Enabling school pupils to interrogate census data

Watson, Deryn M.

ENABLING SCHOOL PUPILS TO INTERROGATE CENSUS DATA

Deryn M. Watson(+)

Abstract: This article reports the work that has been undertaken to develop a data retrieval and interrogation package for use on microcomputers in schools. The aim of this work is to facilitate an understanding by the pupils themselves of the nature of historical investigation by giving them an opportunity to collate, store, search, and interrogate the material. By utilising the enormous power and potential of micros in the classroom it is also hoped that the pupils will be provided with a practical path for an understanding of the workings and nature of information storage and retrieval in a computer age.

INTRODUCTION

Historians and social scientists are increasingly turning towards the utilisation of computers to facilitate the handling of large quantities of historical data. Such an interest however has on the whole been confined to those in the tertiary world of university/college teaching and research, or large archive bodies with access to mainframe computer facilities. This paper is reporting on work that has been done to provide material to enable school pupils, from the age of 11 upwards to actively use a microcomputer themselves to handle data as part of their history lessons. And whereas in many cases the use of a large mainframe for data handling has been closely associated with the application of statistical routines to extract trends from such data, this work on the micro does not refer to statistical routines at all, but rather encourages the pupil in interrogate the data by making secondary searches using conditional entries on a relational data base. The data for which the software is designed is the 1851 Census of England and Wales - the first complete census that was taken for every household. The total unit can be used with small local subsets of this data, and indeed for subsequent years (up to 1901 - due to the operation of the 75 year rule). It is intended to encourage the use of data in relation to a local region and in combination with other local history studies. Thus the purpose is not to extract information on countrywide trends in the population, but to provide a window through which the pupil may look at life in his home town some 130 years ago. This in turn may stimulate questions relating to whether similar households can be found in other areas, so that comparative studies may be generated from the specific.

BACKGROUND

This work has been undertaken by the Computers in the Curriculum Project, funded by the national Government (School Council and Microelectronics Education Programme (MEP)) over a period of 10 years and based at Chelsea College, University of London.(1) This project is funded for the research and development of computer assisted learning (CAL) materials for secondary schools(11-18) in the UK. As a national project, it must consider the needs

(+) Address all communications to: Deryn M. Watson, Computers in the Curriculum Project, Educational Computing Section, Chelsea College, 552, Kings Road, London SW10 OUA, GB

of teachers in many different parts of the country who may be preparing pupils for a diversity of curricula and examinations. There is also an imperative to not only investigate and report possibilities, but also to put our ideas into practice by developing educational software that can be used in schools. To this end, our material is prepared for publication by an educational publisher in the hope that it will increase the quantity of available software and act as an exemplar to stimulate further developments. Although the project works with a variety of disciplines, from physics and chemistry to history, geography and English, its starting point is a concern for education and the application of software to assist the learning process. It is important to stress therefore that the project, although working with computers, is not concerned with an investigation of systems and their relation to applications processes, but rather with the development of appropriate software for utilising the power of micros within the educational framework.

The starting point for all our work has therefore been the school base. In the mid 1970s most historians, and particularly history teachers in schools did not feel that their discipline was numerate and thus appropriate for an investigation of the application of computers. Nevertheless the Project philosophy and model of working made it imperative that we involve history teachers in the work.

THE PROJECT PHILOSOPHY

Although the main rationale of the CIC Project has been described elsewhere it is important to identify certain key concerns as they impinged upon the development of the history data base material.

"1 All investigations are implemented within an educational framework, but not necessarily always in the control of a teacher. We have always firmly believed in the interactive potential and positively encouraged designers to consider the pupil as an active learner when sitting at the keyboard rather than one being passively directed either by the teacher or the software design. Thus the project has aligned itself with the heuristic principles of discovery learning and encouraged the user to ask 'what would happen if ...?' or, more recently, in relation to decision making exercises in the humanities, 'what shall I do?'

2 Accordingly no attempt has been made to look at systems of computer managed learning, nor indeed 'drill and practise' routines (or 'structured reinforcement', as it is often termed). This is not to make any comment on the value or desirability of these approaches, but simply to confirm that the Project considered them outside the remit of their main concerns.

