The computer as a tool in a dialectical approach to data processing

Maria Shinoto
Institut für Ur- und Frühgeschichte
Marstallhof 4
69117 Heidelberg
http://www.uni-heidelberg.de/institute/fak8/ufg

Abstract: For a study of sites and artefacts in a small region of South Japan, the author developed a combination of photography and a relational data base that provides tools for data input, analysis, output and manipulation. New results from analysis in the course of an investigation can change research design within the data base continuously. At the same time the structure is kept consistent. The most efficient part of the data base in this sense is the 'artefact area' with its flexible treatment of artefacts, attributes and classes. The present article focuses on this aspect.

The hard- and software equipment is shortly introduced. Basic theoretical ideas about observations, properties and attributes and their significance for classification shall explain the structure of the 'artefact area' of the data base with respect to statistical, archaeological and technological standards. Two representative examples of application from the case study in Japan are presented.

Key words: relational data base, normalization, continuous adaptation of research design, classification, numerical taxonomy, cladistics, consistency, site, artefact, pottery, Japan.

Introduction

The problem

Collection, analysis and categorize, visual and verbal documentation of data or pictures of artefacts and sites is one basis of our discipline. Although there is a natural sequence in these activities, each step is related with the others. For example, the "analysis" step both influences the later step of ordering and acts as a feedback mechanism on the earlier step of collection. This is the reason for a systematic research problem: how can we adapt our research design to our knowledge, which is inevitably changing in the course of a research project?

Up to now research projects have followed more or less either of two principles. The researcher can start with knowledge from earlier studies in the field and go on, intuitively adapting research strategies to what seem to be new insights or knowledge. From a systematic point of view, this is scientifically not satisfying, although talented researchers have gained great results with studies of this kind. The other, scientifically more acceptable approach is to start with a well defined research design based on knowledge from earlier studies, then collect data, and finally analyze the data in a separate step. A well known epistemological problem becomes more severe than in the "intuitive" approach, since research in a preset framework can give new insights into details, but results of a new quality will be difficult to produce. What is more, the researcher has to wait until the end of the study to understand whether data collection did give new insights or was at least valid.

During a study in South Japan, I encountered this problem when collecting site and artefact data from a small region in order to investigate the settlement activities during several centuries. The study (Shinoto in prep.) is referred to as the "case study" in this article. In the course if the study a solution to the problem had been developed which is mainly based on a data base that controls a continuous modification of the data structure and the research design - both during data collection and afterwards - on a scientifically acceptable basis. This concept has parallels in epistemological ideas like hermeneutics and dialectics, in the realm of physics the idea of self-consistency is utilized in a similiar fashion, in statistics it would be the Bayesian approach. On the one hand these are "handy slogans", easy to misunderstand, but on the other hand they can serve as a starting point for a discussion about whether this approach is appropriate in archaeological research or not (see Shinoto in prep.). For the present article it should be sufficient to summarize that permanent evaluation of research design on the basis of new data is not only most fruitful but also well accepted in other fields of science.

This article cannot present the whole system. Instead I shall focus on a central part of it, the collection and modification of data about artefacts, in order to clarify the idea proper and the task of the computer with examples from the case study. In other words: the system serves a wide range of purposes, but this article focuses on the problem of artefact classification.

Classification background

During the last 40 years, methods based on numerical taxonomy - a form of classification that was predominant in biology during the sixties of the last century - have become predominant in archaeological classification (e.g. Clarke 1962, Clarke 1978 (1968), Hill & Evans 1972, Hodson 1966, Kampffmeyer et al. 1988, Stehli & Zimmermann 1980; see also Dunnell 1971, Whallon 1972, Whallon & Brown (eds.) 1982). At the same time, traditional approaches (Vossen 1970) continued to be used or were further systematized (Adams & Adams 1991).

Adaptation of numerical taxonomy from biology to archaeology was thought to lead to a scientifically satisfying alternative to traditionally more intuitive approaches. But in the late sixties and the seventies, numerical taxonomy lost its dominance in biology and was nearly replaced by another classificatory method called cladistics, which in its original form works without statistical analysis. Both methodological trends have their advantages and disadvantages, they have undergone substantial changes in biology since then (Mayr 1990, Panchen 1992). It is not as easy to transform the ideas of cladistics to archaeology as it was with numerical taxonomy, though it is a very stimulating task. A first attempt concerning methodological and theoretical aspects will be published in another paper (Shinoto in prep.).

For the present paper it should be sufficient to maintain that numerical taxonomy and its developments are not the only scientific way to archaeological classification, there are different models of at least equal scientific standard. The alternative approach deduced from biology needs a clear definition of attributes that can be utilized for classification and their close observation. A clear idea about what is the explanans and what is the explanandum in the classification is inevitable. Evaluation of the observations may involve statistics but does not necessarily do so.

For the artefacts of the case study - mainly pottery of rough making called Narikawa pottery - classifications already exist. Two classifications show a more traditional approach (Ikehata 1980, Tatara 1981) the other is more systematic though it is not a numerical taxonomy (Nakamura 1987). All classifications are based on profile description and cover more than 400 years divided into 4 to 5 stages. To achieve a more detailed classification with information about more than just chronology, the case study had to start with new ideas about research design and new views on the pottery. Therefore, systematic studies on pottery (e.g. Dohrn-Ihmig 1976, Gardin 1976 (1985), Kampffmeyer et al. 1988, Karstens 1994, Orton, Tyers & Vince 1993, Sablatnig & Menard 1996, Shepard 1955 (1980), Traunecker 1981) were adapted to the case in South Kyûshû in order to get a more vivid idea about the ancient society (Shinoto in prep.).

Prerequisites

This article does not comment on the basic rules for structuring data in statistics (see e.g. Shennan 1988 (1990) or any introduction to descriptive statistics). It does not explain the terminology of relational data base theory, the customary drawing of relations among tables, nor does it explain the rules for normalization (see e.g. Adams & Beckett 1997 or any introduction to relational data bases).

System features

Photographs

In the case study, photographs were made instead of drawings. This was one important factor for developing new ideas about the material, mainly pottery of rough making, known as "Narikawa-pottery". Drawings of artefacts can be integrated into the system for documentation and further analysis. On the one hand, photographs are closer to the original, but not a good basis for structured analysis because of their redundancy of individual information. On the other hand, drawings rarely offer new insights since they are standardized with respect to certain traditions, only those features are recorded that are of interest in the framework of this tradition.

The photographs of the case study became a source for new insights and an understanding of the pottery that could not be gained with traditional drawings. The redundancy of individual information was avoided to a certain degree by taking photographs from standardized positions so that photographs of different objects from the same position could be selected in the data base and visually compared. To achieve more than this information, additional spontaneous photographs were taken where it seemed to be of interest (Shinoto in prep.).

Database

The relational data base is designed in order to record information from different regions, sites, documentation in literature, excavations, artefacts and the places where artefacts are stored. The relation consists of several tables which refer to "areas of content" (fig. 1). "Log tables" are not yet implemented; they shall create text data of each data manipulation and structural changes - which is only roughly done by hand at the moment. One area is concerned with the location of the site. Information about administrative units from hamlet up to the national state can be stored in several tables. These are related to tables with information about the natural geography like traditional landscapes, river systems or mountain ranges. A table for maps and automatic mapping may be added in the future and serve as a GIS module together with the analysis tools.

