CREATING A CONSERVATION RECORD DATABASE USING A RELATIONAL DATABASE MANAGEMENT SYSTEM

Y.D.Stallard
W.A.Gray
N.J.Fiddian

Department of Computing Mathematics
University College Cardiff.

Abstract

A computerised system has been developed at University College, Cardiff to manage archaeological conservation records. This system is designed to be used by staff and students in the Archaeology Department who are carrying out conservation. An essential objective in the design of the system was that it could:-

1) deal with current conservation records - keeping track of objects which were currently undergoing treatment in the conservation laboratory and 2) allow records to be retrieved which referred to items which had been previously treated by the laboratory. This second objective meant that a backlog of previous conservation records which were held on cards in a filing cabinet had to be processed so that their data was captured by the new system.

This system was eventually created using existing software on the College's mainframe computer as far as possible, with an intelligent front end being built to create a suitable interface to these packages for the conservators. The most important packages in the system are MRDS (MULTICS Relational Data Store), - a relational data base management system - and its related

interface, LINUS plus LILA. There is also a custom built data capture program which was designed to overcome the problems of capturing the backlog of information already stored on conservation record cards.

This paper will present the objectives of the system and discuss how it was created. Reasons will be discussed as to why alternative methods of construction were not taken. The major alternatives considered were a system based on a modification of the "MDA's" GOS record system, and a system that made use of the data capture package "DATASTAR" which is available on a variety of micro-computers.

How the system was designed, and the approaches taken to help the conservators use it, will be discussed. These design philosophies will be linked to observations on how the system was used in practice and the difficulties which occurred. Suggestions will be made as to how the system could be improved.

1. Introduction

The Department of Archaeology at University College Cardiff has a conservation laboratory in which students and staff conserve artefacts. For each item conserved, the conservator creates a record on a standard record card. These records are handwritten and an example of a card is shown in Figure 1. In 1982 it was realised that the Department had about 3000 of these record cards and that it might be beneficial to create a computerised system which

1) allowed previous records to be accessed by a number of different key attributes,
2) created current conservation records - thus helping to keep track of items currently being treated in the conservation laboratory.

2. System Environment

These conservation records are prepared by the staff of the conservation laboratory - currently four members of staff - and by students reading for the Archaeological Conservation degree who work in this laboratory as part of their training. As it would be impossible for a variety of reasons to employ a person specifically to input information about current conservation tasks it would be the responsibility of these people to enter data to the computer system themselves. While it is reasonable to assume that with time the permanent staff members would gain computer skills, there was no reason to believe that the students would have any particular computing skills. Thus it was determined that the computer system should have an interface which hid computational complexity and made it easy for people who were computationally naive to interact with it.

At the time when the system was being designed none of the members of staff in archaeology had extensive computer programming experience. This meant that they would not be able to create their own system from scratch or support a system designed and programmed by someone else. As no money was available to pay for hardware or a programmer to develop such a system and support it afterwards, it was decided that as far as possible the system should be built using existing software and hardware supported by the College's computing centre.

It was realised that the capture of the information on the existing 3000 record cards would have to be carried out by special staff. The Manpower Services Commission were approached and agreed to create a post for someone to gain work experience by keying in this data. This again meant that the staff keying in this material were relatively unskilled.

3. Available Hardware

The college computing centre has a Honeywell Multics DPS 8/70M computer system with over a hundred terminals distributed throughout the College. Two of these terminals are available to the Archaeology Department. There are also a number of micro-computers in the computing centre which can be borrowed by Departments. It was decided initially that the system would be built on the mainframe with a micro-computer possibly being used for the bulk data capture of previous records.

4. Software Assessment

It was found that no interactive system already existed which specifically catered for "Conservation Records". A possible solution considered at this stage was to create a modified version of the MDA's GOS (ref 1) system. However funds were not available to do this, as it would have involved considerable programming effort. This meant that the only possible solution was to design a system to meet the Archaeology Department's specification using available software, thereby creating a customised system which is supported by the Computing Centre staff. After examining the software available on the Honeywell mainframe, the Multics Relational Data Store (MRDS) (ref 2) was chosen as the most appropriate for creating the system. It had suitable support software which would allow a customised interface to be built for the system in the macro facilities present in its LINUS interface (ref 3).

