

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Not for citation without permission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.

4. ESDS Government: User Survey Results

4.1 Introduction

ESDS Government is the value-added service to support large-scale government surveys, such as the General Household Survey and the Labour Force Survey, which are key data resources for social science researchers. ESDS Government, led by the CCSR will:

- raise awareness of the research potential of the government datasets
- provide user support through a dedicated help line, user groups and FAQs
- exchange information between users and producers
- run training courses on key topics of interest, on specific statistical packages, and on methods of statistical analysis
- provide topic-related online course materials and a range of teaching datasets
- create a number of value-added products providing easy routes through the complex documentation of the government surveys
- work closely with ONS and other data producers to ensure that the complementarity of interest between producers, commissioning government departments and academic users is maximised

The acquisition, processing and delivery of the large-scale government datasets under this service will be undertaken by the UKDA. Data will be made available via the UKDA web-based download system and the Nesstar system for online browsing and visualisation of the data.

4.2 Respondents

One-hundred and ninety-one people participated in the ESDS User Consultation Survey for the large-scale government surveys. More than seven in ten respondents (71%) were from UK Higher Education institutions (see Table 1A in the Appendix)¹. Over a quarter (27%) of respondents were members of Higher Education/Further Education staff and almost equal proportions of respondents were undergraduates, contract researchers and research students (19%, 18% and 17% respectively) (see Table 2A). Respondents within UK Higher Education, UK Further Education and non-UK academic institutions came from a wide range of disciplines; the most frequently recorded discipline was Social Policy (24%) (see Table 3A).

4.3 Current or potential use of data sources

The two most widely used large-scale government surveys were the General Household Survey (40%) and the Labour Force Survey (35%). Other highly used surveys were the Health Survey for England, Family Expenditure Survey and British Crime Survey (Table 1). Sixteen percent of respondents said that they used surveys other than those listed. However, none of the 'other' surveys mentioned were widely used. High proportions of respondents thought that they might use the General Household Survey (66%), Labour Force Survey (52%) and Health Survey for England (52%) in the future.

¹ Note to tables: missing data arising from respondents not giving answers to questions have been exclude from the base numbers shown in tables and the bases used in percentaging. Tables named nA are found in the Appendix to this section, Section 4

Table 1

	Which large-scale government surveys do you usually use?	Which large-scale government surveys might you use in the future?
	%*	%*
None	23	4
General Household Survey	40	66
Labour Force Survey	35	52
Health Survey for England	23	52
Family Expenditure Survey	21	38
British Crime Survey	20	34
Family Resources Survey	13	40
National Food Survey	8	23
Omnibus Survey	8	27
Survey of English Housing	5	26
Other	16	12
N=all respondents	184	184

**Percentages may sum to more than 100 because respondents could give more than one answer*

The status, institutions and disciplines of respondents were examined for each of the five most frequently used surveys. Most of the respondents for each survey were from UK Higher Education Institutions. However, the Family Expenditure Survey had higher proportions of respondents from UK public organisations than any of the other surveys (see Table 4A). The Health Survey for England, General Household Survey, Labour Force Survey and Family Expenditure survey were well used by members of Higher Education/Further Education staff and contract researchers. However, the British Crime Survey was mainly used by undergraduates and contract researchers and was less likely to be used by members of Higher Education/Further Education staff (see Table 5A). There was variation between the disciplines of the respondents who used each of the five surveys. For example those who used the Labour Force Survey were mainly from economics, social policy and management and business studies backgrounds compared with General Household Survey users were from sociology, social policy and political science and international relations backgrounds (see Table 6A).

Over a quarter (27%) of respondents said that they access or order data regularly from the UKDA (or previously MIMAS) and half (50%) do so occasionally. Almost two-thirds (63%) had accessed or ordered data within the last three months and a further 19% had done so more than three months ago but within the last year (see Tables 7A and 8A).

Respondents were asked which data analyses software packages they used on a regular basis. SPSS was the most widely used (78%), followed by STATA (18%). Only small proportions of respondents used SAS, NESSTAR or NSDstat or other packages on a regular basis (see Table 9A)

4.4 Using data in research: barriers and requirements

Sixty-five percent of respondents used at least one of the listed government datasets for research (Table 10A). The topics they were researching are shown in Table 11A.

The majority of the respondents who used the government datasets for research had not experienced any difficulties with accessing the surveys for their research. Only twelve per cent experienced difficulties with registration for data services, 5% with ordering data via the UK Date Archive (UKDA) online system and 18% with downloading data via the UKDA (Table 2). A few respondents (4%) reported other problems with accessing the government datasets. These are listed in Table 12A.

Table 2

Have you experienced any difficulties accessing and using large-scale government survey data in your research?		
<i>Accessing data</i>		
Registration for data services	%	12
Ordering via UKDA online system	%	5
Downloading data via UKDA	%	18
Other	%	4
<i>N=all who use for research (min)</i>		92
<i>Using data</i>		
Conducting statistical analysis	%	16
Data documentation	%	32
Obtaining details about variables	%	41
Merging files	%	19
Weighting	%	24
Handling missing data	%	22
Other	%	2
<i>N=all who use for research (min)</i>		81

The most frequently reported difficulties with using the government surveys for research were with obtaining details about variables (41%) and data documentation (32%). Only two respondents (2%) reported other problems with using the government datasets for research: "the cost of the data and the selective nature of data from agencies" and "reasons for the survey are important as it can affect bias of the outcome".

Respondents were asked to provide information on other difficulties they had experienced with accessing and using large-scale government survey data in their research. A difficulty reported by many respondents was related to the quality and complexity of the data documentation.

Just under a quarter (23%) of respondents thought that the addition of extra derived variables to the government datasets would be very useful for their research (Table 3). The majority of respondents (65%) thought that this would be of some use and a further 12% did not think it would be useful at all. Those who thought that this would be very useful were asked which variables they would like to be added; derived variables related to income were frequently mentioned.

Table 3

Would any of the following be useful in using large-scale government datasets in your research:	
	%
Additional derived variables?	
Not useful	12
Of some use	65
Very useful	23
<i>N=all who use for research</i>	<i>106</i>
A look-up index of variables that are harmonised or comparable between surveys?	
Not useful	3
Of some use	39
Very useful	58
<i>N=all who use for research</i>	<i>109</i>

Almost six in ten (58%) respondents thought that a look-up index of variables that are harmonised or comparable between surveys would be very useful for their research. A further 39% thought that it would be of some use and only 3% thought it would not be any use at all. Those who thought that this would be very useful were asked which topic areas would be most useful. Respondents overwhelmingly reported that they would like an index of variables for income and an index for social capital variables.

Respondents were asked what other services or developments would make it easier for them to access and use the large-scale government surveys. Many respondents expressed a need for access/availability of data and being able to download data. As this service is currently provided, the results suggest that better publicity is required.

4.5 Using data in teaching: barriers and requirements

Twenty-four percent of respondents used government datasets in their teaching. The proportions that reported use of the individual surveys were low. However the most widely used were the General Household Survey (12%), the Health Survey for England (10%) and the Labour Force Survey (8%) (see (Table 13A). Four respondents (3%) reported the use of other government surveys in their teaching including the British Household Panel Survey, the National Diet and Nutrition Survey, the Allied Dunbar National Fitness Survey and the Health of the Nation.

Over half those who used the government datasets in their teaching had used them for postgraduate (56%) or undergraduate (56%) teaching. Only small proportions had used them for teaching in further education (18%) or for professional courses (18%) (Table 14A). The areas in which their teaching was focused are shown in the Table 15A.

Only small proportions of those respondents who used the government surveys in their teaching had experienced any difficulties with accessing the surveys for their teaching (Table 4).

Table 4

Have you experienced any difficulties accessing or using large-scale government survey data in your teaching?	
	%
Accessing data	
Class registration for data services	8
Gaining permission to use data for teaching	15
Ordering via UKDA online system	4
Downloading data via UKDA system	11
Other	-
<i>N=all who use for teaching (min)</i>	26
Using data	
Data documentation	19
Obtaining details about variables	23
Merging files	20
Weighting	24
Teaching statistical analyses with real data	17
Availability of tailor-made teaching datasets	28
Other	-
<i>N=all who use for teaching (min)</i>	56

The difficulties most frequently experienced were gaining permission to use the data for teaching (15%) and downloading the data via the UKDA system (11%). None of the respondents who had used the data for teaching had experienced any other problems accessing it.

With regard to using the surveys for teaching, just under three-in-ten (28%) respondents had experienced problems with the availability of tailor-made teaching datasets, while around a quarter had experienced difficulties with weighting (24%) and obtaining details about variables (23%). The following responses were given when respondents were asked for further details on difficulties experienced:

“Downloaded datasets have not always been easy to decompress for use. Problems have also been caused by the size/disk space needed and the inability to download only selected years of data.”

“Although registration is quite straightforward for myself the registration form is rather intimidating for students wishing to register. A more user-friendly approach for undergraduates would be advisable, in other words, a separate form for undergraduate.”

