

RESEARCH NOTE

VICTORIAN PANEL STUDY

Kevin Schürer

The idea of generating a Victorian Panel Study (VPS) arose from an initiative taken by The National Archives (TNA, formerly Public Record Office) to enter into collaborative agreements with appropriate Higher and Further Education (HE/FE) stakeholders in order to generate new Information Technology and Communications (ITC) resources to the mutual benefit of both parties. In the discussions between TNA and the ESRC that followed, the idea of jointly creating a VPS was first raised. An outline plan was subsequently presented to the ESRC's Research Resources Board (RRB) who initially agreed to fund a pilot project.

In outline the proposed VPS would take as its base the individuals and households recorded in the existing ESRC-funded computerised national 2 per cent sample of the 1851 British census, created by Professor Michael Anderson, and trace these through subsequent registration and census information for the 50-year period to 1901.¹ The result would be a linked database with each census year between 1851 and 1901 in essence acting as a surrogate 'wave'.

This project is timely because great advances have been made in recent years in terms of the creation of computerised and searchable indices to major collections of historical sources. In relation to the proposed VPS, of particular importance are the national databases that exist for the censuses of 1881 and 1901. Following on the success of creating the 1901 internet-based resource, TNA is already committed to a programme of computerising the remaining nineteenth-century censuses for England and Wales. Likewise for Scotland, where the census has been administered separately from that of England and Wales since 1861, plans are well advanced to digitise and index all remaining un-indexed censuses. In addition to this massive programme of census computerisation and indexing, the Office for National Statistics (ONS) has announced its intentions to produce a searchable database of the 'historic' civil registers of births, marriages and deaths.² Although the precise timescale for the completion of these indexing projects is not available at present, it is expected that they will be generated over the course of the next five years. In the case of Scotland, for which all the civil registration material has already been computerised, as has the 1891 census, it is anticipated that work on most

of the remaining censuses will have been completed by the end of 2004. Although these resources are being created primarily with the huge genealogical and family history communities in mind, they potentially offer significant opportunities to the academic research community as well, particularly in facilitating the possibility and potential for tracing individuals over the course of the second half of the nineteenth century.

The primary purpose of the pilot project is to test differing sampling and methodological issues, examine the linking problems involved, investigate the relationship between the VPS and other longitudinal data projects (both contemporary and historical), explore the strategic partnerships which will support the VPS, and moreover, to recommend a framework and strategy for creating a full VPS. In assessing and testing the possibilities and methodological issues relating to the creation of a national VPS for the period 1851 to 1901, the pilot project will encompass the following key aims and objectives:

- 1) to examine methodological issues including different cohort and panel approaches, different 'target' and 'observation' strategies and the problem of refreshment;
- 2) to investigate potential user needs;
- 3) to explore and discuss the potential involvement of other stakeholders and to develop working arrangements with volunteer researchers;
- 4) to prepare initial designs for a linkage database and to investigate possible automated record linkage strategies;
- 5) to assess coding and standardisation schemes used in other longitudinal studies and investigate what other source materials could feasibly be linked to a VPS subset.

An important early task for the pilot project was to identify what parishes or other census tracts it might focus on in order to test linking techniques and other methodologies associated with longitudinal studies. Ideally, the pilot requires areas for which (i) computerised census materials already exist; (ii) the civil registration information for this period has already been indexed; (iii) there is a mixture of settlement 'types', for example, urban, rural, market town, etc, and (iv) where there are active family history or similar societies from which it would be possible to recruit and test volunteers. Thus an initial task for the project was to undertake a survey of what existing materials and sources exist and to select a sample from this for testing purposes. This task also investigated the possibility of including some Scottish materials in the sample.

The pilot will need to examine the possibility of extracting sub-samples from the existing 1851 sample and test the validity and properties of the 1851 sample as a basis of a representative longitudinal database. In particular, the inclusion

of institutions within the 1851 sample needs further exploration as initial investigation suggests that these may be statistically unrepresentative. As part of this task a fact-finding visit was made to Professor M. Anderson, the creator of the 1851 sample, to check various details with him and also to consult additional materials concerning the sample in his possession. Professor Peter Lynn, a recognised expert on sampling, will also be consulted in relation to this task.

The pilot will examine and test various strategies for automated record linkage. In particular, it will investigate the work currently being undertaken at the Cambridge Group for the History of Population and Social Structure on the Kilmarnock/Skye/ Torthorwald reconstitution project, the work of Peter Tilley on the Kingston local history project, the work of the Minnesota Population Centre on linking US census samples, and that of the Swedish Demographic Database. Various methods will be applied to the test data and a longitudinal database will be constructed from the selected test materials. This linkage work will also necessitate work to be undertaken at the Family Records Centre and the purchase of several sample civil registration certificate copies as part of the process. The purchase of the civil registration certificate copies is vital not only to act as a check on the linkage work, but also to assess the accuracy and completeness of the information contained in the certificates. This task will create a series of longitudinal bases for future testing. It will also establish a data design for potential future use in a full VPS. This aspect of the work will be undertaken in conjunction with discussions with experts in this field at the ESRC UK Longitudinal Studies Centre (ULSC) and others working on present day longitudinal databases.

