RESEARCH IN PROGRESS

THE CHESHIRE PARISH REGISTER TRANSCRIPTION
PROJECT 1978-1989

P.E.H. Hair and C.B. Phillips

Unlike neighbouring Lancashire and Staffordshire, Cheshire has never had a parish register publishing society, hence only a very few registers, or segments of registers, have appeared in print. Furthermore, local church opposition has limited the amount of microfilming of registers by the Mormons. However, in the early 1970s the county and diocesan record office began its own programme of microfilming registers. These various factors led in 1978 to the county archivist, Brian Redwood, together with the authors of this note, founding the Cheshire Register Transcription and Computerization Project, a project active since.

The aim of the project is to produce an edition of the parish registers which will not only reproduce all the data in the entries in the Parish Registers (PRs) but will incorporate variant information from the Bishop's Transcripts (BTs). The edition will be stored in the computer and will be available as output from it. The data can be processed to produce, for instance, indexes of names, but it can also be 'interrogated', that is, certain data elements can be sorted out and assessed, for instance, quantitatively. In view of the general scholarly acceptance of the need for a national parish register index - one comparable in range to the post-1837 St Catherine's House index of secular registration - the present note records our progress so far. It indicates some of the very many problems we have faced and have overcome - or are, we believe, overcoming - in the expectation that this report will be of some assistance to later similar projects.

The two operating principles of our project are of equal importance - the adoption of up-to-date computer-based processing methods, and the large-scale involvement of volunteer labour. To date, some 300 volunteers have worked, for varying periods of time, in the project - mainly residents of Cheshire, Greater Manchester and Merseyside, mainly working at home. In Stage One of the procedure, volunteers transcribe from 'hard copy' (reprographic copy) of sections of individual registers, the hard copy generated from the relevant microfilm in the record office. Most volunteers therefore work on post-1700 easily-read material, while the earlier, more difficult but less abundant material is transcribed by a small number of specialist transcribers. The transcript of each volunteer is checked by a fellow worker.
Standardized transcription forms were designed at the beginning of the project, and are issued in ‘pads’, together with a book of detailed standard instructions. Despite the patent theoretical difficulty that the standardization had to precede the mass transcription, that is, before the range of anomalies in the material could be ascertained, in fact both the forms and the instructions have worked very successfully, on the whole. To date only one major revision of the instructions has proved necessary and that was at an early date.

An outstanding feature of the project is that the transcription takes note of the corresponding BT (if extant), this forming Stage Two of the procedure. Without going into details, it must be said that this extension of the evidential base has created a further dimension of difficulty when planning both the standardized transcription and the computer programming. The noting of BT variants (including additions), together with any necessary checking of the original registers (when the reprographic material is difficult or illegible), is undertaken by a team of volunteers working at the record office in Chester.

An absolutely basic point we have had to bear in mind, is that once volunteers have got under way, it is impossible to stop work when an improvement or correction in the system of transcription of the kind that affects all the earlier work is discovered - you cannot take on 100 volunteers and after years of labour tell them that their work is scrapped and they must start again. Fortunately, after the major revision mentioned above, we have needed only minor improvements, and we have a way of introducing these. Pads completed by volunteers and incorporating any BT material are ‘edited’ by a group of trained volunteers working under the direct supervision of the organisers. Since these editors check all pads, partly to deal with transcribers’ queries and to adjudicate on possible BT variants, they are thus in a position to insert minor improvements. This is Stage Three in the procedure. The instructions for transcribers need not change, the instructions for editors can easily be modified, and have been. However, such modifications have arisen mainly in respect of the occasional anomalies that turn up when dealing with hundreds of thousands of PR entries, anomalies that call for an additional rule. Once the transcription has been finalized, we are prepared to consider granting access to these manuscripts, a courtesy that has already proved of some help to a few local and family historians.

In 1978, probably the most novel part of the project was the intention to process and make accessible the transcribed material by computer. A pilot project at Liverpool in 1979 enabled MSC labour to input, by punch card, 10,000 entries, using experimental mainframe programs designed by volunteer programmers. Since then, these programs have been re-designed thrice, and input via micro-computers (currently the BBC series) has come to replace punch cards. Our present computerizing procedure is as follows. Transcribed and edited material is inputted on micro-computers, so far mainly at Manchester, forming Stage Four of the project. In Stage Five, the inputted material, in the form of ‘files’ on floppy discs, is checked (both
machine-validated and visually) and if necessary corrected. **Stage Six** sees the material from the floppy disc transferred to the mainframe, that is, the university computer.

Our intention is that ultimately the data-inputting, like the transcription, will be a home activity, to be carried out by volunteers who own or have access to micro-computers and who will be issued with instructions and program discs. Once again, then, the project is designed to involve and be dependent on volunteers among the general public. For inputting on micro-computers, like the transcription, requires both careful, intelligent attention and a considerable expenditure of time.