“We do fairly simple data analysis for undergraduate and postgraduate teaching and don’t really have the expertise for manipulating and constructing teaching data sets.”

None of the respondents who had used the data for teaching had experienced any other problems using it. However one respondent had not used any of the government surveys for teaching because they had experienced problems gaining access for their students. He/she commented

“the current access requirements make using Government Survey Data for teaching almost impossible. The access requirements need to be changed so they are like the 2001 Census which is excellent and we have already successfully used at MSc level for teaching”.

Over one-third (36%) of respondents thought that themed teaching datasets would be very helpful in their teaching and a further 55% thought that they would be a little helpful (Table 5). Only about one in ten (9%) respondents did not think that themed teaching datasets would be helpful at all. Those who thought these would be very useful were asked which themes would be most helpful and the main responses were health, social exclusion and employment. It was also evident that multi-purpose themes were required.

Table 5

Would themed teaching datasets be helpful in your teaching?	
	%
Not at all helpful	9
A little helpful	55
Yes, very helpful	36
<i>N=all who use for teaching</i>	33

4.6 Online training and support requirements

Table 6 sets out respondents' views on which forms of training they might find useful. Thirty percent of respondents said that they would find courses to raise awareness of the datasets and their research potential very useful. A further 56% said they would find these of some use. Those who thought they would be very useful were interested in health, employment and basic research methods.

Table 6

Which of the following forms of training might you find useful?	
Courses to raise awareness of large-scale government datasets and their research potential	
	%
Not useful	14
Of some use	56
Very useful	30
<i>All respondents</i>	162
Courses on specific aspects of particular surveys	
Not useful	23
Of some use	53
Very useful	25
<i>All respondents</i>	159
Topic-based courses drawing on a range of surveys	
Not useful	11
Of some use	56
Very useful	32
<i>All respondents</i>	154
Online training and support resources	
Not useful	10
Of some use	49
Very useful	41
<i>All respondents</i>	160

A quarter of respondents said that they would find courses on specific aspects of particular surveys very useful. A further 53% said they would find these of some use. Those who thought they would be very useful were particularly interested in the linkage of data over time for the LFS and the hierarchy of the GHS.

Thirty-two percent of respondents said that they would find topic-based courses drawing on a range of surveys very useful. A further 56% said they would find these of some use. Those who thought they would be very useful were particularly interested in employment and the labour market.

Forty-one percent of respondents said that they would find online training and support resources very useful. A further 49% said they would find these of some use. Those who thought they would be very useful were interested in 'how-to' guides, in particular guides that would help them get started with their research.

Respondents were asked to add any other comments about the government surveys. The majority of responses were positive; praising the current work of the UK Data Archive.

4.7 Observations and recommendations

The results of the user survey suggest that many of the large-scale government surveys are widely used and will continue to be widely used in the future. The surveys are used more for research than for teaching. The services currently provided by the Data Archive were praised by users and the following recommendations will build upon what is currently available:

- Employment and the labour market, health and social exclusion were, by far, the most frequently mentioned topics throughout the survey. It is recommended that these themes are used in turn for the first three years of the ESDS large-scale government surveys:
 - 1st year: employment and the labour market
 - 2nd year: health
 - 3rd year: social exclusion
- A look-up index of variables that are harmonised or comparable between surveys would be welcomed by users of the datasets. This should take priority over the addition of extra derived variables to datasets. However, these were also welcomed by users. The first year should concentrate on employment/income variables.
- Themed teaching datasets were popular and should be produced on the same yearly themes as shown above. In addition a teaching dataset on the British Crime Survey should be produced because this is a widely used survey.
- Courses/workshops:
 - Workshops to raise awareness of the datasets and their research potential on the same yearly themes as above.
 - Workshops on 'How do I get started?' to demonstrate searching the ESDS site, exploring with Nesstar, registering for data, downloading and carrying out analyses.
 - Topic-based courses drawing on a range of surveys on the same yearly themes as above.
 - Courses on the hierarchy of the GHS and linking records over time in the LFS would be welcomed by users of the datasets.
 - SPSS is the most widely used analysis package and should be supported in the form of short-courses. Short-courses on introductions to STATA are also recommended.

- Online-training and support was particularly popular, with a particular need for 'how-to guides'. It is recommended that on-line topic based user guides are produced on the same yearly themes as above in order to get people started with using datasets relevant to the topic, including pointing users in the right direction of surveys and documentation. It is also recommended that course materials are made available on-line.
- Respondents expressed a need for the availability of data and the ability to download data. On-line access to data is a fairly new service and ESDS should ensure that it is publicised.

4.8 Appendix: ESDS Government User Survey Tables

Table 1A

Affiliation/Institution of respondents	
	%
UK Higher Education	71
UK Public Organisation	15
UK Further Education	8
UK Private Organisation	1
Non-UK	4
<i>N=all respondents</i>	<i>182</i>

Table 2A

Status of respondents	
	%
Member of HE/FE staff	27
Undergraduate	19
Contract researcher	18
Research student	17
Other student	4
Visiting academic	3
Other	12
<i>N=all respondents</i>	<i>187</i>

Table 3A

Discipline of respondents	
	%
Economics	9
Sociology	9
Social Policy	24
Psychology	2
Political Science & International Relations	7
Education	3
Human Geography	6
Environmental Planning	1
Economic and Social History	2
Management and Business Studies	8
Socio-legal Studies	1
Area Studies	1
Statistics and Computing	5
Librarian	2
Other	21
<i>N=all in UK FE, UK HE or non-UK academics</i>	<i>150</i>

Table 4A

Affiliation/Institution of respondents by survey					
	LFS	GHS	HSE	FES	BCS
	%	%	%	%	%
Non-UK	6	4	2	8	6
UK Further Education	2	4	5	3	6
UK Higher Education	69	74	70	58	74
UK Private organisation	2	3	-	3	-
UK Public organisation	22	15	22	28	14
<i>N=all who currently use the survey</i>	64	73	40	36	35

Table 5A

Status of respondents by survey					
	LFS	GHS	HSE	FES	BCS
	%	%	%	%	%
Member of HE/FE staff	31	34	38	26	19
Undergraduate	17	15	14	21	36
Contract researcher	28	23	21	24	22
Research student	5	7	10	8	8
Other student	1	7	2	3	-
Visiting academic	3	1	2	3	3
Other	15	12	12	16	11
<i>N=all who currently use the survey</i>	65	73	42	38	36

Table 6A

Discipline of respondents by survey					
	LFS	GHS	HSE	FES	BCS
	%	%	%	%	%
Economics	13	4	6	8	3
Sociology	2	14	16	8	7
Social Policy	19	28	35	27	27
Psychology	2	4	3	-	3
Political Science & International Relations	9	11	-	4	13
Education	2	-	-	-	-
Human Geography	6	7	10	8	13
Environmental Planning	-	2	3	-	-
Economic and Social History	-	2	-	-	3
Management and Business Studies	15	5	-	12	-
Socio-legal Studies	-	-	-	-	3
Area Studies	2	-	-	4	3
Statistics and Computing	6	7	6	8	10
Librarian	-	2	3	-	3
Other	23	16	16	23	10
<i>N=all in UK FE, UK HE or non-UK academics</i>	47	57	31	26	30

Table 7A

How often do you access/order government data from the UK Data Archive, or the previous MIMAS service?	
	%
Rarely/never	22
Occasionally	50
Regularly	27
<i>N=all who currently use dataset</i>	143

Table 8A

When did you last access/order government data from the UK Data Archive, or the previous MIMAS service?	
	%
More than 3 years ago	7
More than 1 year ago, but no more than 3 years ago	11
More than 3 months, but no more than 1 year ago	19
In the last 3 months	63
<i>N=all who currently use dataset</i>	122

Table 9A

Which data analysis software packages do you use on a regular basis?	
	%*
SPSS	78
STATA	18
SAS	8
NESSTAR	7
NSDstat	2
Other	13
<i>N=all respondents</i>	153

*Percentages may sum to more than 100 because respondents could give more than one answer

Table 10A

Do you use any of the government datasets listed in Qq2* for research?	
	%
Yes	65
No	35
<i>N=all respondents</i>	182

LFS, GHS, HSE, FES, FRS, Omnibus, SEH, NFS, BCS.