The central "site" table of the "site area" is related to the location tables. Informations related to sites are divided into further tables concerning excavation campaigns and excavation units like layers, artificial units like excavation grids or anthropomorphous house pits etc. - The "artefact area" is related to these excavation units and further divided into attribute and class tables and tables that store pictures or other non-verbal information about the individual find. - The "time" area is concerned with relative and absolute dating.

Information about storage of artefacts could be related to the artefact table - in order to record information about storage for each artefact individually. For practicability, I related the storage area to the site table. The mass of artefacts is stored at the same place, in the case of an exception this can be mentioned separately in a "comment data field". The same solution was chosen for the documentation tables, which store information about notes, protocols, articles or books. It may make sense to relate these two areas to excavation campaigns in the future rather than to the whole site.

The system is designed for international research. Proper names are stored, sorted and searched in the characters of the original language and additional columns (data fields in the actual record of a table) for transcription into Latin characters and for alternative namings. This feature was designed for the situation in Asia. But since anywhere in the world research most often covers several countries, and since places are named differently by different groups, this may be a helpful feature elsewhere as well.

The equipment

The system was developed with standard software that is available on a Macintosh. It can be used on the Windows platform with the same software, which is unfortunately not the case with Linux. When the project started, analogue photography with scanning offered the better quality particularly as far as colour management is concerned. For any new case study to be started nowadays, digital technology should be used in order to reduce cost and increase speed and quality of the research (details in Shinoto in prep.). The whole equipment is easy to transport and can be used on site.

The data base is the centre of the system, it serves data storage, manipulation, analysis. In addition, via ODBC or data export it provides data for more sophisticated analysis. In the future these may be integrated into the software package. The system was developed with the DBMS "4th Dimension", starting with version 3.5. At the moment it is under further development using version 6.58, while the latest available version is 6.71.

This article shall present the ideas as independent from the underlying software as possible. The final aim is to create a software package that is compiled as a stand-alone application which does not need any programming by the end user. When working with the program, the user creates one separate data file for his individual case, independent of the subject of the study. - The aim of this article is to explain the basic ideas of the system in order to enable the reader to create a similiar solution with his own DBMS.

First normalization of artefacts and attributes

One may create columns for attributes in the artefact table. But archaeological practice and the theory for relational data base design shows that this is inappropriate. Using the rules for normalization, one can create a structure for artefacts and attributes that is archaeologically sensible and practicable. - It is important to distinguish attributes and attribute states (Clarke 1978 (1968)) in this context, where each attribute represents a generic term with the attribute states as its concrete manifestations, varying from artefact to artefact.

When designing only one table (fig. 2 a) for artefact and attributes together, different fields will stay empty for each record. If artefacts of different kinds are recorded, like stone tools and pottery, each record has to offer attribute fields for both kinds. In every record the fields concerned with the other kind of artefact will stay empty. In addition, artefacts are not all preserved well. Pottery in particular is most often found as sherds from different vessel parts. Therefore attributes referring to vessel parts that are not preserved cannot be recorded.

To keep the database slim, one may exclude the columns for attributes from the artefact table, creating separate tables for each attribute instead (fig. 2b). If a certain attribute state is observed for an artefact, a record is created in the respective table. Since every attribute requires a table, the data base structure has to be changed for different cases.

The solution in fig. 2c is more elegant. Unique artefact data like ID numbers are recorded in the "artefact table", attribute states are stored in one "attribute states table", regardless the attribute concerned. Several attribute states are related to and specified by one attribute which is recorded in the "attribute table". This forms an n:1 relation between the tables for attribute states and attributes. The number of attribute states is most often limited. Therefore, one attribute state can be assigned to several artefacts, forming a 1:n relation. But at the same time an artefact most often shows more than one attribute, thus forming a 1:n relation with the "attribute states table". Together, both artefact and attribute state form a n:m relation. - In real life, this n:m relation requires an additional table in order to form a 1:n and a m:1 relation as fig. 7 will show.

This data base structure needs no change in the course of research, because attributes and attribute states can be added as records by the user rather than added as tables by the programmer. However, attributes are poorly structured and attribute states are difficult to sort and search.

The same ideas apply to what is called "classes" in this article and may be referred to as "types", "groups" or even "chronological stages" in other contexts. A classification is similar to attributes with the classes being the equivalent for the attribute state. The class name can be stored in the attribute table like an attribute. In many cases, several classification systems exist for the same groups of artefacts, and if we do not wish to favour one, we have to store several class names for one artefact. Again, all these can be stored in the "attribute state table". - When speaking of "attribute tables" and "attribute state tables" in this article, classifications and classes are included.

The nature of attributes

The model proposed is too simple for both archaeological practice and computer technology. The archaeologist as well as the computer needs structured data to work with. In the following sections I will discuss the nature of archaeological observations, the significance of observations for archaeological classification, the nature of variables in statistics as well as the rules for normalization and the basis for encoding data in a computer. This shall lead to a model of an intelligent data base structure that is not limited to a certain case. In other words: the software package shall be applicable to any new project without new programming.

Observations: attributes, properties and classification

Data to be recorded are first of all observations, structured or unstructured. Observations may form attributes or properties. Both may be significant for classification or not as long as they show a certain structure (fig. 3).

Occasional features are recorded as unstructured observations. As soon as they are observed more or less regularly, they should be reconsidered, structured and changed into a structured observation. Observations may concern attributes or properties. Attributes are those features that are formed by intention or without intention as a by-product of intentional work in the course of the artefact production. Thus, they may reflect the intended usage, fashions or stilistic traditions, technological standard and availability of resources - to mention just a few matters that are of interest for classification. The origin of properties is not limited to the stage of production, they can also occur as a result of usage or deposition in the earth. If they occur during the stage of production, they do not reflect intentions of the producer but unintended matters like faulty production. In most cases properties are observed occasionally. If they occur frequently, particularly as usage marks, they may become of interest for classification as well. - In the case study, a certain vessel form is used for cooking in open fire and for storage. Usage marks like soot on certain vessel parts help to understand the usage of the individual pot (e.g. Uno 1999). Comparison of pots with soot and those without helps to find functional differences reflected in attributes that were overlooked before. Besides distinguishing attributes and properties one may distinguish between observations that are significant for classification and those that are not. Of course, unstructured, accidentally occuring observations cannot be significant for classification, but they are nevertheless worth being recorded in the data base and considered for interpretation.

Only few observations reflect the intention of the producer immediately and are as such significant for classification. Some structured attributes are too detailed. In these cases, the intention either has to be deduced from several specific observations, forming a "deduced attribute" or it has to be reconstructed by unifying and simplifying several detailed observations to one "simplified attribute". Two examples may clarify the idea (fig.4).

Fig. 4 shows sherds from five pedestaled bowls, the bottom of bowls from below in the three examples above and two feet below. Two bowls on top still show a clay ball in its centre that functioned as a connection between bowl and foot. So we can record the observation "clay ball" as well as a deduced attribute "adherence technique" with its attribute state "with a clay ball". The bowl in the centre shows a hole instead of the ball, so for the same vessel part we cannot record the observation "clay ball", but the same deduced attribute state "with a clay ball" for the attribute "adherence technique". For classification, only the deduced attribute is of importance. It builds a bridge to another part of the same vessel form: the two examples below show the feet of pedestaled bowls with a hollow on top. It is the trace of the clay ball that connected bowl and foot. Therefore, in case of the feet we observe a third morphological attribute that represents the same deduced, significant attribute for classification. As for the producer of these vessels, she (several studies let us suppose that they were women) made the clay balls intendedly, holes and hollows may be the unintended side effect, but these observations also reflect her intention to connect both vessel parts with a clay ball.