We estimate that the PRs of the ancient county of Cheshire between 1538 and our cut-off date of 1871 contain some three million entries. To date we have transcribed about a tenth of these, covering segments of about one third of the parishes. Computerization made a later start but is now speeding up. Currently we have 40,000 entries on the computer (on ‘tapes’), in a form in which they can be interrogated. A further 30,000 baptism and burial entries have been inputted but await checking and transfer to the mainframe. When the transcribed material has been computerized in sufficient quantities, perhaps in whole parish units, it will be made accessible in appropriate forms. At the moment we favour microfiche, and we have produced a trial fiche (of part of the PR of Macclesfield Christ Church) - now consultable in the record office.

**Problems**

Although Brian Redwood retired in 1986, the Cheshire RO continues to lend the project essential support. But the organisation of the project has always been undertaken by the authors, neither of whom can spare it more than a portion of his academic time. It was therefore necessary to organise the project in such a way that it was largely self-running. We have had to use volunteers almost entirely, because the project has run since 1979 on a shoe-string. Small annual grants from local government sources, together with overheads carried by the record office and our universities and our departments therein, have been its only financial support. Our expenses have been mainly on the printing of the transcription pads and the purchase of computer discs and tapes. The question that anyone considering a similar project in another English county will ask is therefore this - how practical is it to run such an ambitious project with volunteers only and without massive financial support?

Although Cheshire local history and family history societies have shown interest in the project, and do frequently inquire when our material will be available to them, recruitment of volunteers has seldom resulted from direct appeal to the local societies. Recruitment has been principally by personal approach, volunteer to volunteer, although many volunteers turn out to be members of local societies. The modest speed of the project
inevitably causes wastage of volunteers who lose interest, and we regrettably also lose those who move away or die. Yet, in general, our problem has not been that of finding volunteers, but of organising them. Since it would not be practical for us to deal directly with individual transcribers, we recruit local organisers, who then handle local recruitment and who pass the material (pads, instructions, etc) up and down the chain. Shortage of organisers limits the localities in which we have active groups of transcribers. We meet our organisers at intervals, to discuss progress. Editors, who work with us at Liverpool or Manchester on a more or less regular weekly schedule, are more difficult to recruit; and we are deeply grateful for the devotion of our present long-serving colleagues. The authors divide Cheshire and organise the West and East sides respectively; but to ensure consistency we and our editors meet regularly and exchange queries.

Weeding out the occasional volunteer who cannot follow instructions or whose transcription is more illegible than the original takes tact. But almost invariably we gain volunteers who tackle the tedium of transcribing with enthusiasm and care. They are encouraged to work in pairs, which enables easy mutual checking. Although they often prefer to work on the register of their own parish, they must undertake to accept any work they are given, both because the process of obtaining hard copy is such that we cannot guarantee always to have specific material available, and because we wish to concentrate effort on a reasonable grouping of parishes. Whether we will be as successful with volunteers for computerizing remains to be seen - we may have over-estimated the rate of spread of home computers. We certainly under-estimated brand incompatibility.

Certain problems with the computing will be familiar to anyone who has worked on a large-scale computer project, and there is no doubt that we under-estimated these at the beginning. Data input proceeds at a steady pace, but the simple logistics of keeping a check on the whole process whereby a transcription pad moves into and through the computer system, are laborious and time-consuming. The scale of the data files necessitates storage for mainframe use on magnetic tapes. Problems of strained compatibility between our two university computers have proved at various times exasperating. The technology in hardware and software available at the start of the project has significantly improved over the period, yet it has not proved possible to update some of our handling routines to take advantage of the benefits - not at least without paying a price in terms of work by transcribers and inputters that would have had to be scrapped, a price we thought too high. Two changes of mainframe computer have proved devastating, causing delays of probably three years in the development of data handling programs. Finally, there has been a change in the appreciation of computer data by historians: they are now more sceptical of the value of unprocessed data banks. As far as this project is concerned, with its emphasis on producing a computerised edition, this is not a major objection, and in any case we believe that our data files are capable of further manipulation for more specialised pieces of research.
Why bother?

The recent realisation that demographic history provides an entry into the history of the masses in earlier centuries has concentrated attention on the PRs. But to date all historical studies based on PRs have had the limitation that the registers could only be considered individually, and hence multi-parochial conclusions could only be drawn by averaging the individual statistics of individual parishes. Integrating the data available in the registers of a group of parishes was very difficult. Most manuscript PRs have no index; the printed PRs with an index are only a small proportion of surviving PRs; and there are almost no district or regional indexes covering a number of parishes - let alone county or national indexes. Hence inter-parochial migration cannot yet be tackled on any scale, although some valiant pioneering research has suggested that migration may turn out to be the dimension which, when known, modifies certain of our present conclusions about English population history. Hence we in Cheshire are trying to achieve a county index that might be a model (warts and all) for a national index. Our computerized index would be several centuries ahead of the St Catherine’s index, which only alphabetizes surnames - we will be able to sort most categories of data within PR entries. That is, we can isolate and analyse all the entries relating to, for instance, cordwainers, or children dying 0-5.