Table 11A

Research topic	% *
Health	35
Poverty (social exclusion; social capital)	25
Employment and labour market	23
Crime	14
Income	10
Ethnicity	9
Education	9
Youth/children	8
Disability	8
Demographics/households/families	6
Social care/carers	4
Gender	4
Leisure/cultural activities	3
Elderly	4
Other	21
<i>N=all who use for research</i>	<i>113</i>

**Percentages may sum to more than 100 because respondents could give more than one answer*

Table 12A

Other problems with accessing data for research?
Finding data archive material
Knowing where and how to obtain it
Sometimes asked to pay
The closure of online access at MIMAS has caused enormous problems
Used govt sources- it was easier

Table 13A

Do you use any of the following datasets in your teaching?	
	%*
None	76
Labour Force Survey	8
General Household Survey	12
Health Survey for England	10
Family Expenditure Survey	5
Family Resources Survey	1
Omnibus Survey	2
Survey of English Housing	2
National Food Survey	4
British Crime Survey	4
Other	3
<i>N=all respondents</i>	<i>142</i>

**Percentages may sum to more than 100 because respondents could give more than one answer*

Table 14A

In which context have you used the datasets in your teaching?	
	%
Postgraduate	56
Undergraduate	56
Further education	18
Professional courses	18
<i>N=all who use for teaching</i>	34

**Percentages may sum to more than 100 because respondents could give more than one answer*

Table 15A

Area in which teaching is focused	%
Research methods/statistics	39
Health	16
Social studies & related	13
Crime	6
Employment & labour market	6
Other	26
<i>N=all who use for teaching</i>	31

**Percentages may sum to more than 100 because respondents could give more than one answer*

5. ESDS International: User Survey Results

5.1 Response rates

A total of 52 responses were received. Of these, 33 (63%) respondents were from the UK with the remaining 19 (37%) from outside of the UK. Of the 33 UK respondents, 29 (56%) were from the UK academic sector – including 1 respondent from the FE sector. Of those from the UK academic sector, 46% were staff, 29% were undergraduates, with the remainder (25%) being either research students or contract researchers.

The respondents were classified by academic discipline. Excluding librarians and others, 10 different social science disciplines were represented (see table 1A¹). The disciplines with the greatest representation were Economics (15%), Social Policy and Social Admin (15%), Political Science and International Relations (15%), Sociology (8%), Education (8%) and Management and Business Studies (8%). For UK academic only respondents, the disciplines with the greatest representation included Social Policy and Social Admin (17%), Political Science & International Relations (10%) and Economics (10%). Three of the UK academic respondents were librarians highlighting the role that they play in helping academic users to identify and access key data and information resources for research and teaching purposes

5.2 Current or potential use of data sources

A total of 30 (58%) respondents reported using international datasets. The most frequently used datasets included OECD MEI (25%), Eurobarometer (17%), IMF (15%), ISSP (13%), Eurostat New Cronos (12%) and NS Time Series Databank (10%) (see table 1).

Table 1

	Which international datasets do you currently use?	Which international datasets might you use in the future?
	%*	%*
None	42	6
OCED MEI	25	44
IMF Databanks	15	37
UNIDO Industrial Statistics	2	19
ONS Time Series Data	10	25
Eurostat New Cronos	12	42
Eurobarometers	17	42
ISSP	13	52
WEVS	8	52
N=all respondents	52	52

* Percentages may sum to more than 100 because respondents could give more than one answer

The relatively low level of usage reported for the NS Databank may reflect the large number of non-UK respondents in the survey. Table 2A shows current data use by domain. Datasets which were used more frequently by UK academic respondents

¹ Tables named nA are found in section 5.7

included Eurostat New Cronos and Eurobarometer. In terms of UK academic respondents, 45% of all active data users were from Social Policy and Political Science & International Relations. Users from these two discipline areas reported using both macro and micro international datasets.

Datasets from the World Bank, the United Nations, and the ILO were also cited as being used. In addition, two respondents reported use of data relating to the specific topics of tourism and electoral performance. One respondent reported use of JSTOR and Science Direct. (See Appendix 2: q1qa_10 for full textual responses²)

Table 3A provides a breakdown of data use by discipline. Out of the 30 respondents using international datasets, the biggest users were from Political Science (23%), Social Policy (20%), Economics (13%) and Management/Business Studies (10%). Whilst these four disciplines accounted for 66% of all data users, data use is widespread across a range of social science disciplines. For UK academic respondents, the biggest users were from Social Policy (28%), Political Science (17%), Librarians (17%) and Economics (11%).

Out of 48 valid responses, 28 respondents (58%) reported using international data in research. Of these, 14 (50%) were UK academic respondents. The full list of the research topics cited is provided in appendix 2: q6. This list highlights the extent to which international perspectives are being developed for a wide range of research topics and areas. The list of research topics provided by respondents has been classified using the subject categories used by CESSDA (see table 9A). The subject categories receiving the highest number of references were Social Stratification and Groupings (21%); Labour and Employment (15%); Society and Culture (11%); Politics (11%); and Economics (11%).

Out of the 52 respondents, 20 (38%) reported using international data in teaching. Of these, 10 (50%) were UK academic respondents. Use of international data in teaching is spread across a range of social science disciplines (table 6A) although reported use of individual datasets was low. The disciplines recording greatest use of data in teaching were Economics (20%) and Social Policy (15%). In terms of UK academic respondents, data use is also spread across a range of disciplines. Interestingly, 20% of UK academic users of data for teaching purposes were librarians.

In terms of the datasets used in teaching (table 7A) the most popular were ISSP (15%) and OECD (15%). For UK academic respondents, a similar pattern of usage is reported. No respondents reported using the NS Time Series Databank in teaching which is surprising since it is currently used for teaching in a number of institutions. Other datasets reported being used in teaching included: Labour Statistics and DHS data (see Appendix 2: 10_10).

Five respondents (50%) reported using data in a postgraduate context with the remainder reporting use at undergraduate (40%) or FE (10%) level. The subject areas in which teaching was focused included: Social Policy, Politics, Economics, Education, Sociology and Management and Business Studies. One respondent reported using international data to teach data analysis to research students. The full list of the substantive teaching areas is provided in appendix 2: q12.

² Full textual responses are provided in section 5.8

Only 13% of respondents reported accessing/ordering data regularly from the UKDA, MIMAS or r.cade with 37% doing so only occasionally (see table 4A). For UK academic users, the numbers reporting accessing/ordering data on a regular or occasional basis are 12% and 53% respectively. However, in terms of when data was last accessed/ordered from the UKDA, MIMAS or r.cade, 62% of all respondents and 64% of UK academics reported accessing/ordering data in the last three months.

Whilst only 30 (58%) respondents reported that they were currently using international data, 49 (94%) felt that they would use international data in the future (see table 1). In rank order, the datasets with the greatest potential future demand included ISSP (52%), WEVS (52%), OECD MEI (44%), Eurostat New Cronos (42%), Eurobarometer (42%), IMF (37%), NS Time Series (25%) and UNIDO (19%). As can be seen from table 5A, UK academic respondents are more likely to make future use of Eurostat New Cronos, Eurobarometer and ONS datasets than other respondents. However, UK academic respondents were less likely to use the UNIDO data than other respondents. The responses to the future use question clearly indicate that there is considerable scope for increased usage of both macro and micro international datasets.

Other datasets cited as being of potential use in the future included the EC Household Panel (ECHP); Regio; EU Statistics on Income and Living Conditions (EU-SILC); World Bank data; Luxembourg Income Study; UN Public Finance Data and "anything relating to travel and tourism." See Appendix 2: q2_10.

Respondents were also asked to specify whether there are any data on particular countries and/or topics which they would like to use but were not sure where to find them. A total of 25 (48%) respondents answered 'yes' to this question. A complete list of all the responses to this question are provided in Appendix 2: q3_2.

Four respondents had non-specific data requests for named countries or areas (e.g. Egypt, China and the former soviet republics). Eleven respondents requested access to datasets relating to specific topics (e.g. international debt, labour market trends, education, migration and crime) without being specific about countries/regions. Six respondents had requirements for data relating to specific topics for named countries/regions (e.g. panel income data for Sweden). Three respondents required access to particular types of data (e.g. opinion polls and panel studies) for named countries or regions (e.g. developing countries and Scandinavia) without being specific about topics.

The specified topics for which data was required have been classified using the CESSDA subject categories (see table 9A). The most cited subject areas for which data was required were Economics; Labour and Employment; Politics and Law, Crime and Legal Systems. The responses indicate that there is a significant demand for access to high quality/accurate data to undertake comparative research across a range of research topics and areas.

Question 4 sought to identify what software packages were being used to process or analyse international datasets. The most popular packages included SPSS (60%), MS Excel (58%), MS Access (17%), Stata (17%) and SAS (13%) (see table 8A). Very low usage was reported for other general purpose or specialist data analysis packages. Other reported packages used included Minitab and Systat (see Appendix 2: q4_15).

5.3 Using data in research: barriers and requirements

Question 7 sought to identify problems in accessing or using international datasets in research (see table 2).

Table 2

Have you experienced any difficulties in accessing or using international datasets in your research?				
		Non-academic /other	UK academic	Total
<i>Accessing data</i>				
Identifying appropriate datasets	%	64	79	71
Registration for data services	%	29	36	32
Extracting variables or observations	%	50	36	43
Ordering via UKDA online system	%	21	14	18
Downloading data	%	43	21	32
Cost of acquiring data	%	36	57	46
<i>N= all who use data in research</i>		14	14	28
<i>Using data</i>				
Problems in conducting statistical analysis	%	21	29	25
Data documentation	%	43	50	46
Obtaining details about variables	%	71	29	50
Merging files	%	29	29	29
Weighting	%	21	36	29
Handling missing data	%	29	29	29
<i>N= all who use data in research</i>		14	14	28

In terms of data access, the biggest difficulty encountered by all respondents was identifying appropriate datasets. Twenty respondents (71%) reported this as a problem. Other major problem areas included costs of acquiring data (46%) and extracting variables (43%). Registration was still a problem for 32% of respondents but ordering data via the UKDA online system was only a problem for 18% of respondents. However, occasional or regular data users were less likely to experience data access problems, such as registration and data ordering, than irregular users. The exception to this was data acquisition costs.