Beside deducing a significant attribute from several observed and unsignificant attributes, simplification is another important method - as the example "temper" may show. In the course of material collection, one may count and measure temper grains and record their colour or shapes and so on. But this may not lead to a significant attribute for classification, since most often it was not the intention of the producer to add a certain proportion of grains of certain colour and size or shape. Temper was chosen from available resources, perhaps mixed or sieved and added in certain amounts to the clay. It is the task of the archaeologist to find the regularities in the many insignificant but detailed attributes to form groups that reflect the intention of the producer who wanted to produce a certain kind of fabric. Even these simplified temper groups may be too detailed to be significant for classification, since the final intention was to form a certain fabric. The correlation of temper groups with other features of the fabric, like porosity, may be investigated to finally form a significant attribute "fabric".

The last example indicates another basic problem. How do we know that simplification is at its end and that the attribute is significant? Classes are further simplifications of attributes, the border between attribute and class is fluid. As soon as attributes from different realms like "decoration" and "technology" or as soon as properties are integrated, one may speak of classes. Classes may vary with respect to the purpose and nature of the classification, one artefact may refer to several classes at the same time (Shinoto in prep.).

Structure in archaeology, statistics and data bases

Attributes and attribute states have their counterparts in statistics: the variable with categories or values. "Sherd colour" or "rim diameter" are attributes in the realm of archaeology and variables in the realm of statistics. "Red" or "21 cm" would be examples of affiliated attribute states in archaeology, called categories or values in statistics. This structure can be transformed to the structure of relational data bases. This process is called normalization, the structure is reflected in a table for the attribute and another table for the attribute states (fig. 5).

The form of structured attributes varies, as the example from sherd colour and rim diameter above has shown. The variation is structured; like statistical variables they can be on nominal, ordinal, interval or ratio scale. In the case study, I did not encounter attributes on interval scale, for several reasons they should be treated together with attributes on ratio scale. Attribute states on different scale have their counterparts in different data field types of the columns of a data base table (fig. 6).

Observations without statistical scale are best recorded in text fields or alternatively in picture fields or BLOBs. Attribute states on ratio scale are best encoded in data fields of real type, although integers are another possible solution. Attribute states on nominal scale are best encoded in alphanumeric data field types. Attributes on ordinal scale have to be treated different. In the present study, these attribute states are presented in the layouts as alphanumerical variables like "small", "medium", "large", but recorded as an integer "1", "2" or "3" in the respective data field, encoded automatically by the program. This solution is helpful when sorting items with respect to their rank or analyzing them with methods that need data on ordinal scale. While the user can name the attribute without regard to the alphabetical order, the computer is always informed about the rank from the affiliated integer value.

Towards final normalization

The above considerations lead to a new design of the "artefact area" in the data base (fig. 7).

Unstructured observations are stored in the "attribute states (observations) table" and related to keywords in the "attribute without scale table" that may lead to structured attributes in the future. If not, the keywords help to get a better overview of the unstructured observations by searching and sorting. Each keyword forms a record with definitions and comments. Observations on ratio scale are another simple case in terms of normalization. The definition of the attribute is stored in the "ratio scale attribute table", the related 1-table to the "attribute state table". Since attribute states on ratio scale have no previously defined values, each value is stored individually in the "attribute state table", which is again the related n-table to the "artefact table".

The situation is more difficult with attributes on nominal and ordinal scale. In contrast to observations without scale and on ratio scale, here the names and values of the attribute states have to be defined before data input starts. In these cases the values of the attribute states are recorded in the "attribute state table" at the same time when the attribute proper is defined in the "attribute table". Each attribute state forms one record with additional definition.

It has become clear from the above explanations that it is necessary to define attributes and record them before recording of attribute states for the actual artefacts starts. When recording attribute states for a certain artefact in the next stage of research, the user chooses the respective attribute from a drop down list. The computer creates this list from the actual "attribute tables" with regard to the scale of the attribute. If the attribute chosen is on nominal or ordinal scale, the user can choose the attribute state from another list. This list is created from the related records in the "attribute states tables". After the attribute state is chosen, the data base creates a new record in the "attribute states list table" which is the related n-table to the "artefact table". For attributes without or on ratio scale the user can record individual informations and values directly in the "attribute state tables". In other words, the information about the attribute states of a certain artefact is stored in the tables "attribute state list" for nominal and ordinal scale, in the tables "attribute state" in the other cases.

Let us return to the definition of attributes. Ratio scaled data need a unit of measurement which is set in the "attribute table" beside the definition. Values concerning size often consist of up to three dimensions which should not be divided into separate attributes. Therefore the definition consists of three dimensions. This is an important feature for irregularly formed attributes as well, where the dimensions can represent minimum, maximum and normal extension. It was a very important feature in the case study, where rim diameter of the same pot could vary between 27 and 33 cm, to mention just one example.

All tables directly related to the artefact show an additional column called "parameter". This is necessary to explain where the respective attribute is observed. As for diameter, one could form several attributes for rim diameter, neck diameter etc., but in this data base another approach was chosen. "Diameter" would be the attribute, "rim" would be the parameter. Parameters form lists with different values. They are defined at the stage of attribute definition in the "attribute table" and stored in BLOB type fields (binary linked objects). This structure is a good solution as regards the demands of the computer, but it weakens the stringent structure of the archaeological attributes. The problem is controlled by the software when attribrutes are sorted, searched and modified.

Up to now I spoke of attributes alone. All this applies to properties and classifications as well, classification and property being the equivalent to the attribute, class and property state being the equivalent to the attribute state. They are stored in the same tables. Classifications are always structured and on nominal scale as shall be explained below. So all information referring to classes is stored in the tables for attributes on nominal scale (fig. 7). Unstructured properties are stored together with the attributes without scale, structured properties are most likely to be stored in the same tables like attributes on nominal scale. Considering the origin of structured properties (fig. 3), cases for properties on higher scales should be rare. To distinguish classes, attributes and properties, there is another column in the "attribute tables" called "specification" (fig. 7). The user has to decide whether he is creating a new attribute, a class or a property.

The dialectical approach: modifying observations

As explained above, observations have to be modified from time to time in order to achieve significant data for archaeological work. Observations can be individual, very detailed and precise in the beginning, deduction or simplification of many precise observations naturally lead to more vague and global data on a lower statistical level, finally on nominal scale (fig. 8). Although statistical analysis on higher scale offers more possibilities for sophisticated analyses, the final and significant attributes on nominal scale are the bases for further archaeological research, for interpretation of artefacts in social, functional, historical or other contexts.

The process of modification needs to be done carefully in order to keep the data base consistent. Fig. 9 shows the steps from the decision to interrupt the process of data collection in order to evaluate the data and re-design the research down to the step of applying the new structure. Fig. 10 shows the next steps that are necessary to carefully modify the data already collected before new data can be collected with regard to the new structure or rules.