3 In order to reinforce the aims and directions of the software that CIC developed, it has always been considered only one part of an overall educational package. There was a clear imperative to provide related printed materials for both the teacher and pupil which not only placed the software in a context but provided the clear curriculum environment in which it could be used. The development of its 'courseware' takes as much times and is of equal concern to CIC as the software.

4 All the development must take place with the context of the curriculum base of the school, which in the UK still relies on a discipline framework. There seemed to be little mileage in producing software for an esoteric part of the curriculum that would be of little attraction to

the average teacher. Nor was there any attempt to develop materials that would replace perfectly good teaching/learning methods already in use. As a main tenet therefor CIC always asked 'how can the computer aid the learning of a topic in a way that has not been possible before?'"(2)

The model chosen for the response of such questions was the generation of a team of developers who could consist of teachers, programmers, and CAL designers. The work described in this paper is the result of the deliberations of such a group over a period of some 10 years.

THE DEVELOPMENT

Although the Project had initially worked principally in the area of simulations, it was with interest that it received from the initial deliberations of the historians a request to develop a large data interrogation package.

They wished in particular to have the facility to interrogate the 1851 Census with all its myriad of detail. Why was this census of such particular concern? For the first time it attempted to cover every dwelling in England and Wales. Enumerators were employed to make returns that included all households in any one dwelling, the names of all inhabitants and their relationship to the head of the family, sex, age, marital condition, rank profession or occupation and place of birth. The last two sets of information in particular provide the basis for a detailed understanding of the social history of the period.

In the UK in the last 10 years there has been a growing move toward both the study of local history material by pupils as a method of appreciating and learning the history of their forefathers and also toward creating an environment whereby pupils might "experience history for themselves".(3) This has led to an emphasis on the nature of evidential sources and understanding, on hypothesis testing and historical methodology. Indeed a national Curriculum Project in the late 1970s (the Schools Council History 13-16 Project) highlighted these concerns and was partly instrumental in the movement away from a purely factual understanding of history into one of historical enquiry.(4) Pupils are increasingly encouraged to explore historical sources for themselves and draw their own conclusions rather than accept the pre-conceived statements of text books. Such enquiry-based learning of the historians matched the concerns of the CIC Project to develop and foster the use of the micro in the school as a tool for discovery based learning.(5) It is in this interactive mode that we believe its greatest potential lies. The development began from such a matching of concerns. It rapidly became clear that the task was enormous and covered certain stages.

"1 For the purposes of interrogation, the raw data of the enumerators returns needed to be encoded into an acceptable format, without destroying the links with the real history that it contained. Thus if the enumerator used an unusual spelling of a christian name or occupation, the teacher would want that included and be able to recall it in the enquiry. An example of a typical enumerators entry is given in Figure 1.

2 Such an encoding pattern should realistically be the reflection of a variety of historians who should be encouraged to come to some form of agreement about a format which could be widely used. Encoding data is so time consuming that it is important to look at some uniformity in order to gain maximum benefit from the data files with a variety of interrogation packages.