Table 2 also provides a breakdown of access problems by domain. This table highlights the extent to which these data access problems affect UK academic respondents more or less than other respondents. For UK academic respondents, identifying appropriate datasets (79%) and data acquisition costs (57%) are more of a problem than for other respondents. Whilst 36% of UK academic users encountered registration problems only 14% had problems ordering the data from the UKDA and 21% encountered problems with downloading data. Although less of a problem for UK academic respondents, the relatively high numbers may reflect the existence of multiple registration systems for different service providers and multiple interfaces prior to the establishment of ESDS.

The high numbers of UK academic users reporting data acquisition costs as a major problem is not surprising. Whilst some international datasets are available at

zero/low cost to authorised users other datasets are expensive to acquire. However, even if a dataset is expensive to acquire it still may be free to the end-user if made available under a site licence agreement.

In terms of data use (table 2), the biggest reported problems were difficulties in obtaining information about variables (50%) and data documentation (46%). Whilst causing problems for almost 30% of all respondents, the majority did not experience any difficulties in conducting statistical analysis, merging files, weighting or handling missing data. Table 2 also provides a breakdown of usage problems by domain. Problems relating to data documentation, weighting and statistical analysis data were more of a problem for UK academic respondents. Conversely, obtaining variable details was less of a problem.

The following responses (see Appendix 2: q7_14) were given when respondents were asked to provide further details about problems encountered:

“Cross-country definitions are sometimes not clear in the documentation.”

“Difficult to find out what information was available in ISSP or on numbers of respondents until I actually received the data and documentation. Still can't find info on weighting, on what variables are missing for what countries. Income data is a mess (1997) with no clarity as to period covered, units, or currency conversion factors for that year. Also no continuous income variables. Had to abandon any analyses using income.”

Question 8 asked for feedback about proposed services or developments that would be useful to researchers (see table 3). In general terms over 90% of respondents would find these 'of some use' or 'very useful'.

Table 3

Would any of the following be useful in using international datasets in your research:			
	Non-academic /other (%)	UK academic (%)	Total (%)
Guide to comparability across datasets			
Not useful	12	-	6
Of some use	29	33	31
Very useful	59	67	62
N=all who use for research	17	15	32
Alternative systems of classification			
Not useful	12	-	7
Of some use	41	54	47
Very useful	47	46	47
N=all who use for research	17	13	30
Searching facility across datasets			
Not useful	-	13	7
Of some use	40	27	33
Very useful	60	60	60
N=all who use for research	17	13	30

The topic areas which respondents suggested should be covered in guides to comparability included: economic indicators; employment and the labour market; crime; health and aging; and public expenditure by function. Some respondents requested guides covering topics such as linking variables, weighting, and how to compare different data sources. See Appendix 2: q8a_2

Areas where respondents felt that alternative classifications would be of greatest use included: employment and the labour market; health and aging. A couple of respondents requested the use of the EGP classification of occupation and the ESA 95 system for Input /Output Tables. See Appendix 2: q8b_2.

In terms of a search facility, a number of respondents wanted to search across datasets for specific topics such as employment and the labour market; trade and investment; demographics; socio-economic variables; health and aging and crime. See Appendix 2: q8c_2.

Other services or developments that were cited as making it easier for respondents to access and use international datasets in research included: data specific workshops; access to data experts; access to research papers; charts of data for countries; multilingual search facility; advice about the relation between national and international datasets; and freely available historical statistics. See Appendix 2: q9.

5.4 Using data in teaching: barriers and requirements

Question 13 sought to identify problems in accessing or using international datasets in teaching (see table 4). In general terms, 20% to 25% of respondents using international data in teaching reported problems in accessing or using data. No particular problems emerged as being significantly more or less important than others.

Table 4

Have you experienced any difficulties in accessing or using international datasets in your teaching?		
		Total
Accessing data		
Identifying appropriate datasets	%	25
Class registration for data services	%	25
Extracting variables or observations	%	20
Ordering via UKDA online system	%	20
Cost of acquiring data	%	20
<i>N= all who use data in research</i>		20
Using data		
Teaching statistical analysis with real data	%	20
Availability of tailor-made teaching datasets	%	20
Data documentation	%	25
Obtaining details about variables	%	25
Merging files	%	20
Weighting	%	10
<i>N= all who use data in research</i>		20

There was some support for themed teaching datasets. Half of the respondents using international data in teaching would find themed teaching datasets of some help or very helpful. Themes that were requested included: poverty, health, employment, social change, political behaviour, lifestyle and leisure. See Appendix 2: q14_2.

The following response was given when respondents were asked to suggest other services or developments that would make it easier to use international data in teaching:

“It would be really good to have country-level data packaged with individual level comparative data, e.g. key economic and social indicators for the years and countries of the ISSP data.”

5.5 Online training and support requirements

Question 16 asked for feedback about the forms of training that users would find useful (see table 5). In general terms, there was strong support for all the proposed types of training courses and events. For UK academic respondents, the modal class for all types of training courses was ‘of some use’. The most popular training courses required by UK academic users – as measured by the number of respondents specifying ‘very useful’ - were themed courses (52%); courses on using packages to analyse international data (35%); and awareness raising courses on the datasets (30%). There was slightly less support amongst UK academic s for a one-day conference (26%) and courses on advanced statistical techniques (13%).

Table 5

Would any of the following forms of training might you find useful?			
	Non-academic /other (%)	UK academic (%)	Total (%)
Courses to raise awareness of international databanks and their research potential			
Not useful	5	13	9
Of some use	67	56	61
Very useful	29	30	29
N=all who use for research	21	23	44
Introductory and advanced courses using packages in analysing international data			
Not useful	29	13	20
Of some use	38	52	45
Very useful	33	35	34
N=all who use for research	21	23	44
Courses on techniques in analysing international data			
Not useful	14	17	16
Of some use	43	70	57
Very useful	43	13	27
N=all who use for research	21	23	44

Courses organised around a theme which focus on availability of variables and identifying, extracting and describing variables			
Not useful	9	-	4
Of some use	48	48	48
Very useful	43	52	48
N=all who use for research	21	23	44
One day conference on the use of international data in research and teaching			
Not useful	19	22	20
Of some use	52	52	52
Very useful	29	26	27
N=all who use for research	21	23	44

The topic areas that respondents suggested that should be covered in awareness raising courses included: data on public expenditure; basic economic indicators; demographic data; social stratification; family life and ageing. See Appendix 2: q16a_2.

Specific statistical techniques to be covered in advanced courses suggested by respondents included: regression, bootstrapping, proportional hazards and multilevel analysis. See Appendix 2: q16c_2.

Specific themes to be covered in the theme based courses suggested by respondents included: health, ageing, poverty, gender, work history, welfare states, employment and the labour market, family life, political behaviour, social policy issues, and social capital. See Appendix 2: q16d_2.

Question 17 asked for feedback about forms of support that users would find useful (see table 6). In general terms there was strong support for all the proposed support services. For UK academic respondents, the modal class for all types of support services was 'very useful' – with the exception of web materials highlighting key international news stories. The most popular support services required by UK academic users – as measured by the number of respondents specifying 'very useful' – were access to on-line or web-based support resources (70%); introductory guides to the various datasets (56%); advanced user guides (52%) and themed guides (43%). There was less support for web-based international news stores illustrated using international data (33%).

Table 6

Which of the following forms of support would you find useful?			
	Non-academic /other (%)	UK academic (%)	Total (%)
Introductory guides on the structure and content of international datasets			
Not useful	9	4	7
Of some use	33	39	36
Very useful	57	56	57
N=all who use for research	21	23	44
Advanced user guides			
Not useful	9	13	11
Of some use	24	35	29
Very useful	67	52	59
N=all who use for research	21	23	44
Modular themed guides			
Not useful	15	17	16
Of some use	35	39	37
Very useful	50	43	46
N=all who use for research	21	23	44
Online or web-based support resources			
Not useful	5	13	9
Of some use	19	17	18
Very useful	76	70	73
N=all who use for research	21	23	44
Web materials highlighting key international news stores			
Not useful	9	29	20
Of some use	33	37	36
Very useful	57	33	44
N=all who use for research	21	23	44

5.6 Observations and recommendations

The following key findings have emerged from the international data survey.

Usage of international data (macro and micro) in research and teaching is widespread across all social science disciplines. Most use is for research purposes but there is evidence that international data is also used in undergraduate and postgraduate teaching.