To speak in terms of a dialectical approach, the structure during data collection in the beginning is the "thesis". A new idea emerges during data collection, which lead to an "antithesis". This is the hypothesis that can be investigated with statistical or other methods. Therefore, a population in the statistical sense has to be defined (Sahner 1997, Orton 2000) - it can vary with regard to certain vessel parts, artefacts found in certain excavation units and so on. The data base creates a selection that represents the population. Analyses may utilize the whole population. But it may be appropiate to work with random samples that are easy to select in a data base with dialogues where the user can set the parameters. If analyses show that the hypothesis has a high probability, the user has to define in which way the old attributes have to be modified. This can be a simple grouping of many old attribute states to few new attribute states, but it may concern more complex rules. In the case study, the rules were recorded manually and modification was performed manually. It is another task for the future to implement recording and using these "modification rules" in the data base. Finally the new attribute is recorded in the "attribute table". It represents the "synthesis", to complete the comparison with dialectics.

The next step is to adapt the old records to the new research design (fig. 10). For this, all records containing the old attribute have to be selected. They are either deleted and replaced by new records or new records are recorded in addition, and the old records are kept for documentation. This is recommended, since most often the old records offer detailed information that is lost in the new, simplified or deduced attributes.

The flow chart in fig. 10 can be applied to any software or programming language. It shows how each old record is checked by the software, that an equivalent new record is created and that the old record is either deleted or marked as an old record. Deleted records are stored as text in another table and can be saved as text files from time to time. Marking a record as "old" has the effect that - in principle - it cannot be used for another analysis again.

This is an important feature in order to keep research consistent. For example an analysis which uses the attribute "fabric" should not use the attribute "porosity" if this is already part of the "fabric" definition. A contrary example may be the combination of vessel volume and vessel profile. Rim diameter may be used to calculate the volume, but it may be necessary for description of form or profile as well.

This problem concerns the "modification rules" mentioned before. It was controlled manually in the present study. To automatize the process, attributes are now being integrated in a familiy structure with "attribute genealogy". Attributes in the same family cannot be used to form a new level or "generation" of deduced or simplified attributes in the same family. As for the rim diameter in our example, this attribute would refer to two attribute families, and this would allow it to be used twice.

Examples from the case study

Two examples (Shinoto in prep.) may clarify the ideas. The first example concerns "fabric" and shows the emergence of a significant attribute out of non structured descriptions via detailed attributes up to extremely simplified, significant attributes. The second example shows the emergence of an attribute "surface treatment" from spontaneous, "virtual" piling of photographs based on intuition and leading to the definition of an attribute on nominal scale.

Fabric of the pottery in the case study is so irregular that it could not be classified in earlier studies. Detailed observations of features like colour or opacity of temper grains, grain size, porosity and so on (Orton, Tyers, Vince 1993) were recorded in many site reports (e.g. Kagoshima daigaku bunkazai... 1995) but never led to a better understanding of the fabric. In the present study, attributes concerning fabric were observed in an unstructured manner in the beginning. After a while, the observation records were checked for regularities or possibly significant features. These were now transformed to several attributes on varying statistical scales. For testing their validity, the artefacts with unstructured fabric observations were reexamined, the new attributes being recorded for each artefact. When it turned out that the modifications worked, new artefacts were recorded with this structure. The first result showed that sherd colour, feel, grain size distribution, form and quantity of pores and temper proportion seemed to be of importance. In the following sequences, the character and correlation of these features were investigated and simplification went on. Finally, three fabric groups with subgroups could be formed. These proved valid in terms of archaeology, since they show correlation with time and vessel form (Shinoto in prep.) and mineralogically, since several methods of analysis confirmed the groups and could explain their making (Shinoto in prep., Hoffbauer & Shinoto 2000, Shinoto & Hoffbauer 2000).

Fig. 11 shows some of the analyses in the course of the study. The simple table on the left groups artefacts with respect to attributes that seemed of importance in the beginning. The diagramme on the top right shows the relation of temper quantity and the distribution of grain size in the sherds. It was achieved with a quick and easy application of ideas based on fuzzy control rather than measuring and counting. The last table shows the correlation of colour - that was simplified from Munsell Soil Color Charts to three significant groups - with the temper groups that emerged out of the diagramme above. These examples and the other analyses were done with simple but efficient methods, the samples were random samples between 40 and 200 items. The sequential, self-consistent approach on small scale proved more successful than the traditional attempts of the last decades. An example of different nature describes the emergence of surface treatment groups.

The surface of the pottery shows a structure called "hakeme" which is made with various kinds of wooden spatulae (fig. 12). Japanese research on this technique achieved knowledge about how certain patterns were achieved (Sahara 1986). For several years, the width of each spatula and of the incisions as well as the number of incisions and the direction of the original spatula movement was recorded carefully in the excavation reports. Still, these efforts did not lead to any further conclusion and are abandoned from observation nowadays. During data collection for the case study, I got the impression that certain patterns correlate with time in the later period, but exactly like in Japanese research, recording data in detail did not lead to the formation of significant attributes. Recording of spontaneous, intuitive impressions seemed to be more appropriate to understand surface treatment than numbers and exact definitions. Instead of recording data on site, I later checked the records for photographs with details of surface treatment which seemed to be characteristic or interesting for some reason. These photographs were "piled up" in a "photo stacks table" which is an additional table in the artefact area of the data base.

Piling up photographs is a virtual act. By clicking on a button in the artefact table, a record in the "photo stacks table" is created with information about the artefact and picture. A keyword has to be chosen from a pop up menu that is created from a list. The keyword list represents the "stacks", it can be modified at any time. In order to create a more intuitive user interface, "piling up" photographs by drag and drop into several sections of a virtual desk represented by another input layout shall be implemented in the future. However, the less intuitive layout served well in the case study.

From time to time, the stacks can be checked. Are the records in one stack varied, should they be divided into more and smaller stacks, or should several stacks be unified to one larger? Grouping may not be convincing in any way, so records from one pile have to be moved to another. All this can be done spontaneously. Each action is recorded in a "log table", which makes the process transparent later. In the course of this process, a clearer idea about what makes the difference between the groups emerges. When the groups become convincing intuitively, the stacks or keywords can function as category names for attributes on nominal scale, and attribute records can be created for each artefact. Description of each attribute state will not be one single word or value, but a longer description and discussion. Although vague "understanding" of the emerging attribute may be intuitive, this understanding has to be expressed in a clear definition in order to become an attribute.

In the case of surface treatment, this procedure was extremely helpful, surface treatment forms groups that show correlation to time. "Hakeme" became an important factor in structuring the latest chronological phase of the pottery, which may cover up to 200 years and cannot be structured by vessel profile alone.

Summary and prospects

The system presented here is based on researches of the last 40 years which treat problems of data base design, classification in general and in archaeology and systematic approaches to pottery research. References on the cognitive aspect of classification, which becomes important in the "dialectical" approach derived from cladistics, are not mentioned (see e.g. Shinoto in prep.). The data base's basic structure was created before material collection in South Japan started in 1997, but development continued in the course of material collection and analysis. The first analyses were made with exported data and spreadsheet software, implementation of most common analyses and diagramme tools is going on.

The case study (Shinoto in prep.) proved that the consistency of a material collection can be maintained during research with continuously changing research design. More important than that, these changes proved to be most appropriate in understanding the material of the actual study. New insights were achieved that would not have been achievable by dividing research in two separate steps of data collection and data analysis. The flexible approach to research design offers a more vivid image of the ancient society.

The system is not limited to the case study but can be applied to different research problems in other regions and times.