Figure 1 A transcript of the actual census return

	Name of street, place or road	Name and surname of each person in house	Relation to Head of family	Condition	Age M	Age F	Rank, Profession or Occupation	Where born
47	Wordsley	Jane Webb	Daughter			7	Scholar	Stafford, Kingswinford
48		William Leek	Head	M	56		Gardener	Stafford, Wolverhampton
		Elizabeth Leek	Wife	M		51	Laundress	Stafford, Wolverhampton
		John Leek	Son		16		Scholar	Stafford, Kingswinford
		William Burrows	Father-in-law	M	77		Formerly Locksmith	Stafford, Wolverhampton
49		Hannah Oldfield	Head	Widow		65	Kept by son	Salop, Quatford
		William Oldfield	Son	Un	35		Iron Moulder	Stafford, Kingswinford
50		Richard Truman	Head	M	34		Brass founder	Worcester, Dudley
		Elizabeth Truman	Wife	M		34		Stafford, Kingswinford
		James Truman	Son		16		Brass founder	Stafford, Kingswinford
		Clara Ann Truman	Daughter			14	Scholar	Stafford, Kingswinford
		Elizabeth Truman	Daughter			3		Stafford, Kingswinford
		Thomas Caishorn	Father-in-law	M	69		Boatman	Stafford, Kinfare
		Martha Caishorn	Mother-in-law	M		69		Stafford, Kingswinford
51		Benjamin Poyner	Head	M	66		Glassmaker	Stafford, Kingswinford
		Sarah Poyner	Wife	M		66		Stafford, Kingswinford
		Joseph Poyner	Son	Un	25		Glassmaker	Stafford, Kingswinford
		Sarah Poyner	Daughter	Un		22	House servant	Stafford, Kingswinford
		Joseph Walters	Son-in-law	M	35		Coalminer	Stafford, West Bromwich
		Hannah Walters	Daughter	M		36		Stafford, Kingswinford

3 Despite (in the late 1970s) the prevailing problems of batch processing or time sharing systems with a terminal based in the computer department, the historians felt simply that the students should be encouraged to set up their own enquiries of the data and make their own investigations through secondary searches. The interactive potential needed to be fostered in order to gain the maximum historical advantage.

4 Despite the size of the operation a large pack of courseware should be written to

a) hold the hand of the history teacher and pupils who may be unfamiliar with CAL but for whom the package would be an asset
b) explain in detail the nature and wealth of possibilities that such a piece of software can offer.

The notes would need to show not only the mechanisms for setting up a data file, making both the initial and subsequent enquiries, and deciding on the format of the results: they would also need to show the various historical points that can be extracted upon the examiniation of a relational data base, within the context of the history curriculum."(2)

During the development period micros became available; indeed due to Government initiatives in the UK, 96 % of all secondary schools has at least one micro;(6) many have more and they seem to breed fast. Thus it was decided in order to fulfil the requirements of the working group, the software must be developed on disc driven micros, with their relatively limited memory.

The work developd in stages; inevitably an encoding format needed to be agreed before the interrogation software would be designed to accomodate it. At the same time notes needed to be drafted to explain the operation to both the teachers and students. The whole unit was then sent to trial in selected schools and the package subsequently modified in the light of feedback.(2)

