

Accessing and Analysing Social and Economic Data: Users' needs uncovered

A web-based consultation undertaken
by the

Economic and Social Data Service
(ESDS)

Universities of Essex and Manchester

May 2003

Not for citation without permission

CONTENTS

1. The user consultation	
1.1 Introduction	1
1.2 Scope and Methods of Consultation	1
1.2.1 Designing the questionnaires	2
1.2.2 Gathering Information	3
2. Results	
2.1 Current or potential use of data sources	4
2.2 Using data in research and teaching: barriers	4
2.3 User support and training	5
3. Key requirements and recommendations	
3.1 ESDS Data portfolio	5
3.2 Resource discovery and data delivery	6
3.3 Value-added data and documentation	6
3.4 User Support and Training	7
3.5 Conclusion and next steps	8
4. ESDS Government: User Survey Results	
4.1 Introduction	9
4.2 Respondents	9
4.3 Current or potential use of data sources	9
4.4 Using data in research: barriers and requirements	10
4.5 Using data in teaching: barriers and requirements	12
4.6 Online training and support requirements	14
4.7 Observations and recommendations	15
4.8 Appendix: ESDS Government User Survey Tables	17
5. ESDS International: User Survey Results	
5.1 Response rates	22
5.2 Current or potential use of data sources	22
5.3 Current or potential use of data sources	25
5.4 Using data in teaching: barriers and requirements	27
5.5 Online training and support requirements	28
5.6 Observations and recommendations	30
5.7 Appendix: ESDS International User Survey Tables	33
6. ESDS Longitudinal: User Survey Results	
6.1 Introduction	43
6.2 Respondents	43
6.3 Current or potential use of data sources	43
6.4 Using data in research: barriers and requirements	44
6.5 Using data in teaching: barriers and requirements	46
6.6 Online training and support requirements	46
6.7 Observations and recommendations	48
6.8 Appendix: ESDS Longitudinal User Survey Tables	49

(cont)

7. ESDS Qualidata: User Survey Results	
7.1 Introduction	51
7.2 Respondents	51
7.3 Current or potential use of data sources	51
7.4 Using data in research: barriers and requirements	53
7.5 Using data in teaching: barriers and requirements	56
7.6 Online training and support requirements	57
7.7 Observations and recommendations	59
7.8 Conclusion	61
7.9 Appendix: ESDS Qualidata User Survey Tables	62
8. Web Questionnaires	

1. The user consultation

1.1 Introduction

The Economic and Social Data Service (ESDS) is a new national data service that came into operation in January 2003. It is a jointly-funded initiative sponsored by the Economic and Social Research Council (ESRC) and the Joint Information Systems Committee (JISC). It is a distributed service bringing together four centres of expertise in data creation, preservation and use, based at the Universities of Essex and Manchester. These centres are working collaboratively to provide dissemination, preservation, user support and training for an extensive range of key economic and social data, both quantitative and qualitative, across the research, learning and teaching communities.

ESDS incorporates four 'specialist' data services, each of whose prime objective is to provide value-added and support services for particular types of data. These are:

- ESDS Government – for large-scale government data;
- ESDS International – for key international macro and micro data series and surveys;
- ESDS Longitudinal– for the major UK-based panel and cohort studies;
- ESDS Qualidata – for key UK qualitative data sources.

Key deliverables across these four data services are:

- enhanced data and data documentation products;
- online access to data via simple registration procedures;
- improved and tailored user support;
- workshops and training courses;
- online materials and a range of other value-added services to promote and encourage usage in teaching and research.

Further details of the service can be found at the new ESDS web site that is currently under construction (www.esds.ac.uk).

An early task planned for the services was a user consultation exercise. Phase I of this activity was tackled by undertaking a web-based user survey. Conducted in March 2003, the survey aimed to gather feedback from current and potential users of social and economic data on whether the data and support services planned by each data service will meet the needs of the UK academic community (see milestones under the relevant services sections at www.esds.ac.uk).

1.2 Scope and Methods of Consultation

ESDS aimed to gather user-oriented views and feedback about the four distinct kinds of data supported by ESDS value-added data services: government, longitudinal, international and qualitative. Stakeholders targeted were from across the social science communities and across disciplines - active researchers; higher and further education teaching staff; research students; and research support staff (e.g. data librarians).

