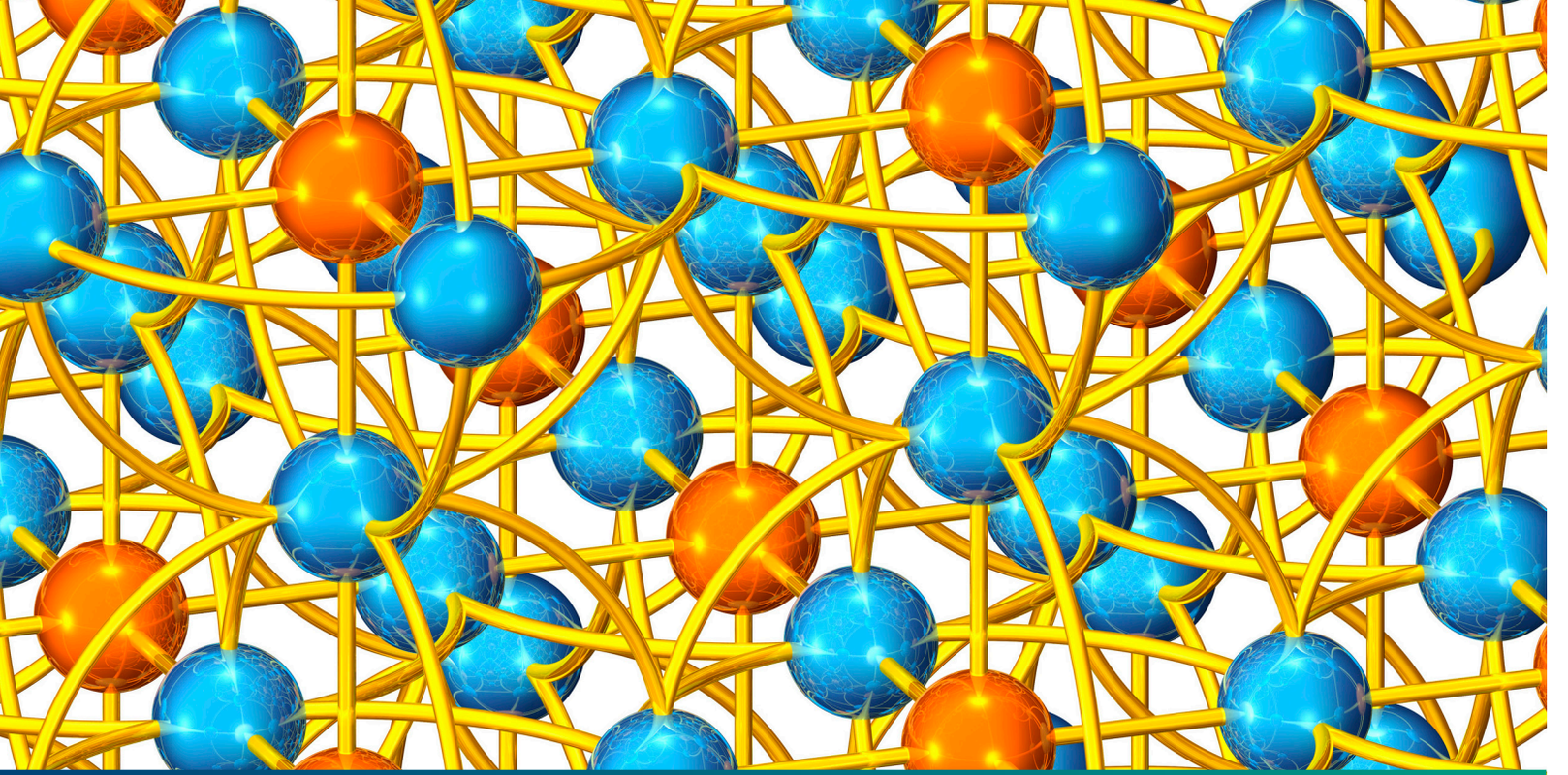

Making the connection

In the face of survey fatigue, how can researchers gather quality information? Data linking could not only add high value to existing surveys, but also create new data resources.
By **Professor Peter Elias**

Survey-based data collection methods underlie much empirical research in the social and economic sciences. Large-scale sampling and surveying of known populations is a tried and tested method used to create many of the major data resources supported by the ESRC. *Understanding Society* is a good example of such a resource, as are the Birth Cohort Studies and, at an international level, the European Social Survey.

Data collected by sample surveys are relatively expensive to compile. Whether through face-to-face or telephone interviews, the process is both time-consuming and costly. Cheaper web surveys are becoming increasingly popular, but for high-quality research in the social sciences these do not provide the same degree of control over the information collected that can be obtained via an interviewer-administered approach.

Response rates across all major surveys are falling, possibly because people are leading increasingly busy lives but also because of 'survey fatigue' – a lower willingness to co-operate as more organisations attempt to gather information from individuals by survey methods.