Archivists are happy to cooperate because parish registers are the most used documents in record offices and the flood of customers is wearing away the originals (but the customers complain about having to use microfilms - or having to make a six-weeks-ahead appointment to get on the machines). Most of the current pressure on record offices comes from the booming interest in family history, and area indexes will be a godsend to the family and local historian, as well as to the academic researcher into population history.

Is the attention we pay to BT variants worth it, particularly since only a small proportion of entries turns out to have variants? There can be no doubt that introducing BTs into our project has slowed it down, and made both the transcription and the computer work more difficult. Our defence is that we have learned that the PR-BT relationship is a complex, inconsistent and often puzzling one. Sometimes the BT is not a pale copy of the PR but an independent source, with entries not in the PR or with additional detail. The differences between these texts may well throw important light on the process of compilation of the PR - and hence confirm the value of the inclusion of BT 'variants' in editions of registers. From the point of view of the project, it might, in hindsight, have meant faster progress if separate files for PR and BT had been provided for and been compiled. Yet this would have meant the gross duplication of many identical items, and complicated, if not inhibited, aggregative analysis of the data.
How scholarly?

The question we are most frequently asked is whether, using volunteers, our transcriptions are accurate. No transcript is a facsimile and therefore literally exact, and no transcription is wholly accurate. Our transcriptions are checked and checked, as are the computer data files, since every stage of processing involves the likelihood of additional error. It is all very tedious, but we claim that our end-product is at least as accurate as the best of other PR editions. Moreover, it is consistent from one PR to another, since we follow the same system of transcription throughout.

Our transcription system, like all such systems, is not wholly literal, in our case partly because we wanted to limit the tediousness to our volunteers. Thus, we do not require them to write 'son' and 'daughter' in every baptism entry, but only M and F (or, in rare cases where gender is ambiguous, U); and a few other categories of data are similarly coded. Names and occupations are copied literally. Variants in placenames are assembled in editorial notes; any non-demographic material inserted in the register is listed in these notes. Nothing that is in a register is omitted, being either transcribed literally, or coded, or noted.

Was our aim too ambitious? Perhaps so, although not necessarily in respect of tackling a whole county, since what has held us up is the planning, organising and programming, as much as the amount of material. Perhaps what we have done will help others to telescope this part of the game. As to the length of period to be tackled, undoubtedly we chose 1871 as the end date for ambitious reasons. Our notion was to overlap the censuses of 1851-71 and the secular registration records of the 1860s - these being more accurate than their predecessors - because we believed that our computerized programme could provide a matrix into which these other nominal-linkage records could be inserted. This would enable a check to be made on the comprehensiveness and accuracy of both the nineteenth century PRs and the secular records. (We went no further than 1871 because at the date of decision no census later than 1871 was open). We also thought that our matrix would accommodate other categories of nominal-linkage records, such as probate records and also certain nonconformist registers. We certainly did not lack ambition, but have of course had no time to attend further to these thoughts.

When available?

The question next most commonly asked of us is this - when will it be available? Even our transcribers are often disappointed when our material does not immediately reach the stage of computer print-out. Our aim is to make all the Cheshire Project material accessible via computer process as soon as is consistent with the progress of the project. To finance the project we visualise a charge for libraries and research institutions wishing to hold large quantities of computerized data, and for those making business profits out of the material. But we hope that this will enable the project to
offer free access to Cheshire local and family historians, to ‘poor scholars’, and to bona fide academic researchers. Producing microfiche will cost money - and threats from various quarters of the imposition of access or handling charges may not always prove hollow. We have therefore to steer a course between maximising scholarly access and bankrupting the project, the latter a tack that would do no-one any good. However, until we have computerized more material, we need take no final decisions on this tricky issue. But the project will not attain its scholarly aims until all its material is available.

It is now clear that the project will take somewhat longer than first envisaged - to be frank, about two decades longer, if we may judge by the current rates of transcribing and computerizing. At the age of one of us this is an interesting forecast. However, if the same calculation of rate of progress be done in relation to other long-term PR transcription projects, no county PR society continuing to use its present method of having registers individually edited and then printed in book form will put itself out of business before the year 2400. The difficulty is, of course, that there are only two of us, and not only our personal but our university life-spans are uncertain. We have therefore thought of turning the project into a PR society, even although we shudder at the thought of the complications in many directions this would cause. Perhaps the present note will produce a rush of good advice - or even a rush of successors.

Note

We cannot list by name our many volunteers, other than our indispensable chief programmer, John Clegg of Manchester Computing Centre. Nor can we mention by name all those in our universities and departments who have advised us or made our travail possible, or all those in the Cheshire Record Office who have taken on extra duties to help us. All the institutions mentioned have provided essential facilities. Our finances, after a 1978 MSC grant, have come from Cheshire County Council, Greater Manchester County Council and Greater Manchester Record Office.