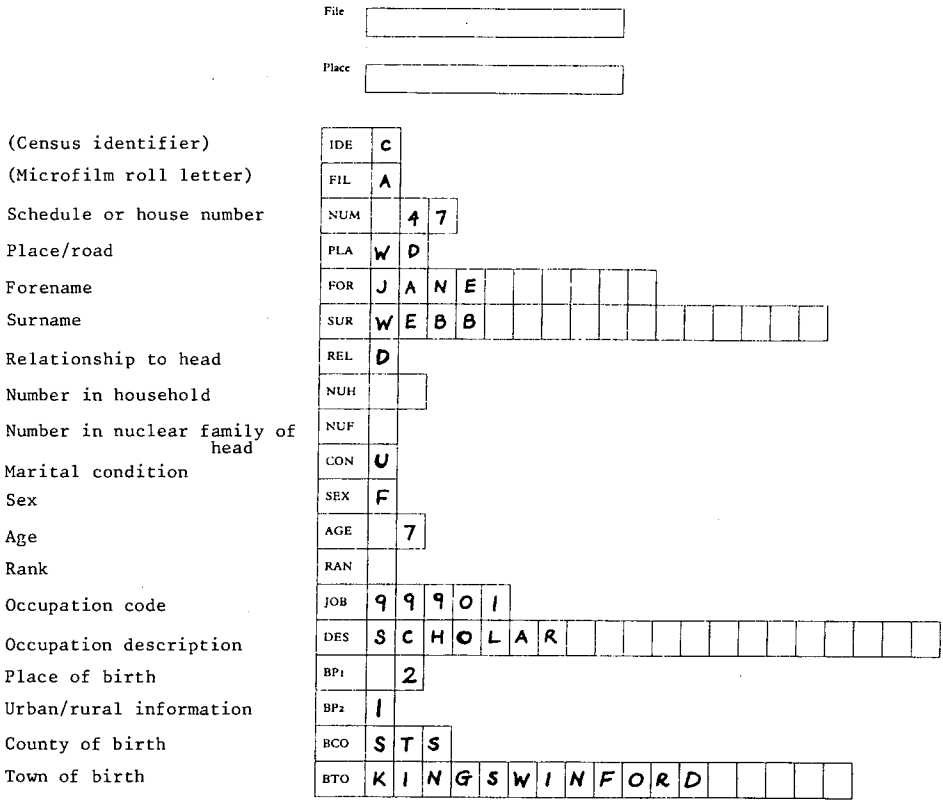
A DESCRIPTION OF THE UNIT

The final unit "Census Analysis" is now published and consists of five component parts.(7)

The Encoding Book

This gives complete details of the encoding framework for the census data, together with students' leaflets on how to put the encoding pattern into operation. The pupils are encouraged to read the original enumerators returns and become familiar with it, including such details as handwriting, how mistakes are noted etc. The aim is to transfer the exact enumerators record across as much as possible and so mis-spellings are to be retained. The enumerators returns are transformed into a line of data made up of 19 fields containing a possible 92 characters in total. Figure 1 gives a transcript of an actual census return; Figure 2 gives entry number 47 encoded. Each field has a 3 letter basic reference identifier; the explanation is given in brackets on the right hand side.

Figure 2 Sample of one entry encoded



This is entry No. 47, from Section 3, encoded.

The history teachers were concerned that the encoding pattern should reflect as much as possible that of the enumerators returns - thus the fields appear in a similar order. One of the key issues was to be able to encode the job description as written by the enumerators in 1851, and yet set up enquiries concerning particular occupations. To this end, a job code was necessary. The Project used the one already available (the Booth Armstrong code)(8) and attempted to provide as much encouragement for users to encode any jobs whose description was not yet categorised e.g. Bonnet Maker, rather than keep them in general categories such as Agricultural - Other. To this end there is plenty of space in the numerical codes for extra identifiers. At the same time it was important to keep space for the actual descripter rather than rely on the code for the record, and so this field (DES) has 20 character spaces available.

Rank covers data given by the enumerator such as Apprentice, Gentry, Owner, Public Servant, and even Pauper. Relation to head to household includes persons such as nephews, residential pupils, servants or lodgers. Perhaps the other most interesting entries suggested by the encoding form however are the last 4 fields, which give information about the place of birth. The code suggested for birthplace (BP1) uses a numerical system arranged so that higher numbers represent more distant places from the census town. Hence the codes are different for each locality and determined by the pupil user in their own location, using a map. Thus 01 represents the fact that the individual was born in the Census town itself. 2-19 cover local towns, 70-79 major cities, 90-99 foreign countries. In this way the users build up their own pattern of information and directory of codes which could act as a basis on local migration studies.

Ideally the publication of such an encoding format will enable pupils and teachers from a variety of schools in various parts of the country to encode their own data and create a file, and arrange for the exchange of their various files for comparative studies. The interrogation software will enable a school in a market town in the South of England to also look at the data of a 19th Century port, railway terminus etc. And such comparative studies will be undertaken from a basis of the particulars rather than the general and may form the basis of defining and testing hypotheses.

Input

Once the data is encoded, it needs to be put onto the magnetic media (a floppy disc) to act as the data file for interrogation. It became clear during the trials of the material that users would appreciate some facility software to speed up the task and attempt to eliminate errors. Thus a program called INPUT was written, which re-creates the pattern of the 19 fields (see Figure 2) on the screen and enables the user to enter the data in the appropriate boxes by use of a moving cursor. Once the entry is complete and the user is satisfied it is correct, it is entered onto the disc. This way the user can enter small batches of material at a time until the total data entry (e.g. for a parish) is complete. The data on the disc is now in the file of material and is ready for use. The published pack of material includes a disc with a data file on it, and it is anticipated that as more become involved the publishers may be able to offer discs of data from a variety of areas.