Acknowledgements

I would like to thank the company AS Verlag & Consulting GmbH, where I could acquire the necessary computer knowledge and which provided me with hardware equipment. AGFA Gevaert AG provided me with slide films and advice in Japan, and the All Nippon Airways AG transported the heavy equipment to Japan and back. The Institute for Egyptian Studies at the University of Bonn gave the permission to test the system with their collection of prehistoric Egyptian pottery, my thanks in particular to Christina Regner. The study has greatly benefited from discussions with people from other disciplines, thanks to Martin Engel and Dr. Giovanni Galizia. Zoltán Böröcz, who commented on a draft of this paper. The views presented in this paper and any mistakes contained therein are the sole responsibility of myself. I gratefully acknowledge the community of Fukiage and the academic institutions in Kagoshima which supported the research most openly and generously.

References

Adams, D., Beckett, D., 1997. *Programming 4th Dimension. The Ultimate Guide.* Foresight Technology.

Adams, W.Y., Adams, E. W., 1991. *Archaeological Typology. A dialectical approach to artifact classification and sorting.*

Buck, C.E., Cavangh, W.G., Litton, C.D. 1996. *Bayesian Approach to Interpreting Archaeological Data.* Chichester et al., John Wiley & Sons.

Clarke, D. L., 1962. Matrix Analysis and Archaeology with particular reference to British Beaker Pottery. 371-382, *Proceedings of the Prehistoric Society 28.*

Clarke, D. L. 1978 (1968). *Analytical Archaeology.* 2nd edition revised by Bob Chapman 1978. London.

Dohrn-Ihmig, M. 1976. Die Anwendung statistischer Prüfverfahren bei der Stilanalyse und bei der Erkennung von Typen am Beispiel verzierter bandkeramischer Gefässe. 1-25, *Praehistorische Zeitschrift 51.*

Dunnell, R. C., 1971. *Systematics in Prehistory.* New York.

Gardin, J.-C. et al.,1976 (1985). *Code pour l'analyse des formes de poteries. Paris, Éditions du Centre National de la Recherche Scientifique.* 1985, dépôt légal initial 1976.

Hill, J. N., Evans, R.K., 1972. A model for classification and typology. 231-273, Clarke, D. L. (ed.) *Models in Archaeology*. London.

Hodson, F. R., 1966. Searching for the structure within multivariate archaeological data. 90-105, *World Archaeology 1*.

Hoffbauer, R., Shinoto, M., 2000. Joint archaeological and mineralogical research on pottery production in Ancient Japan. 989-992, Rammlmair D., et al. (eds.). *Applied Mineralogy in Research, Economy, Technology, Ecology and Culture. Proceedings of the sixth International Congress on Applied Minearalogy ICAM 2000, Göttingen*. Rotterdam, Brookfield.

Ikehata, Kôichi,1980. Narikawa-shiki doki no saibun hennen shian. (Proposal for a chronology of the Narikawa pottery.) 1-41, *Kagoshima Kôko 14*. Kagoshima.

Kagoshima daigaku maizôbunkazai chôsa shitsu (ed.), 1995. *Kagoshima daigaku maizôbunkazai chôsa shitsu nenpô 9-10*. (Annual reports of the Section for Investigation in Archaeological Properties at the University of Kagoshima 9-10). Kagoshima.

Kampffmeyer, U., Zamperoni, P., Teegen,W.-R. , Graca, L., 1988. *Untersuchungen zur rechnergestützten Klassifikation der Form von Keramik*. Helmut Ziegert (ed.): Arbeiten zur Urgeschichte des Menschen 11. Frankfurt a.M., Bern, New York, Paris

Karstens, K., 1994. *Allgemeine Systematik der einfachen Gefässformen. Münchener vorderasiatische Studien*. Münchener Universitätsschriften 16; Philosophische Fakultät 12. München, Wien.

Mayr, E., 1990. Die drei Schulen der Systematik. 263-276, *Verh. Dtsch. Zool. Ges. 83*.

Nakamura, Naoko, 1987. Narikawa-shiki doki saikô. (Re-thinking Narikawa pottery.) 57-76, *Kadai kôko 6*. Kagoshima.

Orton, C., 2000. *Sampling in Archaeology*. Cambridge manuals in archaeology.

Orton, C., Tyers, P., Vince, A. 1993. *Pottery in archaeology*. Cambridge manuals in archaeology.

Panchen, A. L., 1992. *Classification, evolution and the nature of biology*. Cambridge. Reprinted 1994, digital reprint 1999.

Sablatnig, R., Menard, C., 1996. Computer based Acquisition of Archaeological Finds. The First Step towards Automatic Classification. 429-446, Moscati, P., Mariotti, S. (eds.), *3rd Intl. Symposium on Computing and Archaeology, Rome, Vol. 1*.

Sahara, M., 1986. "Nendo kara yakiage made." (From paste to firing.) 27-41, Kanaseki & Sahara (eds.) *Yayoi doki 1*. Tôkyô, Yuzankaku.

Sahner, H., 1997. *Schliessende Statistik*. Statistik für Soziologen 2. Stuttgart, 4th ed.

Shennan, S., 1988 (1990). *Quantifying Archaeology*. Edinburgh University Press. (Reprinted with minor corrections 1990.)

Shepard, A. O. 1956 (1980). *Ceramics for the archaeologist*. Washington D. C., Carnegie Institution of Washington. (Manuscript submitted 1954, published 1956, reprint 1980.)

Shinoto, M., Hoffbauer, S., 2000. Kagoshima ken Fukiage chô Tsujidôbaru iseki no Narikawa shiki doki no kôbutsugakuteki kenkyû. (Mineralogical studies on the Production Process of the Narikawa Pottery from the Tsujidôbaru Site in Fukiage, Kagoshima prefecture." 91-104, *Jinruishi Kenkû – (Journal of the Society of Human History) 12*.

Shinoto, M., (in prep.) *Neue Forschungen zur Kofunzeit in Südjapan. Die Siedlungskammer um Tsujidôbaru und ihre Keramik*. Arbeitstitel Dissertation Heidelberg.

Stehli, P.; Zimmermann, A. 1980. Zur Analyse neolithischer Gefässformen. 147-177, *Archäo-Physika 7*.

Tatara T. 1981. Narikawa-shiki doki no kentô. (Investigations on the Narikawa pottery.) 89-116, *Kagoshima Kôko 15*. Kagoshima.

Traunecker, C. 1981. Code Analytique de Profils de Céramique de l'Ancienne Égypte. 49-77, Arnold D. (ed.) *Studien zur altägyptischen Keramik*. Mainz.

Uno, Takao. 1999. Kofunjidai chû/kôki ni okeru shokki, chôrihô no kakushin. Ritsuryôseiteki shokki yôshiki no kakuritsu katei. (Innovations in Food Vessels and Cooking Styles in the Middl and Late Kofun Period. The Process of Establishing Food Vessel Styles of Ritsuryô Society.) 25-42, *Nihon Kôkogaku 7*.

Vossen, R. 1970. Klassifikationsprobleme und Klassifikationssysteme in der Amerikanischen Archäologie. 29-79, *Acta Praehistorica et Archaeologica 1*. Berlin.

Whallon, R. 1972. A new approach to pottery typology. 13-33, *American Antiquity 37-1*.

Whallon, R., Brown, J. A. (eds.). 1982. *Essays on Archaeological Typology. Kampsville Seminars in Archeology 1*. Evanston/ Illinois.