International data – macro and micro - is used for a wide range of research topics and areas. This diversity of use should be reflected in the training and support materials developed by the IDS.

There is considerable potential for promoting increased usage of international datasets (macro and micro) across all social science disciplines. A key priority for the IDS is to improve awareness of international datasets and how they can be used in research and teaching across a range of social science disciplines.

There is considerable demand for data on particular countries and/or topics to support research across a range of areas. A key task for the IDS is to develop a responsive data acquisitions strategy that seeks to enhance the current portfolio to enable comparative research to be carried out across a range of topics and areas.

Respondents reported experiencing a range of problems accessing and using international data in research and teaching. For UK academic users, difficulties in identifying appropriate datasets, the cost of acquiring data and problems with data documentation were cited as particularly significant barriers to use of international datasets in research. Other reported problems included data registration and extracting variables. The establishment of the one-stop registration system for ESDS and the negotiation of data redistribution licences for key datasets, such as the IMF databanks, should help to address problems relating to registration and data acquisition costs. In terms of the other problems, a key priority for the IDS is to develop a searchable guide to international data sources; improve access to data documentation and provide web-based access to macro datasets via Beyond 20/20.

There was support for value added services to support research use of international datasets – particularly a guide to comparability and a dataset search facility. These guides to comparability should address particular themes as well as covering generic issues, such as international classification systems. There is also a need to make it easier for users to locate variables relating to specific topics. This will be addressed through ESDS metadata initiatives as well as the provision of a search facility within Beyond 20/20.

SPSS and MS Excel were the most popular packages for processing and analysing international data followed by Access, Stata and SAS. There was very little reported use of more specialist time series analysis packages. The development of training courses and support materials should reflect the most popular packages. Further consultation may be required before developing courses and materials for more specialist packages.

There was very strong support from UK academics for the proposed range of support services – especially web-based support resources; introductory guides to the various datasets and advanced user guides. There was slightly less support for web-based international news stories. A key priority for the IDS is to develop web-based introductory support materials relating to the initial portfolio of international datasets

There was strong support from UK academics for the proposed range of training courses – particularly themed courses; awareness raising courses on international datasets and courses on using packages to analyse international data. There was less support for courses on advanced statistical techniques and an annual conference. The key priority for the IDS is to develop introductory courses on international datasets (macro and micro) and data analysis. In terms of addressing reported data access/use problems, these courses should have strong practical component and demonstrate how international datasets – macro and micro - can be used to address substantive research questions across a range of themes. Feedback

about the provision of support services indicates that these courses should be made available via the web.

There was some support for themed teaching datasets but more consultation will be required to identify the required themes, data types (macro and/or micro) and software supply formats.

Librarians appear to play a key role in assisting academic users to identify and access data for research and teaching purposes and should be targeted with information about international datasets.

5.6 Appendix: ESDS International User Survey Tables

Table 1A

Discipline of respondents		
	Count	Col %
Economics	8	15.4%
Sociology	4	7.7%
Social policy	8	15.4%
PolSci & Int Rel	8	15.4%
Education	4	7.7%
Human geography	1	1.9%
Econ & social history	1	1.9%
Management	4	7.7%
Area studies	1	1.9%
Stats & computing	2	3.8%
Librarian	3	5.8%
Other	8	15.4%
N= all respondents	52	

Table 2A Current data use by domain

	Status			
	Non-academic/Other		UK academic	
	Use currently?		Use currently?	
	Yes		Yes	
	Count	Col %*	Count	Col %*
OCED MEI	5	22%	8	28%
IMF Databanks	3	13%	5	17%
UNIDO Industrial Statistics	1	4%	-	-
ONS Time Series Data	2	9%	3	10%
Eurostat New Cronos	1	4%	5	17%
Eurobarometers	3	13%	6	21%
ISSP	3	13%	4	14%
WEVS	2	9%	2	7%
N=all respondents	23		29	

* Percentages may sum to more than 100 because respondents could give more than one answer

Table 3A Current data use by discipline

	Count	Col %
Economics	4	13.3%
Sociology	2	6.7%
Social policy	6	20.0%
PolSci & Int Rel	7	23.3%
Education	2	6.7%
Human geography	1	3.3%
Econ & social history	-	-
Management	3	10.0%
Area studies	-	-
Stats & computing	-	-
Librarian	3	10.0%
Other	2	6.7%
N=all current dataset users	30	

Table 4A Frequency of use by domain

	Status				Total	
	Non-academic/Other		UK academic			
	Count	Col %	Count	Col %	Count	Col %
How often do you access these data products?						
Rarely	9	69.2%	6	35.3%	15	50.0%
Occasionally	2	15.4%	9	52.9%	11	36.7%
Regularly	2	15.4%	2	11.8%	4	13.3%
N=all current dataset users	13		17		30	
When did you last access/order these data products?						
More than 3 years ago			1	9.1%	1	6.3%
More than 1 year ago, but no more than 3	1	20.0%			1	6.3%
More than 3 months, but no more than 1 yr	1	20.0%	3	27.3%	4	25.0%
In the last 3 months	3	60.0%	7	63.6%	10	62.5%
N=all current dataset users	5		11		16	

Table 5A Future data use by dataset by domain

	Status			
	Non-academic/Other		UK academic	
	Count	Col %*	Count	Col %*
OCED MEI	11	47.8%	12	46.2%
IMF Databanks	9	39.1%	10	38.5%
UNIDO Industrial Statistics	6	26.1%	4	15.4%
ONS Time Series Data	5	21.7%	8	30.8%
Eurostat New Cronos	8	34.8%	14	53.8%
Eurobarometers	9	39.1%	13	50.0%
ISSP	13	56.5%	14	53.8%
WEVS	12	52.2%	15	57.7%
<i>N= all respondents</i>	23		26	

* Percentages may sum to more than 100 because respondents could give more than one answer

Table 6A Data use in teaching by discipline

	Count	Col %
Economics	4	20.0%
Sociology	1	5.0%
Social policy	3	15.0%
PolSci & Int Rel	2	10.0%
Education	1	5.0%
Human geography	1	5.0%
Econ & social history		
Management	2	10.0%
Area studies		
Stats & computing		
Librarian	2	10.0%
Other	4	20.0%

Table 7A Data use in teaching by dataset

	Count	Col %
OCED MEI	3	15.0%
IMF Databanks	2	10.0%
UNIDO Industrial Statistics	-	-
ONS Time Series Data	-	-
Eurostat New Cronos	2	10.0%
Eurobarometers	1	5.0%
ISSP	3	15.0%
WEVS	2	10.0%
N= all who use for teaching	20	

Table 8A Software packages used by domain

	Status				Total	
	Non-academic/Other		UK academic			
	Count	Col %*	Count	Col %*	Count	Col %*
EViews	1	4.3%	-	-	1	1.9%
GAUSS	-	-	1	3.4%	1	1.9%
MATLAB	-	-	2	6.9%	2	3.8%
MS Access	3	13.0%	6	20.7%	9	17.3%
MS Excel	13	56.5%	17	58.6%	30	57.7%
NSDstat			2	6.9%	2	3.8%
NESSTAR	1	4.3%	2	6.9%	3	5.8%
Ox	-	-	-	-	-	-
PCGive	-	-	-	-	-	-
RATS	1	4.3%	-	-	1	1.9%
SAS	3	13.0%	4	13.8%	7	13.5%
SPlus/R	2	8.7%	1	3.4%	3	5.8%
SPSS	13	56.5%	18	62.1%	31	59.6%
Stata	5	21.7%	4	13.8%	9	17.3%
N=all respondents	23		29		52	

* Percentages may sum to more than 100 because respondents could give more than one answer

Table 9A Subject classification of research topics and data needs

Subject Categories	Research use (%)	Data needs (%)
Economics	11	22
Trade, Industry & Markets	6	
Labour and Employment	15	17
Politics	11	13
Law, Crime & Legal Systems	4	13
Education	4	4
Information & Communication		
Health	2	4
Natural Environment		4
Housing & Land use Planning	2	
Transport, Travel & Mobility	2	
Social Stratification & Groupings	21	
Society & Culture	11	9
Demography	4	4
Social Welfare Policy & Systems	8	9
Science & Technology		
Psychology		
History		
Reference and Instructional Resources		
N=all referenced topics	53	23

5.8 Appendix 2: Textual Response to Questions

q1a_10 Which other International datasets do you currently use.

Liberal Democratic Performance
WDI, WB data, UN statistics, ILO,
WHO violence against women study
Jstor, Science Direct,
Information pertaining to tourism
World Bank Economic Indicators

q2_10 Which other International datasets might you use in the future.

European household panel survey
REGIO, ECHP, EU-SILC
Luxembourg income study
United nations Public Finance data
Household Labour Survey
Anything pertaining to travel and tourism
World Bank

q3_2 Are there any data on particular countries and/or topics you would ideally like to use but are not sure where to find?