5. System Analysis

The data available on the record cards and the set of questions supplied as examples of the types of retrieval by the Archaeology Department were analysed. As a result of this analysis a normalised data model was created, which allowed all the suggested questions to be answered in a simple manner. This model, consisting of ten relations, was then created in the computer as a secure database with sub-models for the archaeology users. The relations of this data/model are shown in figure 2.

As designed there are two major components in the computer system - the data capture sub-system and the question answering sub-system. Ideally both of these components should be created using the facilities of the MRDS system. However a closer examination of the data capture processes revealed that there were two distinct types of data capture occurring. When capturing the backlog of information, complete records were available and would be entered as units whereas the information about new items being conserved would be entered over a much longer time scale and in sub-units of the complete record. It was clear that MRDS could meet the latter of these requirements, but in common with most data base management systems it was difficult to design an appropriate system for capturing a backlog of data that was suited to naive users.

6. Data Capture

If we examine the relations of the datamodel (figure 2), we see that certain attributes are repeated in separate relations. This repetition allows the relations to be linked on these key attributes during retrieval. However it complicates the data capture as it means that a data value may have to be replicated in several relations.

The following principals were adhered to during the design of the data capture sub-systems.

a) The values for all attributes were to be entered once only, and replicated if required by the system.

b) The structure of the database was to be invisible to the user.

c) During the data capture the information was to be entered in a natural way from the record cards, i.e scanned from left to right if practicable.

d) Any data which was repeated for all sub-objects on the same record card (eg. treatment) was to be input only once and then automatically copied to all tuples involved.

e) The system should convert certain types of data eg. dates into standard representation.

6.1 Data capture for items being conserved

Initially the two data capture facilities provided in MRDS/ LINUS were considered. The first of these is the LINUS "store" command which allows data to be entered into a relation. In response to a store command the user is prompted for each value that has to be entered. A sub-system for capturing information about items being conserved was built around this command. It is assumed that some data about the object being conserved will be entered immediately it arrives at the laboratory eg. labno, description. There will then be a period of inactivity with respect to this object as far as data capture is concerned while it is undergoing treatment and examination. Further information will then be entered. This pattern of long periods of inactivity between data capture will be typical of the data entry for items being conserved. This type of data capture is suited to an on-line system built on facilities provided by the store command.

A tutorial explaining how this data capture could be achieved was written for the students. The tutorial was available as a printed document or could be accessed at the terminal in machine readable form. It explained the concepts of relations and tuples and the meaning of those attributes and relations in their database.

By typing the name of a relation, the student obtains a screen display of the relation's attributes, along with rules for the entry of data under each column. These rules may take the form of a specified format for the data (labno/subdivision) or may refer the user to a thesaurus of "preferred terms", which currently takes the form of cardboard alphabetic lists, stored at the side of the terminal and standing on an axis so that they may be "flicked" through. They are presently being extended and it is hoped that when they are finished they will be stored in the computer.

The main problems discovered when data was entered in this way were:-

a) Spelling mistakes - especially of importance when key words were spelt incorrectly.

b) Uncontrolled use of terms - the thesaurus was not used to discover the "preferred term".

c) Deviation from required format - X-ray/photo nos. in particular were entered in many forms, leading to difficult retrieval of one no. from a list which could be stored as 187/01, 187/02, 187/03
or 187: 01-03
or 187: 01, 02, 03.

Future extensions for error-correction in this type of data-entry might include:-

a) Use of domain options in the data model source so that only words from given lists may be entered as data for a selected attribute.

b) The thesaurus could take the form of numbered lists so that only a number need be typed into the system and the relevant data automatically entered. This would eliminate spelling mistakes, but the user should be allowed to alter his selected number, as careless mistakes must be avoided at all costs.

c) Any word entered would be automatically looked up and its "preferred term" entered instead - great care must be taken here as accuracy is essential for correct "look-up".

But these can only be of use when the present lists are complete.

6.2 Data Capture of Backlog Records

In this data capture system it was essential that a large amount of information could be captured quickly in a natural way using relatively unskilled people (MSC trainees). It was realised that the system described above was not appropriate.

Initially a system was developed based on the LINUS file transfer command. To use this command, data for each relation must be prepared in separate files - one for each relation. Within a file the data items must appear as a delimiter separated list in the order in which the attributes appear in the corresponding relation. Each tuple is entered as a separate line in this file. This data is then imported into the database relations from these files using the LINUS store-from-file command. This part of the capture can be simplified by creating a macro which captures all ten files. However the initial preparation of the data in the files is clumsy and unnecessarily complex for the users if it is done using the system editor.