Data linking provides another approach that has the potential not just to add high value to existing surveys, but also as a means of creating new data resources which place no reliance whatsoever on survey-based methods. It takes advantage of our 'digital footprint' – the records we create as we go about our daily lives. These may record transactions we engage in, registrations we undertake, our communications with each other and web-based activity. Examples include administrative records (eg, social security payments, income tax payments, school and college enrolments, hospitalisation and GP records), transaction records (eg, mortgage payments, electricity billing, use of loyalty cards when shopping), internet activity such as web searching and use of social media, and remote sensing records (eg, road transport sensing devices). While these examples reveal the diversity of the electronic information we generate, they share certain common characteristics:

> "Issues must be resolved before data linkage is a regular part of developing data for research"

- they are, in most cases, personal records – they may contain detailed information about an individual that could be misused if placed in the wrong hands;
- they are not designed specifically as research resources even though the information they contain might have research value;
- they belong, in general, to big datasets covering large segments of the population.

Some countries have developed their use of administrative records to the point where these are now routinely substituted for the more traditional ways of generating data resources for social and economic research.

In Finland, for example, the population census is conducted not via the use of a census form delivered to households but from 30 different registers and administrative files. In the UK, it is within the devolved administrations that data linking has progressed to the point where it significantly enhances existing surveys and/or facilitates linkage across different types of administrative records. While these examples indicate the value of linked data, the approach towards more routine research use of data linkage procedures is rightly moving cautiously given the risks and problems involved. A number of important issues must be resolved before data linkage becomes a regular part of the process of developing data for research purposes. These include:

Maintaining security, preventing abuse of linked data

The personal nature of many of the types of data which constitute our digital footprint means that they could be misused in some way if they fall into the wrong hands. However, researchers require access to these data at a very detailed level if they are to be able to link them together and to understand their value as research resources. Careful controls over the conditions of access must be put in place to ensure that the data concerned are not used for inappropriate purposes or by unauthorised persons.

The need for consent to link data

For some categories of data and in some situations, legal and/or ethical considerations require that the individuals or organisations whose data are to be linked in some way should consent to this process. There is a need for clarification of this requirement and harmonisation of best practice in this area.

Assessing the quality of linked data resources for specific research purposes

It is often assumed that data from administrative sources give an error free, accurate record of the process from which they are drawn. This may not be the case. Given that such data are not designed for research purposes, their fitness for purpose as research resources must be carefully assessed.

Gaining access to and use of data which may have commercial value

Certain types of electronic records have significant commercial value. A good example is the loyalty card data generated by shoppers. Organisations may be reluctant to make such data available for research if they feel that this value may be exploited by competitors.

These are not easy issues to resolve. There has already been some progress on laying the foundations for what looks set to become a major new approach to data construction in the social and economic sciences. The ESRC has funded the Administrative Data Liaison Service and the Secure Data Service. The former service, managed by the University of St Andrews in conjunction with the Universities of Oxford and Manchester, houses and makes accessible to researchers a lot of information about the wide range of administrative data sources which are potentially available in the UK for research. The latter, which forms part of the ESRC-funded UK Data Archive at the University of Essex, provides a secure environment for data linkage and analysis, with remote access from UK universities.

Government departments and some private sector organisations have developed protocols that are designed to protect confidentiality and prevent inappropriate use of data in their safe-keeping. However, there is, as yet, no systematic cross-departmental approach to issues such as the situations in which consent is required for data linkage, the procedures through which researchers may gain access and the requirements placed on researchers that will ensure that data security is maintained.

For this reason, the ESRC, together with the MRC and the Wellcome Trust, has set up an Administrative Data Task Force (ADT) designed to address these issues across a range of administrative data types which have potential research value when linked to other sources of information. Chaired by Sir Alan Langlands (Chief Executive, HEFCE) and with strong support across government, the ADT aims to make rapid progress to resolve these issues in ways which will help promote their safe and efficient use for research. A particular concern shared by all those organisations that hold personal data is that the public should be confident that their information is held securely, used appropriately and that no harm should ever arise from misuse of their data in any way.

This cross-departmental and cross-funding agency activity should be seen as the very beginning of the ways in which these new forms of digital information will change the landscape of social and economic research. While no-one is predicting that the traditional approaches to data collection will disappear, innovative use of the many new types of data that we create will be high on the research agenda over coming years. ■

Professor Peter Elias, ESRC Strategic Advisor (Data Resources)
Email peter.elias@warwick.ac.uk
Web securedata.data-archive.ac.uk

Examples:
Scottish Health Informatics Programme (www.scot-ship.ac.uk), the Welsh programme for Secure Anonymised Information Linkage Databank (www.healthinformatics researchlabs.swansea.ac.uk/sailproject.html) and the Scottish Longitudinal Study (www.lscs.ac.uk/sls)
Web www.adls.ac.uk
www.understandingsociety.org.uk
www.cls.ioe.ac.uk
www.europeansocialsurvey.org