	1996/97	2001	Request for future research
Computer	PowerBook 5300 (48 MB/2 GB)	PowerBook G3/400 (320MB/6GB)	notebook with any processor architecture
OS	MacOS 7.5	9.04	UNIX based OS
Photographs	Nikon FM	=	digital, 48 bit, more than 7 Mega-pixels; add-on for camera control from inside the data base software
	135 mm	=	=
	55 mm Macro	=	=
Paper scans	Ricoh FS2	Heidelberg 1450	-
	Cirrus 2.0	Linoscan 6.x	-
Slide scans	Nikon LS2000*	"	-
	LaCie SilverScan 4.x*	"	-
Digital imaging	Photoshop 3.0	Photoshop 4.0	-
CAD (maps/GIS, processing of artefact data)	MiniCad 5.0	VectorWorks 8.5	CAD-add-on or data exchange module for CAD packages independent of the software
DBMS	4th Dimension 3.5	version 6.56	newest version or SQL data base
Statistics	4th Dim, Spreadsheet software		add-on for advanced analyses in the data base or dynamic data exchange

*Table 1. Equipment used at the beginning of the study, at present and requests for future development. (= same as earlier; " same as above; - not necessary; * bought after material collection in 1998).*

Figures

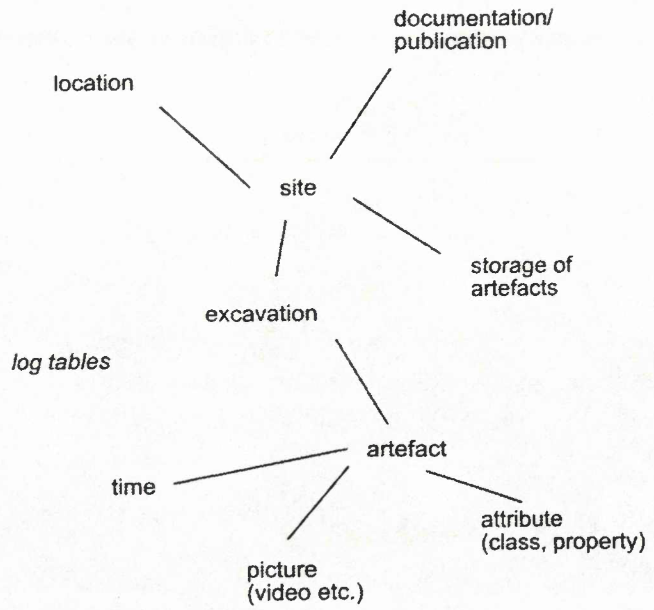


Figure 1. Overview of the data base structure. Each term represents an "area", i.e. a normalized structure of tables that store informations about the respective topic. Lines indicate a relation, where the central table of one area relates the area with the central table of another neighbouring area. At the moment, the data base consists of about 40 tables in 9 areas.

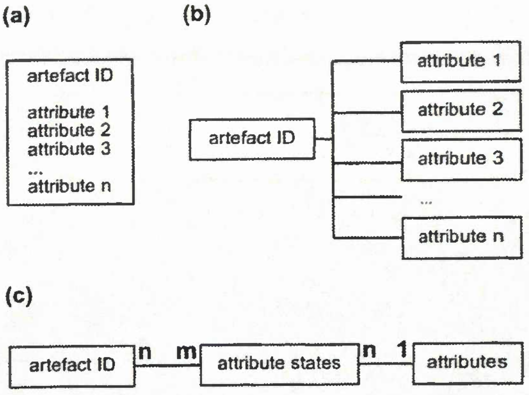


Figure 2. Examples for simple organization of artefact and attribute state relations, the starting point to develop the relation in Fig. 7. - (a) Flat file solution with individual fields for each attribute in order to record one attribute state if it is observed. (b) Each attribute is assigned to one specific table. If an artefact shows a certain attribute state, one record is created in the respective table. New attributes require an additional attribute table. (c) Relation where artefacts, attributes and attributes states are assigned to specific tables. For each artefact several attribute state records are created in the"attribute states table".

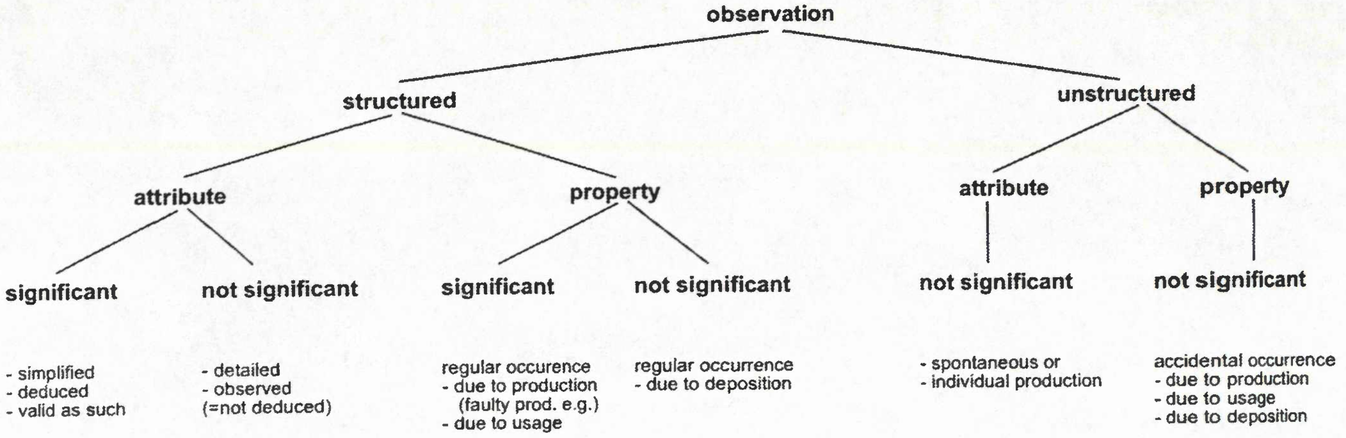


Figure 3. Classes of observations and their significance for systematic artefact classification.

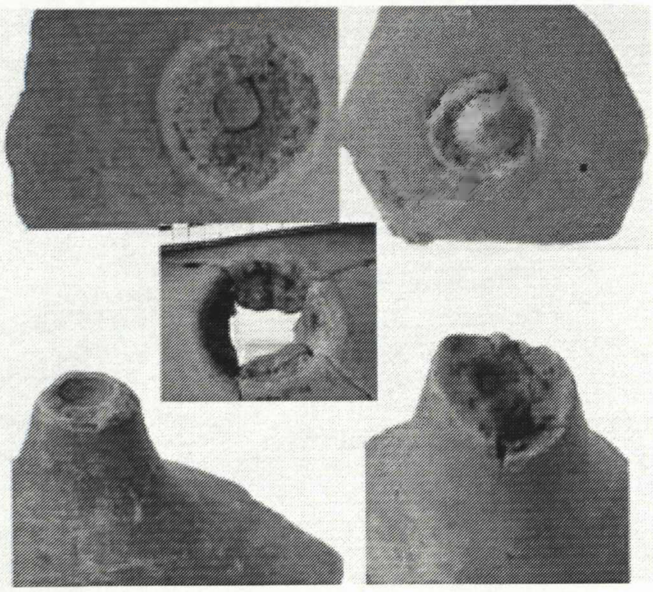


Figure 4. Attribute "adherence technique" deduced from several attributes at dufferent vessel parts.

archaeology	statistics	relational database (normalization)
attribute	variable	table "attribute"
attribute state	category / value	table "attribute state"
artefact	unit	table "artefact"

Figure 5. Structural parallels between archaeological attributes, statistical variables and a normalized relation.