We looked at a number of alternative ways of creating the ten files. A very suitable approach was to use a specialised data capture package "Datastar" (ref 4) which could be run on a Superbrain micro-computer. In this approach a form was generated on the

micro-computer screen which was identical with the record card. The cursor was controlled so that it passed from field to field in a natural way and the user typed in the data from the record card at the appropriate point. Datastar creates a file of comma separated data from this entry process with each record creating a separate line in the file. Files created in this way were then transferred to the MULTICS mainframe using a locally supported file transfer program CTP (ref 5) developed by Henry Thomson of the computing centre. This data was processed on the mainframe by two specially written programs which transferred the data into ten files ready for import into the database by the LINUS command described above.

An alternative approach using the dBASEII database (ref 6) on a Superbrain was also considered. In this dBASEII captured the data into two relations for each record - one corresponding to the front of a record card and the other the back. A command file was then set up which performed ten applications of the copy command, each application creating a data file corresponding to one of the ten relations. These files were then transferred using the CTP program to the MULTICS system and imported into the database. This approach meant that there was no requirement to write special file conversion programs as in the "Datastar" method.

After consultation with the Archaeology Department and the Computing Centre it was decided that the Datastar based approach was best. Unfortunately at the time when the system was ready for operation, it proved impossible to get an appropriate micro-computer in the Archaeology Department. At this point it was decided that a system programmer (Jennifer Wallis) in the Computing Centre would write a special purpose data entry program which simulated many of the features of the Datastar system and ran on a MULTICS terminal. It was this system which was finally used to capture the backlog of data.

7. Question Answering

Around sixty queries were formulated by the Archaeology Department. These were regarded as the "most typical" type of questions to be asked of the stored data. It was realised that these questions could be parameterised and stored as macros in the LINUS interface to MRDS. In this way it would be easy for the users to ask a variety of similar questions of the database.

When a user wishes to interrogate the database, a list of numbered questions is displayed on the computer terminal. The user selects the type of query required by typing its number and enters the arguements (parameters) for the question. An example of such a question is shown in figure 3.

This macro question/answering system was set up so that all the computational complexity was hidden from the users. This simplified and minimised the set of commands that had to be explained to the users (e.g It is not necessary to explain how to open and close the database, set scopes and call the LINUS sub-system). This system also displayed the retrieved data in an appropriate default layout. However if a user wishes to by-pass this sub-system he can use the LINUS system himself.

As with the data capture sub-system a document was prepared describing how to use this system, and it was made available in a machine readable version so that users could access it from their terminals.

Some drawbacks of this system are that

1) The macros do not prompt the user for parameters. However, the documentation explains what these parameters are, and there are never more than five.

2) Any error creates difficulties for the user as he is left in LINUS with no indication and cannot ask another question without

first leaving LINUS.

When a user asks a question a natural language version of the query is displayed, so that they can check that it is the query they intended. This is displayed with the output from the query.

8. Outcome

a) Given the constraints under which the system had to be designed it is a reasonable working solution to the problem of managing the conservation records. A number of ways in which the system could be improved are identified in this paper. Although it was anticipated that it would be easier to maintain the system if it was built from standard software supported by the computing centre, in practice there have been some problems due to changes in the operating system which have affected the conservation record system in unpredictable ways.

b) Most people do not read off-line documentation until they hit a problem and they then prefer to talk to someone who knows the system rather than read the text.

c) Although the system is specifically designed to make it easy to answer sixty questions, it is possible to ask further questions using the LINUS interface. This would involve a user in learning more about the computer system.

9. Acknowledgements

Miss Stallard would like to acknowledge with thanks the financial support received from the United Kingdom SSRC in the form of a post-graduate studentship. We are also grateful to Miss K.Hunter, Dr.D.Leigh and Mr.P.Price in the Department of Archaeology for their help during the designing of this system.

REFERENCES

1. The "GOS" System — MDA, Duxford Airfield, Duxford, Cambridgeshire

2. Honeywell, (1982) — Multics Relational Data Store (MRDS) Reference Manual

3. Honeywell, (1982) — Multics Logical Inquiry And Update System (LINUS) Reference Manual.

4. Micropro (1980) — Datastar (Release 1.1) User's Guide

5. Thomson Henry (1981) — C.T.P (Cardiff Transfer Program) Computing Centre University College Cardiff.

6. Ashton-Tate (1983) — DbaseII (Version 2.4) Assembly Language Relational Database Management System. Reference Manuals.