For each data service, the survey had the following specific objectives:

- review current and future use of datasets in research and/or teaching;
- identify research topics and areas in which data are being used, or where data and documentation products would be desirable;
- identify any gaps in the initial data portfolio as part of the process of reviewing the ESDS data acquisitions policy;

- review what software is currently being used for the analysis of datasets;
- identify any problems currently encountered in accessing or using datasets in research and/or teaching;
- obtain feedback on the services or data related developments that would make it easier for users to access and use data in research and teaching;
- obtain feedback on the training requirements for data handling techniques and data analysis;
- obtain feedback on the specific training and support services to be provided by the service to help establish priorities.

1.2.1 Designing the questionnaires

With these objectives in mind, a set of simple survey questions was drawn up and divided into four sections covering each of the data types. While the majority of the questions were common to each type of data, additional questions that related to specific issues were also added. The surveys can be found appended at the end of each of the service-specific sections.

The content of the questionnaires was drawn, primarily, from the deliverables of each of the data services. ESDS's prime mission is promoting and supporting the use of data, in the areas of both research and teaching (in this case seen as distinct activities). Each data service has a provisional five-year plan with staged milestones and defined deliverables, based largely on knowledge built up about users and user requirements, by each of the ESDS constituent institutions. Between them, Essex and Manchester have over 50 years experience of servicing and supporting social and economic data users in the UK.

The consultation process aimed to help re-affirm, or where necessary, modify these planned deliverables and help set priorities in areas that were perhaps less well defined in the initial tender bids, submitted in 2002. First, one of the areas in which the services were particularly interested was the subjects or topics of greatest popularity amongst data users. Thematic data samplers feature heavily in the services' anticipated deliverables and it is important to select these according to user demand. Second, how useful do the kinds of data enhancements proposed by the services seem, and how should the work be prioritised? Third, providing quick and easy access to data is a key aim of the ESDS. What, if any, difficulties are currently faced by users in gaining access to data or using data? And what kind of data formats do users prefer?

Finally, a further important area is the provision of training and documentation to support secondary data analysis. In what areas of data handling or analysis techniques is help most needed? In what substantive areas would courses be best designed? What kinds of support materials do, or would users find most beneficial?

Each of the four service's questionnaire was thus divided into five sections:

- use of data sources and software;
- use of data in research;
- use of data in teaching;
- training and support for using data;
- general open comment.

The surveys were developed and piloted over a period of two to three weeks - internally among ESDS staff and via peer networks. Given the known problems of conducting successful web surveys, the surveys were kept simple and short with no more than 20 questions making up each section. Closed questioning was used for the majority of questions, with follow-up text boxes for some questions to elicit specific responses, and to enable respondents to provide free comment if desired.

1.2.2 Gathering Information

With ESDS collective branding in mind, the user consultations for each value-added data service were conducted alongside each other, presented in a common attractive format, layout and style. The surveys were mounted on the UKDA web site as html pages. Users were invited to complete the surveys online over a period of two weeks, with a prize draw of Amazon vouchers being advertised as an incentive. A single entry point to the surveys ensured that users with different analytic or teaching interests had the opportunity to answer more than one of the surveys (e.g. distinctly for government data and longitudinal data).

The survey was advertised on a number of relevant Jiscmail discussion lists and via the UKDA and MIMAS newsletters. A total of 367 responses were obtained across the four surveys, with the government data survey taking the lion's share (52%) of the total response, followed by the qualitative data survey (20%), and international and longitudinal data surveys (14% each).

Data Service	Number of responses
Large-scale Government	191
Qualitative	73
International	52
Longitudinal	51
Total	367

Responses to the specific surveys were collected, collated and analysed by staff within each of the four services. Section 2 sets out a summary of the overall findings across the four surveys, while sections 3-6 offer a more detailed but mostly descriptive analysis of the responses, reported by data type. Authors by section were:

- Executive summary and introduction – Louise Corti, UKDA
- Government data – Vanessa Higgins, CCSR
- International data– Keith Cole, MIMAS
- Longitudinal data – Jack Kneeshaw and Louise Corti, UKDA
- Qualitative data – Louise Corti, Karen Dennison and John Southall, UKDA

2. Results

This section summarises the overall findings across the four surveys. The surveys were particularly helpful in:

- pinpointing the most popular data sources;
- elucidating barriers to accessing and using some types of data;
- drawing out the key substantive themes of interest to users of data - researchers and teachers alike;
- Signalling demand for particular modes of user support and training resources.

2.1 Current or potential use of data sources

The results of the user surveys suggest that many of the most prominent data sources asked about in the survey are widely used. There is clearly also considerable scope for future use of both existing data within the ESDS remit and those data planned for acquisition by ESDS.

