



social research UPDATE

- Databases and directories are useful tools that can be considered for methods of sociological investigations.
- Whilst such products are useful, the process of creating them is fraught with practical and theoretical challenges.
- Drawing from our experience in researching sections of the voluntary sector, namely faith based groups and women's organisations, we discuss some of these issues.

Tracing and Mapping: the challenges of compiling databases and directories

Andri Soteri-Proctor and Greg Smith

Databases can be seen as electronic filing systems used to store and retrieve information for both practical and analytical purposes. In social research, for example, databases have been used as a tool for sampling or to analyse information. Commercial companies use databases to create products such as the *Yellow Pages*, and libraries use them to keep records on, for example, books and book loans. In the voluntary sector, practitioners and researchers have compiled and used databases in 'tracing' and 'mapping' exercises to examine the size, nature and value of the sector, and to identify gaps and unmet needs to inform policy making (Marshall 1997; Lewis 2001). Other practical uses of databases include building information and networking resources, as databases or directories in electronic and paper form.

Technological advances, increasing skills and better access to equipment have provided the potential to add, amend and merge existing information and have enabled a multitude of opportunities for sharing data, which can be accessed locally, nationally and internationally. There are opportunities for storing, retrieving and analysing data through a number of software applications. These include simple non-relational (flat file) databases and spreadsheets, relational databases such as MS Access and similar products, complex quantitative statistical packages such as SPSS, and network

mapping packages (such as Pajek, <http://vlado.fmf.uni-lj.si/pub/networks/pajek/>), which is useful for showing relationships between, for example, individuals or organisations. There is also a significant development of Geographical Information Systems (GIS), which enable visual mapping of services or resources, with data located at points in space and collected by a variety of methods, from asking interviewees for their home postcode, to photography from "spy in the sky" satellites (Marble and Peuquet 1990; Goodchild, Steyart et al. 1993). For an extensive resource list about GIS see <http://www.geo.ed.ac.uk/home/giswww.html>.

Databases are, however, simply a tool and they do not eradicate the challenges involved in the processes of compiling and using data. For the purpose of this *Update*, these processes will be referred to as tracing and mapping exercises.

Tracing is the process of finding or tracking information for the database and, ultimately, for the research. Depending on the access, amount and quality of existing information, this process can be time and labour-consuming. It is predominantly investigative and involves making a number of decisions about, for example, what or who to look for, what information is available and how to find it.

The analysis of information collected and stored in a database is called the *mapping* process, at least in the context of the voluntary sector (Lewis 2001).

Greg Smith is Senior Research Fellow at the Centre for Institutional Studies, University of East London. He is undertaking research examining children's understanding of faith and the role that it plays in their lives. He has researched and published extensively on urban religion and community development. During the 1980s he worked as a member of the Linguistic Minorities Project, which produced the standard work on "The Other Languages of England".

Andri Soteri-Proctor is Senior Research Fellow at Centre for Institutional Studies, University of East London. She is currently undertaking research examining children's understanding of faith. She is also a PhD student at Manchester University, evaluating the social and community benefits of aspects of the employment initiative, New Deal for Young People. Previous to this, she carried out research examining issues related to funding and capacity building of women's organizations.

social research UPDATE

While mapping may conjure up images of geographical maps, it can also include analysis of relationships between the actors and their literal or metaphorical construction of space and place, although sometimes it does not even involve this.

Challenges

This *Update* focuses on issues related to tracing rather than mapping. However, rarely do these processes exist independently when undertaken for research. The aims of mapping data often influence the content of information collected for the database and the type and quality of information collected influence what can be mapped or analysed. Some of the challenges involved in compiling databases are specific, relating to the equipment and skills available to the researcher, other issues are more general. Defining the target group is a common challenge that cuts across most research undertaken; and defining who or what is to be included and excluded is an essential part of compiling a database, especially one that is being used for sampling purposes.

To ensure the maximum inclusion of target groups or individuals, especially the hard-to-reach, can require a huge amount of work, irrespective of whether this information is tracked first-hand or through existing directories and databases. Determining whether those tracked are members of the target population often requires more information than that collected and it is therefore not unusual to have to undertake enumeration inquiries to check for their eligibility. This can help determine eligibility for those already listed, but it is not always easy to ensure the inclusion of all potentially eligible research subjects.