Access to developing country panels.
training and employment, crime rates ...
Central and Eastern Europe - not all countries have accurate data
Dementia and Alzheimer s disease; cohort studies of the elderly and their patterns of institutionalization
There is a need to negotiate the purchase of both the ECHP and EU-SILC for the UK academic community along with the Eurostat harmonised Household Budget Survey and Labour Force Surveys if possible. The results of these surveys are being used as the basis for European policy and there is very little UK Social Science analysis of these very important data
Egypt
International debt
Interpersonal Violence data - especially European
Comparision of school qualifications what is offered (in a comparison table) and what students in different countries actually got.
former Soviet republics
time use
nationalism
environment, human rights
labour market trends
I am interested in crime rates of crimes involving guns and how they relate to the lack of gun ownership by law-abiding Britons and the government's inability to control illegal gun ownership by thugs.
China
Data on attitudes toward democratization, political and economic reforms in Central and Eastern European countries.
childcare provision in european union to compare individual member states
Swedish panel income data; more data on the Scandinavian states in general that has been translated into English. Also seems like there are few of the major German

dataset with English codebooks. Any Western, Industrialized country's national opinion surveys translated into English would be VERY helpful.

rural development

British outbound tourism information regarding Central America

UK - how many people are employed in each sector

Foreign Direct Investment, International Migration

data on political instability

q4_15 Which software packages do you use on a regular basis for processing or analysing data?

Atlas TI

Minitab, Systat

Amelia

Adobe, Word, Outlook

q6 What topics are you researching?

intergenerational poverty

population behaviour

Elections and Democratization

tenure, wage gaps, training, employment flows

poverty, social exclusion, (un)-employment, civil society, urbanisation, gender,

development, social policy and social welfare

Aging, dementia, institutionalization predictors

Income dynamics, poverty

trade

retirement, class, gender, marital inequalities and differences among welfare regimes

public expenditure by function

violence against women and children

impact of European integration on the UK and France - the euro, enlargement etc

Organizational Socialisation, education in FE, Inclusive Education

I help researchers find information, so it can be a wide variety of topics

time use, work-life balance, exercise, working conditions

bilateral trade flows

cross-national comparisons in attitudes to a range of social issues

Effect of gun control on crimes involving weapons

organization, employment, ethnicity

Living conditions of older people

democratization, political attitudes

employment and the labour market

childcare provision

Welfare state provision, especially as it relates to gender and family. Also looking at impact of political makeup on policy outcomes,

Gender Discrimination and Gap in Pay

Travel and tourism

working time and community involvement

appeal of parties to their traditional constituency

trade

inequality, foreign direct investment, economic growth,

q7_13 Have you experienced any difficulties using international datasets in your research?

we have computer savvy people on staff who help with this sort of thing

q7_14 Please provide further details on any difficulties accessing or using data
Cross-country definitions are sometimes not clear in the documentation.
Difficult to find out what information was available in ISSP or on numbers of respondents until I actually received the data and documentation. Still can't find info on weighting, on what variables are missing for what countries. Income data is a mess (1997) with no clarity as to period covered, units, or currency conversion factors for that year. Also no continuous income variables. Had to abandon any analyses using income. Wanted to analyse ECHPS but as far as I could see it was very expensive.

I am at the early stages of researching the comparability of data on public expenditure by function. I may well have many of the problems mentioned.

q8a_2 Topic areas for guides to comparability across datasets

linking variables, weighting ...

violence against women and children

Health and ageing

This is essential in order to use these data. On most data sets no variables are defined in the same way in different countries not even age. This problem is not well understood

public expenditure by function

economic indicators

To compare correctly different sources

employment and the labour market, Eurostat weighting methods

childcare provision

A comparison of the liberal/conservative nature of the specific country's political parties, that is, what is liberal in the US is fairly conservative for many Europeans and surveys should help researchers reflect this. Also the meaning of certain policies regarding gender and family. For example, in ISSP many Swedish are not in favour of financial benefits for childcare, while many Americans are. This is not because the Americans are more liberal, but they are more comfortable with that type of provision. Researchers need to be aware of this.

q8b_2 Areas where alternative systems of classification would be of greatest use

Health and aging

EGP class

general government v. central/regional/local government

ESA 95 v. earlier classification

employment and the labour market

q8c_2 Across which areas would like to be able to search

linking observations

Demographics, SES vars, and health and aging

All datasets

Interpersonal violence

employment and the labour market

trade, investment

q9 What other services or developments would make it easier to access and use international datasets for research?

having workshops such as the recent BHPS (excellent) workshop
free access to university researchers
Chart of data/variables available by country *and year*
It would be useful to be able to email someone with knowledge of each dataset, e.g. a (volunteer) researcher for each dataset who would be willing to answer basic queries
Advice on relation between international data sets and the country data on which they are based.
list of research papers previously done in specific areas
Many of the country's data archives do not tell you right off whether the survey is in English or not. You should be able to type in a search option to find all surveys in English
more freely available historical statistics

q10_10 Which other International datasets do you currently use in teaching.

DHS
Labour Statistics

q11_6 Discipline teaching context

Political Science

q12 In what substantive areas is your teaching focused?

social research
political science, methodology, regime studies, elections
social policy [broadly defined]
Policy
development economics
organization, entrepreneurship, stratification
early years education
Political Sociology and Gender Roles
Microeconomics and Labour Economics
Data analysis for research students

q14_2 What themed teaching datasets would be most useful?

Regimes, elections, democratization
Social policy [broadly defined]
Poverty, Health, Employment
inequality, daily behavior, organization, employment
Social change; work and organisation; consumption, lifestyle & leisure

q15 What other services or developments would make it easier to access and use international datasets for teaching?

It would be really good to have country-level data packaged with individual level comparative data, e.g. key economic and social indicators for the years and countries of the ISSP data

There used to be a dataset called "The world handbook of social and political indicators". It's very out of date now, but was a great resource.

q16a_2 Topics to be covered in awareness raising courses

Public expenditure data (for political and economic research)

Interpersonal violence

Basic economic indicators

Demographic and socio-economic topics especially marriage/family and social stratification

Ageing

q16c_2 Specific techniques to be covered in courses

Regression, jack-knife, and bootstrapping, proportional hazards, transitional probabilities

bootstrapping

Multilevel analysis focusing on incorporating sample design into the analysis model

q16d_2 Specific themes to be covered in courses

Health and aging; aging in place

Poverty, Health

gender and work history

Welfare states

Poverty/social security

Interpersonal violence

Marriage/family, social stratification

political participation, human rights, violence, genocide, inequality, employment, etc

Ageing

democratization, attitudes toward political and economic reforms

employment and the labour market

childcare provision

Community, social capital

6. ESDS Longitudinal: User Survey Results

6.1 Introduction

ESDS Longitudinal is the value-added data service that aims to provide value-added data enhancements, user support and training for a range of the most popular UK based panel and cohort studies.

These include:

- the British Household Panel Survey (BHPS);
- the National Child Development Survey (NCDS);
- the British Cohort Study 1970 (BCS 70); and
- the more recent Millennium Cohort Study (MCS)

It will also encourage linkage with other datasets not directly supported by ESDS, such as the ONS Longitudinal Study and in conjunction with the ESRC National Strategy Committee, aims to further identify new longitudinal data collections suitable for inclusion, such as the English Longitudinal Study of Ageing (ELSA), the Avon Longitudinal Study of Parents and Children (ALSPAC), and the 1946 Birth Cohort. The service will create additional derived variables, enhance data and documentation standards across all studies, and improve information on weighting, missing data, and statistical adjustments. An online system for visualisation of documentation, subsetting and data matching is also planned as well as mounting online sampler files in Nesstar to facilitate the exploration of the content of data.

The work of the Longitudinal Data Service will be undertaken jointly by the UKDA and ISER (which houses the ESRC-funded UK Longitudinal Studies Centre) at the University of Essex. Further details can be found at www.esds.ac.uk/longitudinal/introduction.asp.

6.2 Respondents

There were 51 responses to the longitudinal user survey. A large majority of the respondents (36 or 72%) were based in UK HE institutions (Table 1A in the Appendix)¹. The next largest group of respondents were based outside the UK (18%). The remainder of the valid response (10%) are UK-based and split between the FE sector, private organisations and public bodies. Forty-five percent of respondents were members of HE/FE staff, a quarter were research students and 22% were researchers (contract or full-time) (Table 2A). Only 6% were undergraduate students.

A breakdown of the response by discipline shows that just under half of the respondents who provided information (21) work in the area of sociology or social policy (Table 3A). Human geographers (6), economists (5), psychologists (4) and management/business researchers (4) together accounted for a further 40% of the valid response.

6.3 Current or potential use of data sources

Two-thirds of the respondents (34) currently use longitudinal surveys (Table 4A). Table 1 indicates that of these 34 respondents, 56% currently use the BHPS, 27% use the NCDS and 21% use the BCS70. A small number of other longitudinal

¹ Note to tables: missing data arising from respondents not giving answers to questions have been exclude from the base numbers shown in tables and the bases used in percentaging. Tables named nA are found in the Appendix to this section, Section 5

surveys are being used. These are: the 1946 National Birth Cohort Study, the European Community Household Panel (ECHP), the DWP/PSI Families and Children Survey (FACS) and a Spanish pseudo panel of cultural consumption.