The program CENSUS

This is the core to the whole unit. The program has been written to enable the interrogation of the data file. It operates through the search of va-

rious conditions that are identified by the user. The material is a relational data base and operates accordingly. The five conditions are equals, not equals, greater than, lesser than and between. A pupil may want to find out how many girls were at school in the area in 1851. The search would be set up by asking for SEX = F, AGE< 16, JOB = 99901 (the job code for a scholar). The data would be searched and a secondary file created of all entries for girls under the age of 16 who were scholars. The pupil may then want to know what sort of household such a girl came from, and has the opportunity to find the rank or profession of the fathers of these girls. This in turn may highlight the restricted type of households of these girls (Gentry, Doctor, Army etc.) and at the same time prompt the pupil to notice what a small number there were. Another search may now be made of the data to list all the girls under the age of 16 and note their occupation. The software places the interrogation tools in the hands of the pupil; the facility is there to make se-condary 'hit files' and follow through the implications of the first results by making further searches, either for more detail of the subsets or making wider 'sweeps'.

In order to encourage use by teachers and pupils, much attention had to be paid to the screen design of the software. Users are addressed in English, rather than "computerspeak" and invited to respond to questions such as "specify the characteristics of the people you want picked out from the CENSUS" and "Do you want information about their households as well?" In the trials version of the software the team was so concerned to "hold the hand of the user" that we perpetually put up messages asking if they were sure this was a correct entry. Trials feedback reported irritation with these messages because they slowed up the interaction and pace of the enquiry. It was also pointed out that if we are attempting to produce pupils who appreciate the need for accuracy in each field when encoding they must also appreciate accuracy when interrogating. Editing procedures, while important, must not obtrude so much as to suggest to the pupil that they need not have considered their enquiry carefully before inputting at the keyboard, or interrupt the flow and sequencing of any search.

The software itself is designed around an optional structure. The 4 main component parts are - to set up an enquiry, choose the format to display the results, search the data to create the hit file, and output the results. Each is a self-contained component of the program, and the first three can be returned to in any sequence for editing or "changed minds" before the results of the search are implemented.

The actual speed of the enquiry is significant, and trials responses suggest that the operation was slow. Much effort has gone into modifying the software to improve this aspect while working within the constraints of the main micros found in secondary schools, i.e. RML 380Z, Apple II and BBC Model B.

The actual layout on the screen also gives some physical identification to the actual process that is involved. Indeed the whole unit has aimed to highlight rather than hide the actual nature of information capture, encoding and retrieval, and this issue will be dealt with more thoroughly in the next Section.

Teachers Guide and Students Leaflets

Notes for the teacher accompany the CENSUS program, explaining its rationale and how it may be used. A large document, it is important as it explains how the unit was developed, places it in the context of current historical concerns and classrom practise. Thus it aims to cover both the pedagogical

and methodological issues that relate to computer interrogations in the average classroom. It has been written deliberately to enable the teacher, who may be a total novice to the concept of using the computer in the classroom, to feel comfortable and confident that this may be a useful teaching and learning resource. This Guide was written by the principal authors who formed the core of the history working team. Thus it is written not by a curriculum development project, nor by programmers, but by individuals whom the teacher can respect i.e. other history teachers who face the same issues that they face.

A separate booklet contains the students' leaflets. These take them step by step through the software's potential use, including examples of the type of enquiry they may wish to make. They cover the background principles to the sorting of data, the actual procedures that they need to follow for the operation (with many photos on the screen images at each stage) and various suggested routes of operation. The Leaflets do not have to be used by either teacher or pupil, but are a method whereby we can indicate the various ways the unit can be used within the confines of a history classroom, or enquiry. The language of the students leaflets is geared towards a pupil of about 13 years and attempts have been made to clarify the sequence and purpose of each running stage. Examples from questions on the leaftlet introducing the Census includes "What sort of things do you think opposers to the census were afraid of". "Who might find a census useful apart from the Government?" "Which was the first census to include details about you?" On other leaflets questions may be found such as "Plan how to use the computer to study up to what age sons commonly lived with their parents" and "Make a hit file of all the people who work on farms. Work out a useful way to tabulate the information".(7)

Thus the whole unit consists of both written (paper based) material and the magnetic material.