Transient populations and those that lack documentation are particularly difficult to trace. In the voluntary sector, for example, it is relatively easy to obtain details about registered charities from the Charity Commission website and other databases and directories (some of these sources are listed in Soteri 2002). However, community-based, small, and informal organisations that are non-profit, but are not registered, are difficult to track down and probably constitute the majority of such organisations

(Halfpenny and Reid 2002). Researchers often need to consider ways of finding these groups, going beyond existing directories. There are a variety of ways of doing this and researchers' existing knowledge and networks can help, particularly if the research is being carried out by someone well established in the locality or the field. For the *Newham Directory of Religious Groups*, two decades of local involvement in the faith sector contributed to its success (Smith 2001). While such efforts can be manageable for locally specific research, it is more difficult to sustain with larger or more general target groups. Snowballing and/or network sampling techniques can be used, but there is a danger that the coverage is then determined by the starting point. Substantial alternative networks or agencies that do not link to the starting point may be excluded.

Throughout the process of tracing, it is essential to refer to data that has already been collected in order to avoid duplicating records. Cross-referencing and checking records that may appear to be duplications is not rocket-science, but such exercises can be time-, cost- and labour-intensive, particularly when dealing with large data sets. In our research projects it was not unusual to come across organisations with the same name. Possible reasons for this are that there is more than one organisation with the same name; that an organisation has more than one address; and an organisation has moved from one, or even both, of the addresses. It is not unusual to encounter contradictory information from different, or even the same sources. A useful discussion about the challenges of sampling frameworks and using 'imperfect frames' can be found in Kish (1965).

Financial resources often dictate the method of approaching the target group and postal (or, nowadays, email) questionnaires are commonly used. However, the level of response associated with these methods is often low. While re-mailing and intensive telephone follow-up can raise response rates, this can be expensive and time consuming. Researchers need to consider whether the cost and effort can be justified. Determining this point of

social research UPDATE

diminishing returns is dependent on the purpose of the work, time and budget. There is also the issue of how to interpret non-responses, which could be due to, for example, an organisation closing or moving, as well as choosing not to respond.

It is difficult to calculate accurate and realistic costs for compiling databases and directories since they are not easily separated from other parts of the research process. Costs will also vary from project to project. The scale of such an enterprise can however be judged by the work carried out to create the *Directory of Religious Groups* in Newham. This involved the work of a research officer roughly half time over a year (say £15k including overheads); approximately one day a week over the same period of an IT professional seconded from the local authority (say £10k including overheads); costs of design and printing of 500 copies of the directory (£3k). In addition the project had the unpaid help of a student on full time placement for most of the academic year, a team of a dozen or so American students working over a two month period; and several occasional community based volunteers. While the project was never costed as a specific budget item, it is clear that including everything from basic research to printing and not very extensive marketing of the directory, such a project would needed a grant of between £50 -100k to be adequately funded.

Discussion

The usefulness of information technology in the research context is undeniable. IT has opened a world of opportunities for researchers. However it would be unwise to assume that it has contributed to making tracing and mapping exercises less labour and cost-intensive. Initially at least, they rely heavily on the manual efforts and 'legwork' of researchers and administrators to search and collect data. In local studies practitioners, networkers and community based volunteers may be better placed to gather the information but they are not easy to recruit, recompense or manage.

One way to counter some of these problems is to develop inter-agency partnerships. Other benefits of this

approach include the forging of mutually supportive relationships, the sharing of information and helping to minimise the duplication of work. Indeed the process of information gathering, if conducted in a collaborative and participatory fashion, can become a useful process of community development. However, the potential of mixed-purpose research and the creation of joint ventures between researchers and practitioners may open up further practical and political challenges. Researchers and practitioners (and service users) may have different priorities and values that are not obvious during the early stages. Although a clear contract between parties can help, conflicts can emerge.

The use of IT has also raised increasingly complex political, social and legal issues about data protection and copyright in relation to users, the data stored and the sources of information. Issues of access are more about the functions of data regulation than technological capabilities. Current legislation can be bewilderingly complex for researchers and data users in the voluntary sector and needs to be simplified by some form of overarching body. It is not always obvious who ultimately owns the data or who should be registered as the data holder and whether the data may be given away or sold to all comers.