Table 1

	Longitudinal surveys usually used	Longitudinal surveys might use in the future
	%*	%*
None	33	4
BHPS	56	69
NCDS	27	57
BCS70	21	59
MCS	-	47
ELSA	-	41
Other	12	10
<i>N = longitudinal data users</i>	51	51

*Percentages may sum to more than 100 because respondents could give more than one answer

The same 34 respondents were asked how often they accessed or ordered data from the UK Data Archive. Around a quarter use the UKDA 'regularly', around a half use the UKDA 'occasionally', leaving a further quarter whose use of the Archive falls into the 'rarely/never' category (Table 5A). Fewer respondents (21) answered the question 'When did you last access/order any longitudinal data from the UK Data Archive?' Of these valid responses, 8 respondents had obtained data from the UKDA in the last three months and a further 9 had obtained data in the last year (Table 6A).

All 51 respondents were asked which longitudinal surveys they might want to use in the future. Of the studies listed, the BHPS is most likely to be used in the future (69%), closely followed by the BCS70 (59%) and the NCDS (57%) (Table 1). The two studies not held at the time of the survey by the UK Data Archive – the Millennium Cohort Study² (MCS) and the English Longitudinal Study of Ageing (ELSA) – were expected to be used by just under half of respondents (47% and 41% respectively).

A number of other studies that are not currently archived were mentioned as potentially useful for the future. These are: ALSPAC, a panel study of lone parents (initiated by the DSS), the ECHP and the New Earnings Survey longitudinal database. One respondent was also interested in several international longitudinal studies on ageing. Several respondents stated that they were 'not sure' which data might be useful in the future, with one respondent stating that 'it is difficult to know what is available if it is not archived'.

All respondents were asked to provide information on their use of data analysis software packages. The results presented in Table 7A indicate that SPSS is the most popular package (used regularly by two-thirds of the respondents). One-third of the respondents are STATA users and there is a small constituency of SAS users, over half of whom are based outside the UK.

6.4 Using data in research: barriers and requirements

Just under half of the respondents (25) use at least one of the longitudinal datasets held by the UKDA in a wide range of research areas (see Table 8A column 1). Social stratification and groupings, health, and labour employment were the topics most researched.

² The first sweep (interim version) of the MCS has recently been acquired by the UKDA, SN: 4683

These 25 respondents were asked whether they have experienced any difficulties in accessing/ordering longitudinal data. The results in Table 2 suggest that very few respondents had difficulty registering, ordering, downloading or with charges, with one respondent reporting that 'Access to and documentation of the BHPS is excellent, plus good support'.

Slightly larger numbers had difficulty with the data themselves. Around a third of the respondents had difficulty with each of 'conducting longitudinal statistical analysis', 'data documentation', 'obtaining details about variables' and 'merging files'. 'Handling missing data' proved most problematic: almost half of the responses (9 out of 19) to this question flagged the problem. One respondent suggested that:

"Since the use of longitudinal datasets is quite a complex task, I wish there was some sort of support for researchers handling the data."

Table 2

Difficulty accessing/ordering/using longitudinal data (research)	% *
<i>Accessing data</i>	
Downloading data via the UKDA	13
Cost of acquiring data	8
Registration for data services	8
Ordering via UKDA online system	8
<i>Using data</i>	
Handling missing data	47
Merging files	38
Data documentation	35
Conducting longitudinal statistical analysis	33
Obtaining details about variables	29
<i>N = use long data for research</i>	25

*Percentages may sum to more than 100 because respondents could give more than one answer

The same 25 respondents were asked which value-added resources they would like to see to facilitate the usage of longitudinal datasets. Table 3 suggests that additional derived variables and indices of harmonised/comparable variables would be resources best received by users, with as many as two-thirds of respondents suggesting that these would be *very useful* additions. Eight out of ten users would find online visualisation of data structure useful, but less, two-thirds, advocated the availability of sampler data files.

Table 3

Support for value-added resources (research)	%
Additional derived variables	
Not useful	4
Of some use	29
Very useful	67
Index of harmonized/comparable variables	
Not useful	13
Of some use	21

Very useful	67
Online visualization of data structure	
Not useful	17
Of some use	44
Very useful	39
Sampler data files	
Not useful	38
Of some use	29
Very useful	33
<i>N = use long data for research</i>	25

When asked in which topic areas the proposed resources might be useful, a range of responses were collected, many of which fitted into broader areas of social stratification and groupings, economics (income in particular), health, and education (Table 8A, column 2). Respondents were also asked to suggest other services or developments that might be useful. There were only four responses to this question but two of these mentioned further training on research methods specific to the use of longitudinal data, and two suggested that:

“information on panel attrition, drop outs, imputation techniques for missing variables.”

“the issue for us is where the coverage of cohorts reaching age 60 and over? A brief summary of numbers by sex and birth cohort across the surveys would be interesting.”

6.5 Using data in teaching: barriers and requirements

Only eleven respondents (22%) claim to use longitudinal data in their teaching. However, only 8 respondents provided details on the datasets used and information on the type of class taught. Of these 8, two use the NCDS, three use the BCS70 and three use the BHPS; three teach at the postgraduate level, two at undergraduate, two teach on professional courses and one teaches in the FE sector. Areas taught tended to be sociology, economics and research methods (Table 8A, column 3).

Those respondents who use longitudinal data for teaching only (4) were asked what difficulties they have experienced in accessing/ordering longitudinal data (7 respondents had already answered these questions, see Table 3). None of these respondents had experienced problems registering classes. A single respondent had experienced difficulty gaining permission to use data, ordering, downloading and the charges associated. The same respondent was also alone in having problems with the documentation and with merging files – though this respondent was joined by one other in having difficulty when it came to applying weights.

Four of the five respondents who answered the question ‘Would themed teaching datasets be helpful in your teaching?’ perceived them to be helpful.

6.6 Online training and support requirements

Sixteen of the 51 respondents (31%) had been on a training course on longitudinal data. Four respondents had attended the Essex Summer School course on the BHPS, and a further three mentioned the multi-level training course run by the Institute of Education.

All respondents were asked which training courses they might find useful in the future. The results presented in Table 4 indicate that large majorities considered each training course to be useful.

Topic-based training courses, drawing on a range of surveys were found by all but 6% to be useful. Online training and support materials (including 'how to' guides) also came out as one of the most popular types of training, with just under half saying they would find such resources *very useful*. Awareness training forums were found to be less popular although respondents saw this activity as particularly salient:

"The service is amazing. Anything you can do to raise the profile of the people who use these data sets would help."

Two respondents mentioned the possibility of an introductory day on longitudinal data sources in the area of child development.

Table 4

Support for particular training courses	%
Training course on a topic, drawing on a range of surveys	
Not useful	6
Of some use	63
Very useful	31
Online training and support (inc. 'how to' guides)	
Not useful	11
Of some use	44
Very useful	46
Training course on specific aspects of particular surveys	
Not useful	13
Of some use	60
Very useful	28
Training course to raise awareness	
Not useful	23
Of some use	55
Very useful	21
<i>N=all respondents</i>	<i>51</i>

Training courses on specific aspects of particular surveys would also be welcomed by eight out of ten respondents. Matching data files came out as an area requiring training while other comments highlighted further possible options:

"All the major longitudinal surveys, focusing on their structural idiosyncrasies."

"In general I think giving users an overview of the data structure and a sense of how to correctly merge data is of most use."

Finally, another respondent felt that further training was useful but should:

"complement rather than compete with the courses that other centres ... are already running."

6.7 Observations and recommendations

The results of the user survey indicate that there is a small constituency of longitudinal data users across a wide range of academic disciplines. Significantly, however, these users intend to access a wider number of longitudinal studies in the future, including ELSA which is not currently held by the ESDS but is a planned acquisition. We might extrapolate from this finding that resources directed at the supply and support of longitudinal data are well targeted.

In common with the large-scale government data user survey, longitudinal survey data are used more for research than for teaching. The following recommendations for future service developments reflect this:

- It is recommended that an important ‘first step’ is to raise awareness about what the ESDS longitudinal data service currently provides and is likely to provide in the future. To this end, we propose to:
 - Enhance the online information currently available on longitudinal data: specifically, to produce an FAQ resource and a web-based database of existing longitudinal data collections.
 - Produce promotional material for distribution to known longitudinal data users and at longitudinal data courses/workshops.
- A further important first step is to provide users (and potential users) of longitudinal data with online access to sampler datasets. This innovation will not only allow users to see what topics are covered by certain surveys but will also present users with the opportunity to visualize the structure of the data files. The release of a sampler BHPS dataset is planned for later this year.
- It is also clear that users would like to see a look-up index of variables that are harmonised or comparable between surveys and would welcome the provision of additional derived variables. The latter is an on-going process and will continue. The resources required for the production of an index of harmonised variables – and several other potential enhancements to the longitudinal data and metadata – will be explored in the coming months. Sampler data files are seen as less essential resources than these data enhancements.
- Courses/workshops (including those already established) should focus on the following themes:
 - Topic-based longitudinal data training: especially in the areas of health, income or education.
 - Specific aspects of particular surveys: for example, merging files and weighting data.
 - Research methods: a recurring theme for some respondents was the lack of opportunities to obtain specialist instruction in the analysis of longitudinal data (as opposed to data analysis in general).
- Online-training and support was popular and ‘how-to guides’ would be welcomed.

