tens of meters in the middle latitude zone). If such a precise value is unavailable, we have to identify the longitude and latitude of the site by reading maps in the excavation report with reference to georectified maps and satellite imageries provided by Google Earth or other GIS applications. This procedure requires a certain amount of time, remote sensing skill, and experience and thus it would be rather difficult for archaeologists to do themselves. Therefore, the team takes advantage of a network database in which multiple editors can simultaneously input data. If the editors find it difficult to identify the geolocation, they are asked to upload scanned maps. A GIS and remote sensing specialist identifies it for them.

Descriptive information of cultural layers

Descriptive information of cultural layers provides a main component for the subsequent data analyses. The record of a cultural layer contains five chronological indicators: 1) name of the layer (Layer 1, Level II, and Phase 3, for instance), 2) name of the chronocultural entity, or lithic industry (Mousterian, Aurignacian, and Gravettian, for example), 3) the marine isotope stage (MIS, formerly called the oxygen isotope stage or OIS), 4) name of the local chronological indicator (such as specific palaeosol and tephra), and 5) absolute date [ybp] assessed by radiometric dating results (Fig. 5). These different indicators are employed to relate the archaeological chronology with the palaeoenvironmental one using different terminology. A record of a layer also contains a summary of the unearthed materials such as palaeoenvironmental samples, human fossils (Homo neanderthalensis and Homo sapiens), and symbolic artefacts (carved bone objects, rock arts, beads, pendants, and ochre, for instance).

Radiometric dating

Radiometric dating is a critical technique to observe the diachronic distribution of
archaeological sites. The information on dating comprises: 1) laboratory number as an unique identifier, 2) mean age and its standard deviation [ybp], 3) dating method such as AMS radiocarbon, thermoluminescence (TL), optically stimulated luminescence (OSL), electron spin resonance (ESR), and uranium series (U-Series), 4) type of the sample (such as burnt lithics, bone collagen, and shells), and 5) taxon of the sample, if identified (Fig. 6). These data will serve for future reassessments of chronology.

Bibliographical reference

A typical bibliographical reference includes name of the author(s), publication year, title, place of publication, publisher, journal or series title, volume and issue number, pages, and a unique identifier (Fig. 7). The style of bibliography follows the international standard ISO 690-1. Unique identifiers such as ISBN (International Standard Book Number) and DOI (Digital Object Identifier) allow editors to skip inputting the detail because it is easy to specify a unique source by using an online search engine for academic literature, such as Google Scholars and ISI Web of Science. For non-English literature, the name of the author(s) and the title are translated into English for convenient retrieval. The editors are asked to upload PDF files if available, in order to share them with other members. In the near future, the PDFs will be integrated into a digital repository of the project, which will be managed by the National Institute of Informatics (Mori et al. 2011).

Wiki-like encyclopaedia of the lithic industry

In addition to recording the information on sites and radiometric dates, the database team is compiling an encyclopaedia of lithic industries. This sub-project is intended to facilitate an inter-regional and/or diachronic comparison of lithic traditions. Such a comparison has
been demanded for a long time but is difficult
to carry out because the definitions, in terms
of typological (Bordes 1961), technological
(Inizan et al. 1992), and behavioural (Torrence
1989) aspects, are too diverse. Therefore, we
plan to first collect the original descriptions
and then analyse them by means of ontological
approaches such as morphological analysis
and network graphing, in order to quantify
and visualise the similarity, difference, and
ambiguity in lithic industries.

This encyclopaedia is incorporated into the
Neander DB as a wiki-like knowledge base.
The definition and characteristics of each lithic
industry are quoted from the original texts or
summarised by the specialists referring to the
typical specimens from representative sites
(Fig. 9). The quotations and summaries are
explicitly distinguished. Description follows
the processes of production and includes
information regarding: 1) raw materials, 2) core-
reduction technology, 3) retouch technology, 4)
hafting methods, 5) tool-maintenance, among
others. Similar to the relationship between site
and bibliography, a table of citation is inserted
as a hub between a record of lithic industry and
a bibliographical reference.

Discussion and Conclusions

This paper has reviewed the scheme and
practical application of our Neander DB, a
network-based RDBMS for Neanderthal and
AMH sites in Africa and Eurasia. For a year since
November 2010, approximately 3,400 layers of
1,255 sites, with more than 4,700 radiometric
dates including European data published by
the Stage 3 Project (van Andel and Davies
2003) and PACEA (D’Errico et al. 2011), have
been recorded in the Neander DB. There is no
doubt that the RDBMS with advanced network
computing technologies has successfully
facilitated inter-institutional collaborators to
assemble and share one master database. It has
contributed towards not only reducing errors
and redundancies in database editing but also
facilitating database maintenance. In fact, minor
revisions of the database scheme, GUI, and
hardware have frequently been carried out as
per users’ requests without any troubles in data
backup and versioning. Continuous updates
of the system and contents are provided. The
database will be open to public access after the
scientific achievements are published.

The Neander DB is characterised by an
explicit relationship between bibliography,
archaological records, and wiki-like
encyclopaedia of archaeological objects. Such
an integrated system is useful to reorganise
data in a flexible manner, and it may broaden
opportunities to discover overlooked
relationships in archaeological concepts. In the
case of the lithic industry encyclopaedia, we
will be able to clarify the similarity, difference,
and ambiguity between lithic industries using
ontological approaches.

The collected data are exported to geographical
information systems (GIS) for on-demand mapping (Fig. 10). The maps are ready to be
published online through the ArcGIS Server
hosted by the Center for Spatial Information
Science (CSIS), University of Tokyo. In the near
future, the spatio-temporal distributions of sites
will be analysed with agent-based evolutionary
models and a palaeoenvironmental dataset
including elevation, slope, distance to water
sources, temperature, precipitation, and
vegetation, provided by other groups of this multidisciplinary project. It is expected that large-scale data mining in combination with different data sources would enable us to explore unknown archaeological patterns so as to reconstruct a detailed spatio-temporal process of the replacement of the Neanderthals by AMHs and to discover possible explanatory factors that differentiate the ecological niches and behavioural strategies of these two human species.

Through this database project, the authors have gradually noticed that network computing can potentially change the way of archaeological thinking itself. First, as pointed out above, it enables inter-institutional collaborations to expediently develop new models which explain an archaeological phenomenon. Second, the database-oriented data processing lets us canonicalise and quantify the attributes of archaeological objects such as sites, built structures, and artefacts; moreover, it also provides us with an opportunity to rethink the definitions of the objects themselves. Third, an integration of library-oriented systems such as bibliographical databases and museum-oriented ones like a relic catalogue and site database allows archaeologists to retrieve and reorganise data more quickly and effectively. These advancements may contribute to the discovery of new research issues and ideas for archaeologists in the next generation.

**Acknowledgements**

The Neander DB project is undertaken by the research group ‘Archaeological Research of the Learning Behaviours of the Neanderthals and Early Modern Humans’, directed by Yoshihiro Nishiaki, as a part of the planned research project ‘Replacement of Neanderthals by Modern Humans’, directed by Professor Takeru Akazawa (Kochi University of Technology). It is financially supported by the Grant-in-Aid for Scientific Research on Innovative Areas, Ministry of Education, Culture, Sports, Science and Technology, Japan. An early version of this paper was presented as a poster at JINMONCOM 2010, an annual conference on Humanities and Computer Science, held at Tokyo Institute of Technology on December 11 and 12, 2010. The authors are grateful to two anonymous reviewers for their constructive comments. The specification of the database is as of November 30, 2011 and is subject to change.

**Bibliography**