- The major government series are amongst the most heavily used data in the ESDS collection, with equally great use anticipated for the future.
- For longitudinal data users, many expected using in the future data not yet currently held by the ESDS, such as the Millennium Cohort Study (MCS) and the English Longitudinal Study of Ageing (ELSA), both of which are destined for acquisition. There was further demand for access to some key MRC studies, such as the ALSPAC and 1946 Birth Cohort, currently only available via negotiation with the principal investigators (and relatively high fees for ALSPAC access).
- Equally, for international macro data the costs of data acquisition are a major barrier to use which will be overcome by ESDS through the negotiation of UK wide academic redistribution licences with organisations such as the IMF.
- Resources directed at the supply and support of government, longitudinal and international data are currently, then, well targeted.
- Qualitative data sources are less widely used for research, as already suspected, but many respondents indicated that they would consider using archived qualitative data in the future. Low usage figures are due to the lack of sources available in areas of greatest demand, together with the hitherto inaccessibility of some qualitative data collections due to their non-digital paper format.
- Overall, UK government and longitudinal data, and international macro data series are used more for research than for teaching. However, qualitative data are used to a greater extent for teaching than are government or longitudinal data.

2.2 Using data in research and teaching: barriers

While the services currently provided by the UK Data Archive were praised by users, significant numbers, mostly less regular users, reported experiencing a range of problems accessing and using data in research and teaching.

- The main difficulties cited were: identifying appropriate datasets; problems with the complexity of data and with data documentation; accessing data using online systems, adequate data manipulation and analytic skills; and the cost of acquiring international macro data.

- As anticipated, there is an overwhelming call from users for quick and easy access to data, and online access to data is seen to be one way of meeting the needs of today's time -pressured researchers.
- The most popular substantive research themes identified by actual or potential data users were remarkably common across the UK data types: health, employment and the labour market, and social exclusion for government, longitudinal and qualitative data. International data users called were researching in the areas of economics, labour and employment and politics.

2.3 User support and training

Across the range of users represented by the survey responses, there appears to be considerable scope for increasing use of data by providing enhanced data support resources. The lack of widespread usage of data in the classroom is already known, as highlighted recently in a report resulting from a JISC project, 'Taskforce on The Use of Numeric Data in Learning and Teaching'¹. Whilst individual datasets are used extensively in academic research they are significantly under used in learning and teaching programmes within Higher Education, and rarely used in Further Education.

- Overall teachers, at both undergraduate and post-graduate levels, require more help in obtaining suitable materials for teaching.
- The most popular kinds of support advocated are: tailor-made, annotated training datasets, 'how-to guides', case studies of secondary analysis, online learning materials and tutorials.
- For teaching key methodological and analytic skills, training is often seen, and arguably should be essentially viewed, within a substantive context. As such, themed datasets were very popular.
- Across all surveys positive views were expressed towards the prospect of more training being made available via ESDS. Training on particular data types, datasets, software packages, data manipulation and analytic skills are in demand, as are introductory and 'getting started' courses.
- Topic-based or thematic contextualised courses are considered to be a good way of approaching training in basic and more advanced data analysis skills.

3. Key requirements and recommendations

There is considerable potential for increasing usage of four kinds of data types datasets across all social science disciplines.

3.1 ESDS Data portfolio

There are some gaps in the data collection currently provided under ESDS.

- There is evidence of an unmet demand for access to international data to undertake comparative research. ESDS International is addressing this shortfall

• ¹ Rice et al (2001), *Report on Use of Data In Teaching and Learning*, Edinburgh University [<http://datalib.ed.ac.uk/projects/datateach.html>]

by negotiating access to some of the most important macro data series that are currently prohibitively expensive for the majority of users to acquire.

- ESDS should help negotiate access to data from some of the key MRC sponsored longitudinal studies, including the ALSPAC and the 1946 Birth Cohort. The UKDA has recently undertaken a consultancy for the MRC as part of their project on establishing a policy on Data Sharing and Preservation, one of the aims of which is to facilitate access to these key high investment longitudinal data sources.

3.2 Resource discovery and data delivery

- The ability to find data efficiently is seen as a priority for international and qualitative data, suggesting a need for improved finding aids. Work in this area should include improving introductory descriptions of data, the provision of FAQs, and easier ways to search and browse for particular kinds of data or data series.
- Browsing and introductory descriptions of the most popular datasets, together with an overview of data by theme are advocated.
- A web-based database of data collections held in other locations would also be welcomed. The universal portal for UK-based social science data planned for ESDS is likely to be viewed as a good investment by these users.
- Data should be organised in a systematic and familiar way, with thorough and clear documentation. Help in navigating through complex documentation would be beneficial, particularly for users of government and longitudinal data series.
- Users prefer instant desk-top access to data, but are not all aware of the services currently offered for UK data. Online access to data, for example via the UKDA Web-based instant download service and NESSTAR should be better publicised.
- Enabling interactive online access to qualitative data, i.e. by browsing and exploring the content of data via a web browser, would be welcomed by users and should be considered as a priority for research and development work for ESDS Qualidata.
- Data should be free, where possible, from restrictive access conditions and ethical constraints and, in the cases where permissions must be sought from data depositors, the administrative process taken to gain authorised access following a data request should be efficient and timely.
- Promotional and awareness raising materials for distribution to known data users and at key data courses/workshops around the country would be beneficial.