IMPLICATIONS FOR HISTORIANS

My purpose in giving such a detailed description of the unit is to highlight the reality of translating the CIC concerns in general into the particular as it affects the historian. The move to take as much historical methodology as is practical into the average school must surely be welcomed by those in the tertiary world. Through the use of the unit within an educational environment pupils are given an opportunity to become familiar with the real evidence as well as the questions that the evidence raises. A brief scrutiny of such records by the pupils without the benefit of the micro usually results in them picking out the unusual or odd, in other words those facts which catch the eye. Using the power of the micro enables the individual to submit the data to a truely historical analysis. Indeed the unit was trialled in one primary school with 10 year old pupils with significant success.

However the implications of developing software and notes that are specific to a particular set of data are severe. The CIC Project worked on this problem since 1975. Since 1979 there have been the equivalent of 4 man years of resource as well as the freely given time of many serving teachers and access to the skills of the project team as a whole. This implies a significant level of funding. The resource cannot be condensed into a shorter time period; much of the refinement has been as a response to periods of reflection or trials. Nevertheless it does point a way for the potential to handle other records e.g. parish records, trade directories or probate inventories. The value of such work will hinge upon the premium which historians may place upon the availability of such a facility. Other questions, such as the

use of general purpose, rather than customer built specific purpose, data interrogation packages will then need to be considered.

WIDER ISSUES

The potential of such material however does not lie just with its historical significance. School pupils are entering an age of Information Technology. The UK Government has been responsible for two initiatives in the early 1980s designed to bring the regular use of computers to the awareness of every pupil. They aim for every child to leave school and enter society and (hopefully) the world of work with an understanding of the use and potential of these machines. Rather than simply teach computer studies as a "new" discipline, it is clearly desirable for pupils to see it in operation in a variety of scenarios, such as the school office and the average classroom. Use of the computer in the geography or history classroom thus reinforces this point. But use of a data interrogation and retrieval package may point pupils in the direction of understanding the mechanism of data handling and sorting (and thus overcoming the mystique of the machine) and also the question of what data is appropriate for such handling. In a field of such rapid change, the access to larger and more remote data bases through the medium of telesoftware will shortly become relatively commonplace. Tomorrows citizens should feel confident of the mechanism and value of being able to "download" subsets of data from a variety of sources geographically spread across the world. With the parallel methodological lessons they have learnt in their history classes they should also be able to bring their critical faculties to bear upon the nature of the data, its reliability and its applicability in each particular case. The unit Census Analysis has been designed to facilitate an understanding of both the process and use of information technology.

FOOTNOTES

1) R. Lewis, D. Want - "Educational Computing at Chelsea 1969-79". Computer Assisted Learning - Scope, Progress and Limits. eds. Lewis, Tagg. Heinemann, London 1981.
2) D.M. Watson - "A Census Data Retrieval and Interrogation package developed for use by pupils in schools". Proceedings of International Conference on Data Bases in the Humanities and Social Sciences. Rutgers, 1983, Lexington Press, USA 1984 (in press).
3) P.J. Rogers. "The history - theory into practise". The Historical Association. T.H. 44, London 1980.
4) Project Director - Dennis Schemildt. Trinity and All Saints College, Horsforth, Leeds, UK.
5) D.M. Watson. "Humanities Teachers Write Imaginative CAL". Proceedings of 6th International Conference on CAL in the Humanities. eds. Burton and Short, Computer Science Press, Maryland, 1983.
6) Department of Industry - "Micros in Schools", DoI Information Leaflet 1981.
7) I. Killbery, R. Mason, D. Watson. "Census Analysis", Longman, Harlow, 1983. (Software available for Apple II, RML 380Z and BBC Model B (all disc driven)).
8) Wrigley E.A., "Nineteenth Century Society", C.U.P. 1972.