An important task for the pilot project will be to use the linked test data to explore the feasibility and sampling implications of differing observation rules. This task will form a large component of the work of the pilot in the second half of the project. The linked test data will also be used to explore and measure the sampling implications of differing data refreshment strategies. One particular aspect to be studied under this task are the possibilities and problems associated with the cluster nature of the 1851 sample, with whole enumeration districts as the sampling unit creating potentially a geographical component to the VPS, absent in most other longitudinal studies. Not only are there potential advantages in making the VPS comparable, where possible, with other historical national longitudinal databases, but in order to facilitate the use of the VPS in conjunction with present day longitudinal data sources to examine social scientific issues in the longer-term, it is also important to examine the potential for harmonisation with other present day databases. This task will focus on the potential database design of the VPS and investigate the possibility of mapping coding schemes between the VPS and other longitudinal studies. It will be undertaken in consultation with experts working on present-day longitudinal studies based at the Institute for Social and Economic Research, University of Essex and at the Institute of Education, University of London.

The success of a full VPS will in part depend on data materials being made available from a number of key stakeholders including:

TNA, regarding the availability of census data and indices for 1861 to 1891;

ONS, regarding the proposed changes relating to changes to the historic civil registration information;

The General Registry Office, Scotland (GROS), regarding access to the Scottish census and civil registration materials;

QinetiQ, regarding access to the 1901 census material;

FreeBMD, regarding access to their database of the indices to the English and Welsh civil registers, as well as their related census database;

Genealogical Society of Utah, regarding access to their database of parish register materials;

Discussions are being held with these various organisations in order to establish their future plans and time scales, and to develop future working relationships.

Obviously the main purpose in creating a VPS would be to promote and produce high quality research. The pilot study will undertake a potential user consultation exercise, including the wider social scientific community in addition to economic and social historians. Work will be undertaken to try and ascertain research requirements from the VPS database so that these can feed back into the pilot project and specifications for a full project. This consultation work will be undertaken chiefly in two phases. An initial phase will be undertaken through virtual email based surveys and follow-up meeting, as appropriate, when the test database is being constructed, in the hope that it will feed back into this process. The second phase will consist primarily of a user consultation workshop to be held in London in the later stages of the pilot project, at which preliminary outcomes of the project will be presented and feedback and responses sought.

In planning the VPS it is desirable not only to learn from the experience of other national historical longitudinal database projects but also to ensure where possible that the VPS is designed in such a way as to allow cross-country analyses to be undertaken. For these reasons the pilot project will investigate the work of other national historical longitudinal database projects, in particular the Historical Sample of the Population of the Netherlands and the project linking census samples from the US censuses based at the Minnesota Population Center. From a research point of view it may prove desirable to link (implicitly or explicitly) records from the VPS to other economic and social historical materials. The pilot will investigate the potential for this and suggest what might be considered as candidate materials.

Due to the effort potentially needed to create the VPS in terms of person hours, it is proposed that volunteers from local history groups and family history societies be enrolled to help with the task of checking and verifying links and possibly capturing additional civil register and census information. This model

has previously worked satisfactorily in related projects based at the Cambridge Group for the History of Population and Social Structure and the Universities of Hertfordshire and Lancaster. The research team will seek to identify groups and organisations that would potentially support and promote the VPS. It will also work with a small body of initial volunteers to develop prototype data entry software that they might use for capturing data, together with a set of procedures and instructions for managing the work of volunteers.

There will be two main deliverables from this pilot project: a detailed report and a series of linked databases. The detailed report will be submitted to RRB in order that the Board can consider what the next steps for the proposed VPS should be. Drawing on the experience of the pilot project, this report will make specific recommendations on how the VPS might proceed, including specifications for a full project, management requirements and projected costs, in a form that could be used as a basis for a future call to tender. The VPS pilot project will create several linked longitudinal databases, developed as part of the testing procedures. Although limited in scale and scope, these will have research and teaching potential beyond this pilot project and therefore will be offered for deposit with AHDS History (formerly the History Data Service), one of the five centres of the Arts and Humanities Data Service located within the UK Data Archive.

NOTES

1. ESRC award RH00232032. The data resulting from the project are available as Anderson, M. et al. *National Sample from the 1851 Census of Great Britain* [computer file]. Colchester, Essex: UK Data Archive [distributor], 1979. SN: 1316. The original data deposited with the UK Data Archive have subsequently been completely checked, cleaned, recoded and reformatted at the Department of History, University of Essex to make them directly comparable with the available data for the 1881 census. See Schürer, K. and Woollard, M. *1881 Census for England and Wales, the Channel Islands and the Isle of Man (Enriched Version)* [computer file]. Genealogical Society of Utah, Federation of Family History Societies [original data producers]. Colchester, Essex: UK Data Archive [distributor], 2000. SN:4177 and Schürer, K. and Woollard, M. *1881 Census for Scotland (Enriched Version)* [computer file]. Genealogical Society of Utah, Federation of Family History Societies [original data producers]. Colchester, Essex: UK Data Archive [distributor], 2000. SN:4178.
2. See Office for National Statistics, *Civil registration: delivering vital change*, (London, 2003). It is not known at this stage what will be included within the 'historical' category, and this will most probably vary depending on the nature of the event, but certainly all records relevant to the VPS would be 'historical'.