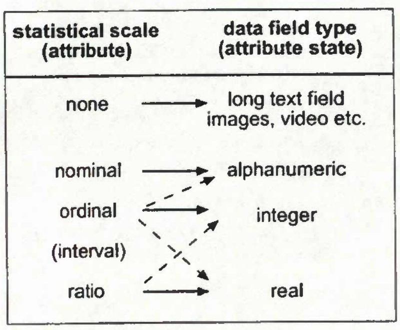


Figure 6. Assigning attributes on a certain statistical scale to data field types.

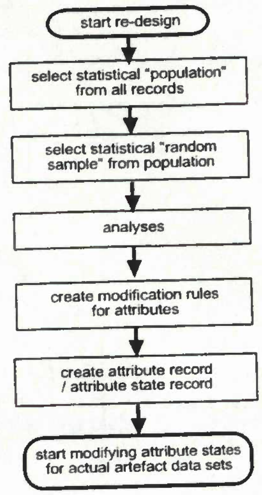


Figure 9. Sequence of steps that lead to modification of the research design based on new knowledge from new data.

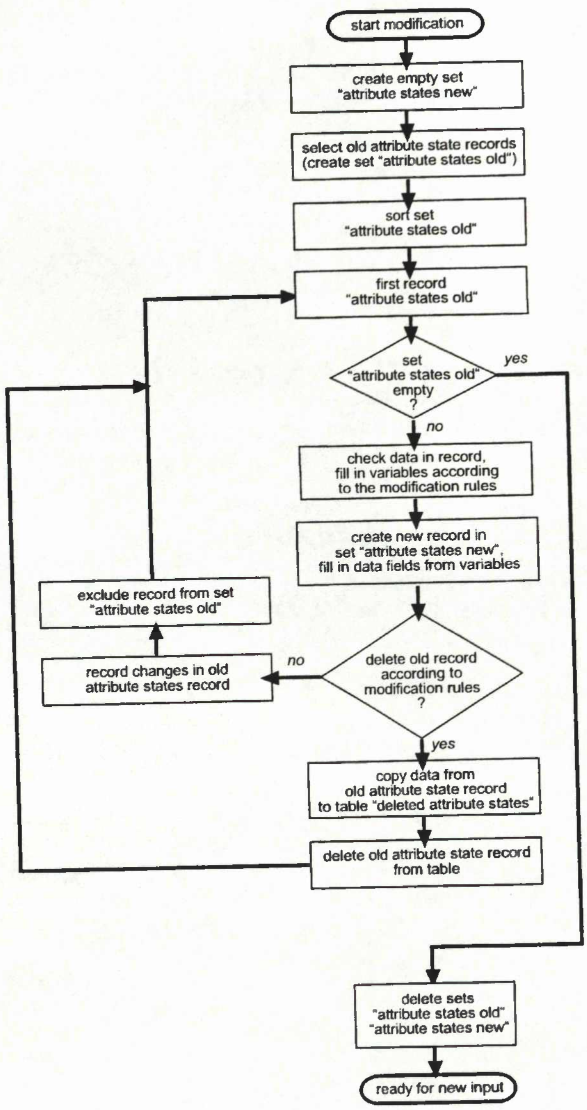


Figure 10. The process of modifying attribute states and property states of exsiting records after definition of the new attribute, property or class. Sets represent the selection. Deleting a selection of records means deleting the actual records whereas "sets" means deleting the representation without effect on the records.

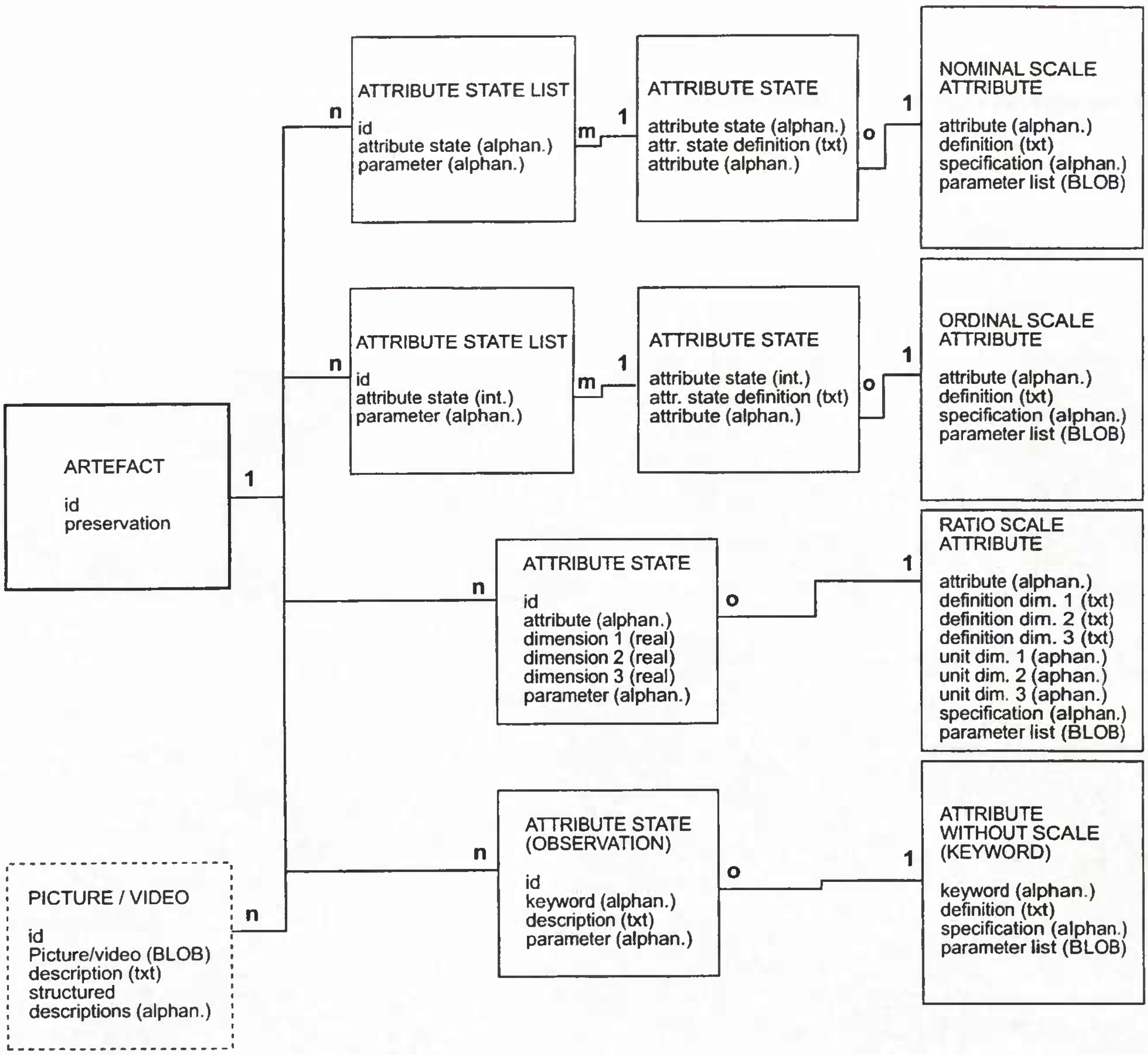


Figure 7. Tables to store observations, properties, attributes and classes (represented by the term "attribute") in the artefact area of the database. Pictures can be understood as nonverbal observations beside documentation, therefore they are included into this diagramme. – "1" and "n,m,o" represent the respective sides of a 1:n relation. The n-part of the relation is represented with different letters to show comparable levels.