7. Date C. J — An Introduction to Database Systems (Second Edition) - I.B.M (U.K) Lab. Ltd. Addison-Wesley Publishing Co.

8. Honeywell, (1982) — Multics Programmer's Manual Commands and Active Functions Reference Manual.

UNIVERSITY COLLEGE CARDIFF — Dept of ARCHAEOLOGY — D. O'CONNOR. — CONSERVATION RECORD

Material	Brief Description	Owner, Site.
Ae	4 coins (see list below)	Glamorgan/Gwent Site 21

Date Received	Suggested Completion Date.	Laboratory Number
24.7.79		LCC 2153
Work Approved By ------ Date ------	Removed By ------ Date	Owner's Particular Requirements.

Drawing Scale ..... Life

21.7501.01

21.7501.02

21.7501.06

21.7501.05

	DATE	Numbers, Results.
X-Ray	10.12.81	266/17, 266/18
Photos B & W	10.12.81	221, 222/1, 222/2
PHOTOS COLOUR	10.12.81	277, 278.
Analyses		
Expert Report		
Other		
Cost/Time		

UNIVERSITY COLLEGE CARDIFF — Dept of ARCHAEOLOGY — Record of EXAM & TREAT — LABORATORY NUMBER 2153

Date	Conservator	
9.8.79	S.O'C	ALL CLEANED, MECHANICALLY, AT FIRST, USING SCALPEL & GLASS BRISTLE BRUSH. STABILIZED BY IMMERSION IN A 3% SOLUTION OF BTA IN IMS AND FINALLY LACQUERED WITH A 50% SOLUTION OF INCRALAC IN TOLUENE USING SANTOAL AS THE MATTING AGENT. * 21.7501.05 ON CLEANING IT WAS DISCOVERED TO BE A VERY BASE SILVER COIN. FORMIC ACID WAS USED, APPLIED LOCALLY ON SARABS TO REMOVE THE COPPER CORROSION PRODUCTS FROM THE SOFTER SILVER SURFACE. * 21.7501.02 MECHANICALLY CLEANING REVEALED THE REMAINS OF SILVER PLATING. * 21.7501.06 CLEANING SHOWED THIS TO BE VERY BASE SILVER. THE COPPER CORROSION PRODUCTS WERE THEN REMOVED USING DIL (30%) FORMIC ACID; ALSO 15% AMMONIUM THIOSULPHATE AQUEOUS SOLN WAS APPLIED TO REMOVE THE SILVER CORROSION PRODUCTS. HOWEVER THE COIN REMAINED VERY DARK IN COLOUR & FURTHER CLEANING WAS ABANDONED SO AS TO PRESERVE THE DETAIL OF THE COIN.

Figure 1.

RELATIONS IN THE ARCHAEOLOGY DATABASE.

OBJECT.....	labno *	Contract	excavator	excavating -unit	owner	site	initial-date-received.

SUB-OBJECT.....	labno/ subdivno *	layerno	small-findno	museum-accno	primary-material	secondary-material	description	period	dimensions	latest-date-received

PREVIOUS-VISIT.	labno/ subdivno *	date-received *	previous-treatment	previous-conservator	institution	previous-date	ref

LOCATION.....	labno/ subdivno *	date-received *	location -prior	location -during	location -after

PROGRESS.....	labno/ subdivno *	date-received *	start-date	conservator	completion -date	removal -date	remover	collecting-institution

VISIT.....	labno/ subdivno *	date-received *	xray -nos	B&W-photo -nos	colour-photo-nos	hours-spent	cost

TECHNICAL-EXAM.	labno/ subdivno *	date-received *	method *	part *	investigator	exam-ref	result	technical-exam-date

OBJECT-CONDITION	labno/ subdivno *	date-received *	part *	material	completeness	condition

CONSERVATION...	labno/ subdivno *	date-received *	part-&-material *	treatments *	materials -used	date-of-treatment

DOCUMENTATION..	labno/ subdivno *	document-description

* indicates key attribute

Figure 2.

List all objects ever treated in the lab.

Where

a) DESCRIPTION = A (Rel Sub-Object)

b) PRIMARY MATERIAL = B

c) PERIOD = C

Printing:

All Object data

All Sub-Object data

X-Ray numbers

Date-received in lab.

TO USE type:

EXEC qll A B C "SORT"

(Where "SORT" is replaced by any column name occuring among those to be retrieved).

Figure 3.