Other issues include whether it is ethical and/or legitimate to disclose or publish details of those on the database. In the case of voluntary organisations, for example, it is not unusual for some groups to operate from a residential address or to work on sensitive or controversial services such as women's refuge, racial harassment, gay helpline services, or abortion clinics. If the information is being collected directly by questionnaire a clear contract can be spelled out with responding groups, but where secondary sources are drawn on or there are possibilities of third party access and re-dissemination of the data these issues are not always easy to resolve. Of course such considerations also apply to paper publications, but the ease of electronic data reproduction and transmission make them more salient than before.

We have already argued that costs of tracing

social research UPDATE is distributed without charge on request to social researchers in the United Kingdom by the Department of Sociology at the University of Surrey as part of its commitment to supporting social research training and development.

Contributions to **social research UPDATE** that review current issues in social research and methodology in about 2,500 words are welcome. All **UPDATE** articles are peer-reviewed.

social research UPDATE

and mapping exercises are very high. It can therefore be difficult to obtain grant funding for high quality comprehensive information gathering work on a sustainable basis. In the commercial world, up-to-date databases do seem to have a market value as companies will often pay to be listed and others will pay a high price for access to information. Even where there is web access, restricted subscription services generate income for information providers. However, the culture and poverty of the voluntary and some other sectors tend to prevent sales at a market price which will cover costs. Finally, there is a glut of freely available information of questionable accuracy, currency and relevance, which tends to drive down the price of data. The growth of the web, in which almost every organisation now has its own web pages to describe and publicise its work in its own words, may make external compilation of information less necessary.

Conclusion

The sophistication of information technology has contributed to databases becoming an invaluable tool for social research for sampling and for storing information for research analysis. In this *Update* we have discussed some of the challenges involved in the compilation of databases and directories in the context of research that we have undertaken in the voluntary sector. However, they are not exclusive to the voluntary sector. For example, Bulmer et al. (1998) discuss similar problems in a different field. If such issues are not adequately addressed, the quality of information collected and stored will be limited and questionable and, ultimately, will affect the quality of analysis undertaken. A similar point is made by Dey (1993) in his discussion about the role of computers in qualitative research.

Researchers considering compiling their own database should err on the side of caution. Whether it is worth undertaking such a costly and time-consuming task is difficult to answer, because it is so dependent on the purposes of the particular tracing and mapping exercise and because the value of the product that is created may only become clear retrospectively.

Bibliography

- Bulmer, M., W. Sykes, et al. (1998). *Directory of Social Research Organisations in the UK*. London, Continuum.
- Dey, I. (1993). *Qualitative Data Analysis. A user-friendly guide for social scientists*. London and New York, Routledge.
- Goodchild, M. F., L. I. Steyart, et al. (1993). *GIS and Environmental Modelling. Progress and Research Issues*. Oxford University Press.
- Halfpenny, P. and Reid (2002). "Research on the voluntary sector: an overview." *Policy and Politics* 30(4): 533-550.
- Kish, L. (1965). *Survey Sampling*. New York, Chichester, Brisbane, Toronto, John Wiley & Sons.
- Lewis, G. (2001). *Mapping the contribution of the Voluntary and Community Sector in Yorkshire and the Humber*. Yorkshire and the Humber Regional Forum.
- Marble, D. F. and D. Peuquet (1990). *Introductory readings in geographic information systems*. London, Taylor & Frances.
- Marshall, T. (1997). *Local Voluntary Activity Surveys (LOVAS)*. Research Manual, LOVAS PAPER 1, Home Office, Research & Statistics Directorate.
- Smith, G. (2001). "Religion as a source of social capital in the regeneration and globalisation of East London." *Rising East* 4(3): 128-157.
- Soteri, A. (2002). *Funding in London women's organisations*. London, Centre for Institutional Studies, University of East London.

social research UPDATE

(ISSN: 1360-7898)

is published by

Department of Sociology
University of Surrey
Guildford GU2 7XH
United Kingdom.

Tel: 01483 300800

Fax: 01483 689551

Edited by Nigel Gilbert

(e-mail: n.gilbert@soc.surrey.ac.uk)

Winter 2003 © University of Surrey