3.3 Value-added data and documentation

- Data enhancements focusing on look-up indices of variables that are harmonised or comparable between surveys would be welcomed by users of the quantitative datasets, as would additional derived variables. Online access to sampler datasets for longitudinal data and for qualitative data is a priority.
- Provision should be made for enabling users to import data and coded data into CAQDAS packages. It is recommended that work on a DTD for defining a standard and exchangeable format for data and associated encoding is pursued.

- It is recommended that enhanced data products and support materials for each specialist data service are based on the substantive topics and themes most popularly requested.
- Data samplers are deemed to be less essential for researchers of complex survey data, but useful for training purposes. The creation of digital samplers for qualitative data users are advocated, particularly for scanning larger paper-based 'classic' collections.

3.4 User Support and Training

- Introductory workshops on 'How do I get started?' would be supported by users, to demonstrate searching the ESDS site, registering for data, exploring data online, downloading and carrying out analyses.
- SPSS and STATA are the most widely used statistical analysis packages for survey data users and should be supported in the form of short-courses. Data handling and analysis in SPSS and MS Excel would be beneficial for users of international macro data. Workshops that combine training in the use of CAQDAS packages and qualitative data analysis would be useful, suggesting collaborative events with the ESRC CAQDAS Networking project would be fruitful.
- For government and longitudinal data, courses on the hierarchy of the data, linking records over time and weighting data would be welcomed. Training in key analytic skills should be built in to the programme of training, and would be perhaps best undertaken, from a financial consideration, in collaboration with other training providers.
- Topic-based courses drawing on a range of data based on the key themes would be popular. It is further recommended that online topic based user guides are produced on the most popular themes in order to get people started with using datasets relevant to the topic, including pointing users in the right direction of surveys and documentation.
- Themed training/teaching datasets are also seen as valuable resources and should be produced on the same yearly themes as selected for data enhancement work.
- Online-training materials, such as those from courses, are seen as a priority when considering methods of providing support for data and secondary data analysis. 'How-to guides', case studies and exemplars of re-use would also be well received by users, and by information intermediaries, such as subject librarians.

The creation of online resources, as demanded by users, can be prohibitively expensive, and ESDS should explore how to best foster its relationship with both current initiatives in this area and tutors in HE or FE so as to maximise the use of archived data in mainstream teaching. Current initiatives include: the ESRC Research Methods Programme Phase II with an emphasis on research methods and data analysis training (and also incorporating graduate level training)², led by Angela Dale of CCSR; the JISC 5/99 programme under which an excellent set of teaching and learning materials have been created (with UKDA and CCSR involvement) based on the Contemporary and Historical Census Collection (CHCC)³; and the JISC Exchange for Learning Programme that aims to 're-purpose' existing digital JISC-based resources (including datasets) for

² The ESRC Research Methods Programme Phase II specification can be found at <http://www.ccsr.ac.uk/methods/>

³ CHCC www.chcc.ac.uk

teaching and learning, and under which the UKDA currently has a grant to create a flexible teaching resource based on some of the most popular UKDA data series.

The following sections, 3-6 report a more detailed but mostly descriptive analysis of the survey responses, reported by each data type: government data; international data; longitudinal data; and qualitative data. Only key tables are reported in the body of the reports with further tables found in the Appendix to each section, as indicated in the body of the text. Missing values and Ns for column percentages are not reported in tables.

3.5 Conclusion and next steps

The results of the surveys are much welcomed by the ESDS team. Not only have they helped confirm and consolidate the deliverables already drafted for the services (see www.esds.ac.uk milestones for each service) they have helped pinpoint more precisely priorities for both data enhancement work and for planning support activities and resources. ESDS will consider carefully updating the milestones/deliverables in consideration of some of the more pertinent findings from this consultation.

The establishment of JISCmail lists for each of the specialist data services will be one method of keeping in touch with users and obtaining feedback on development plans. ESDS Qualidata will be hosting a further user consultation over the next six months in order to help plan development work for the provision of online data browsing systems.