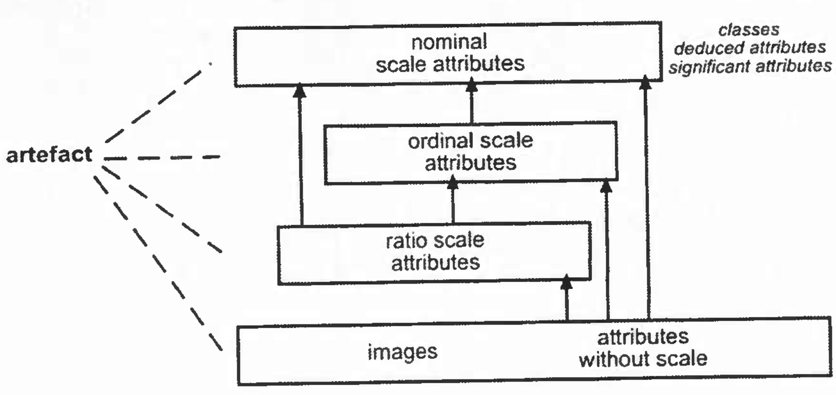
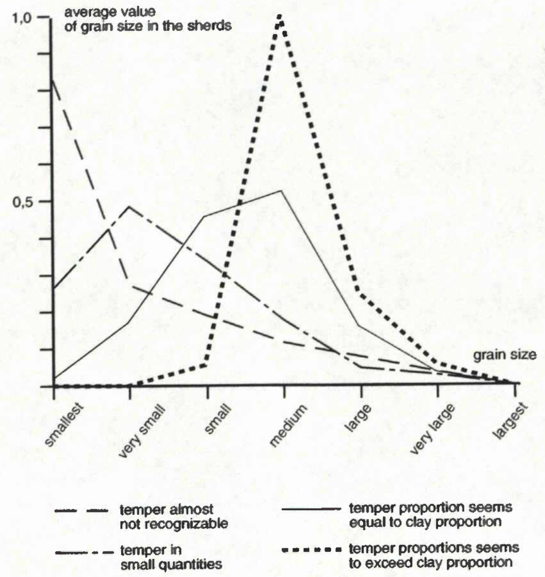


Figure 8. The sequence of scales in forming few, vague but significant attributes and classes from many precise, individual properties or attributes.

Probe	Farbe Braun	Sand	Menge km	w	n/m	e	ü	hauptsächliche Partikelgröße Kst	SK	K	M	G	SG	ÜG
112		o	o					o						
881		o	o					o						
825		o	o					o						
815		o	o					o						
229		o	o					o						
234		o	o						o					
236		o	o						o					
705		o	o						o					
843		o		o					o					
931		o		o					o					
826		o		o					o					
853	o			o					o					
1064	o			o					o					
821	o			o						o				
873	o			o							o			
332	o			o							o			
726	o			o							o			
735	o			o							o			
706		o		o							o			
709		o			o				o					
356	o				o					o				
736	o				o						o			
422	o				o						o			
592	o				o						o			
811	o				o						o			
195	o				o						o			
340	o				o						o			
331	o				o						o			
182	o				o						o			
949	o				o						o			
1098	o				o						o			
829		o			o						o			
1043		o			o						o			
832		o			o							o		
208		o				o					o			
186		o				o					o			
198		o				o					o			
371		o				o					o			
824		o				o					o			
836	o					o						o		
779		o				o						o		
385		o				o						o		
830		o				o						o		
810	o						o			o				
831		o					o				o			
787		o					o				o			
831		o					o				o			
797		o					o				o			
816		o					o						o	



	fine (fine)	(coarse)	medium	coarse	*sum*
light yellowish brown with red	8	4	4		*16*
"dirty" brown with red		7	11	2	*20*
pale yellowish brown without red				13	*13*

Figure 11. Some analyses on random samples that led to the three fabric groups. Left: table of observations that were structured from unstructured observations; n=50, 1 sample without data. Top right: correlation of grain size distribution and temper quantitiy leading to temper groups; n=200, 7 samples without data. Bottom right: correlation between temper groups and colour groups simplified from detailed observation with colour charts; n=50, 1 sample without data.

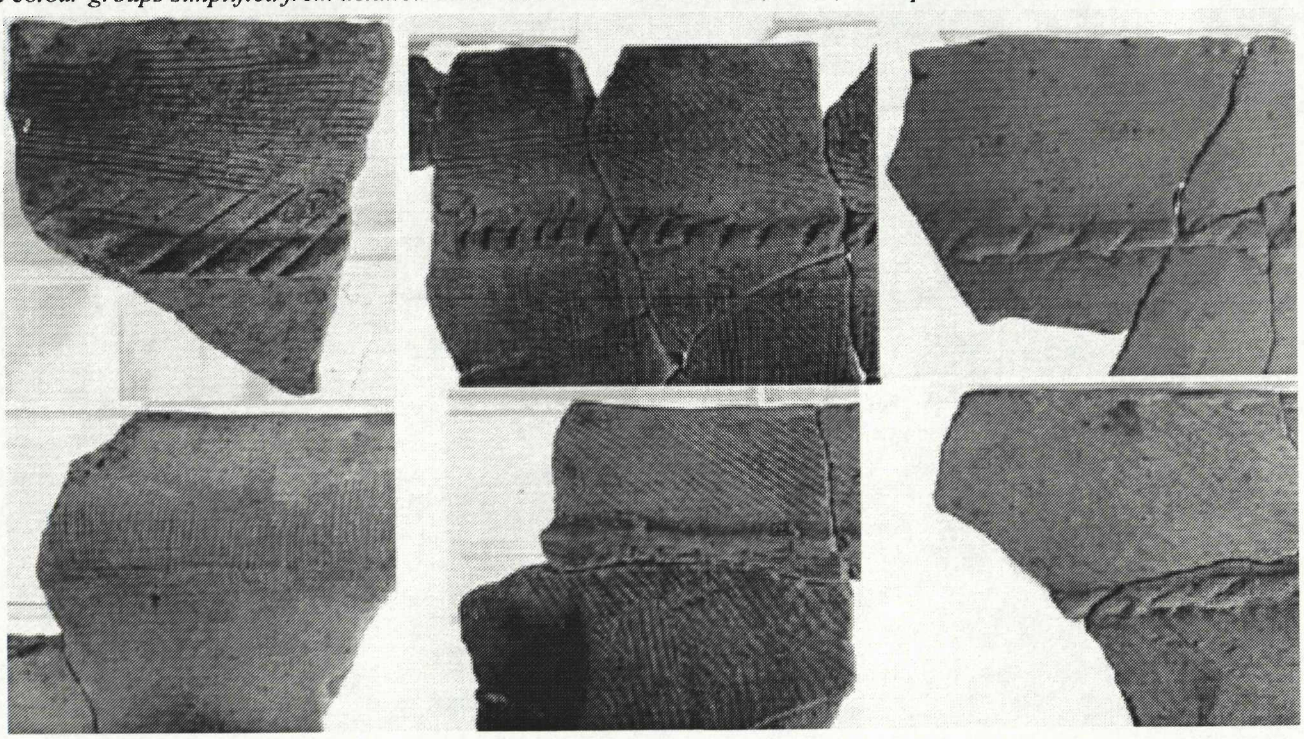


Figure 12. Some characteristic examples of surface treatment of pottery in the case study.