6.8 Appendix: ESDS Longitudinal User Survey Tables

Table 1A

Institutional status	N	%
UK HE	36	72
Non-UK	9	18
UK private organisation	3	6
UK public organisation	1	2
UK FE	1	2
<i>N=all respondents</i>	50	100

Table 2A

Status of respondents	%
Member of HE/FE staff	45
Research student	26
Researcher (contract or full-time)	22
Undergraduate	6
Visiting academic	2
<i>N=all respondents</i>	51

Table 3A

ESRC primary discipline	%
Sociology	23
Social policy	21
Human geography	13
Economics	11
Psychology	9
Management and business studies	9
Stats and computing	4
Librarian	4
Area studies	2
Education	2
Political science and IR	2
<i>N=all respondents</i>	47

Table 4A

Use longitudinal surveys	%
Yes	67
No	33
<i>N=all respondents</i>	51

Table 5A

Access/order data from UKDA	%
Rarely/never	28
Occasionally	48
Regularly	24
<i>N = users of long data</i>	29

Table 6A

UKDA - most recent use	%
> 3 years ago	10
> 1 year ago < 3 years ago	10
> 3 months ago < 1 year ago	43
< 3 months ago	38
<i>N = users of long data</i>	21

Table 7A

Analysis software used	% *
SPSS	67
STATA	33
SAS	14
<i>N = all respondents</i>	51

**Percentages may sum to more than 100 because respondents could give more than one answer*

Table 8A

Research areas/themes	Being researched	Areas useful for added enhanced data products+	Areas taught/might use data teaching-	Training areas requested
	N*	N*	N*	N*
Economics	6	8	1	
Labour and employment	8	3	1	3
Politics	1	1	1	
Education	6	4	1	1
Information and communication	3	1		
Health	10	7	2	4
Social stratification and groupings**	21	10	4	4
Society and culture	5	3		1
Demography	1	1	1	1
Social welfare policy and systems	1			
Psychology	3	1		

+ additional derived variables, indices of harmonised/comparable variables, sampler data files.

** respondents often noted more than one area*

7. ESDS Qualidata: User Survey Results

7.1 Introduction

ESDS Qualidata, the value-added service to support qualitative data replaces and augments the former ESRC-funded service previously offered by Qualidata. A key feature of the new service is to provide a range of enhanced and accessible digital qualitative data and documentation resources for research and teaching purposes. In addition it will significantly enhance user support activities relating to these data. The service will promote a training programme designed to generate an increase in the use of these datasets and an appreciation of secondary analysis of qualitative data. Further details can be found at www.esds.ac.uk/Qualitative/introduction.asp. The work builds on the Qualidata unit's expertise and international reputation in this area.

7.2 Respondents

Seventy-two people participated in the ESDS User Consultation Survey for qualitative data.

More than six in ten respondents (61%) were from UK Higher Education institutions (see Table 1A in the Appendix)¹. Just under one third (23) of respondents were members of Higher Education/Further Education staff, while a quarter were undergraduates (19) and almost equal proportion were contract researchers and research students (18% and 17% respectively) (see Table 2A). Respondents within UK Higher Education, UK Further Education and non-UK academic institutions came from a wide range of disciplines (Table 3A); the most frequently recorded discipline was sociology or social studies (21%), followed by health studies and medicine (11%), psychology and social policy (both 10%).

7.3 Current or potential use of data sources

Four out of ten respondents (29) said they had undertaken secondary analysis of qualitative data. Of these, almost three-quarters (20) had consulted qualitative data for secondary analysis in the last 3 months, and a further quarter within the last 3 years (Table 4A).

Of those who had ever undertaken secondary analysis, text materials were consulted by many more of the sample than audio-visual materials. The most widely used types of data were in-depth structured interviews and semi-structured interviews, consulted by approximately half of this group of respondents (52% and 48% respectively) - see Table 1. This was closely followed by case study notes and interview summaries and notes (both 38%) and press clippings (35%). A further quarter had used field notes and minutes/records of meetings and group discussion transcripts (each with 24%). No respondents had used kinship diagrams or other anthropological materials. Six percent mentioned that they had consulted other textual data sources in addition to those listed, which included policy documents and websites

Turning to audio-visual data sources, these had hardly been consulted at all, with under one in ten having accessed data in this format. The material used most by this group was naturally occurring speech/conversation (by 10% or 3 respondents).

¹ Note to tables: missing data arising from respondents not giving answers to questions have been excluded from the base numbers shown in tables and the bases used in percentaging. Tables named nA are found in the Appendix to this section, Section 7

Table 1

What kind of materials did you consult?	Text	Audio-visual
	%*	%*
In-depth/unstructured interviews	52	3
Semi-structured interviews	48	3
Interview summaries or notes	38	
Case study notes	38	3
Press clippings	35	3
Group discussions	24	
Open questions in structured interview surveys	24	3
Field notes	24	3
Minutes/records of meetings	24	
Observational recordings	21	3
Participant observation notes	17	3
Correspondence relating to research projects	17	
Personal documents	17	
Psychological test data	14	7
Photographs	14	7
Unstructured/semi-structured diaries	7	
Naturally occurring speech/conversation	10	10
Other	6	
Kinship diagrams/other anthropological materials		
N= respondents who had have consulted sources	29	29

*Percentages may sum to more than 100 because respondents could give more than one answer

A large majority (70%) of respondents thought they might use sources of qualitative data in the future (See Table 5A). While few specific databanks were mentioned, such as CHILDES² and TALKBANK³ databases, data covering a range of topics would be considered as useful sources. Data sources in the areas of 'social stratification and groupings' (including social exclusion and the family) and health were most often cited. When asked why they would not consider using sources of qualitative data in the future, the few respondents that answered, felt that their own data sources would satisfy their own research needs, and that other sources might be less relevant to their own work.

Respondents were asked which qualitative data analysis software (CAQDAS) packages they used on a regular basis. Under half (44%) of all respondents said they used a CAQDAS package on a regular basis (Table 6A). Packages used regularly were ATLAS-ti (11%), NVivo (11%), N4/NUD*IST (10%), followed closely by winMAX (8%). None of the sample used C-I-SAID on a regular basis and very few cited using Qualrus (3%) or Ethnograph (1%). Other softwares mentioned were: answr; CLAN, TASX, Mineset (a data-mining software), SPSS and Excel.

² CHILDES is the Child Language Data Exchange System which provides tools for studying conversational interactions, including database of transcripts, programs for computer analysis of transcripts, methods for linguistic coding, and systems for linking transcripts to digitized audio and video. childes.psy.cmu.edu/

³ TalkBank is an interdisciplinary research project supported by the National Science Foundation to Carnegie Mellon University and the University of Pennsylvania, aiming to foster fundamental research in the study of human and animal communication. www.talkbank.org/

When asked whether they would expect to use a CAQDAS package to re-analyse qualitative data, only 4 out of 10 respondents said that they would.

7.4 Using data in research: barriers and requirements

Although there was an interest in re-using material, in practice, only 40% (27) of the sample had undertaken secondary analysis of qualitative data in their own research. While 70% of these had revisited their own data, there was a greater incidence of using colleague's data (44%) rather than acquiring archived data via a dissemination service (33%).

Just under half of respondents (13) said they had consulted data for either descriptive purposes or to undertake comparative research, a restudy or a follow-up study. 41% said they had used data either for secondary analysis, research design or methodological advancement, or for teaching and learning. Only 22% had consulted data for the purposes of verification. The topics they were researching are listed in Table 5A. Health and social stratification and groupings are the most popularly researched fields.

Respondents were asked to provide information on other difficulties they had experienced when accessing and using qualitative data in their research. A common view was that it was generally difficult to obtain or access relevant material. Of those who have undertaken secondary analysis of qualitative data for research, 87% experienced difficulties locating appropriate data (Table 2).

Table 2

Have you experienced any difficulties accessing and using existing sources of qualitative data in your research?	
Accessing data	%*
Locating appropriate data	87
Format of data e.g. paper	67
Archival organisation of data	62
Access conditions	62
Ethical constraints	47
Other	18
<i>N=all who use for research (min)</i>	22
Using data	
Time to consult sources thoroughly	80
Adequate data or project documentation	68
Re-analysing qualitative data	33
<i>N=all who use for research (min)</i>	22

*Percentages may sum to more than 100 because respondents could give more than one answer

While this is probably largely due to a lack of availability of material across topics of research interest, it may also be a result of poor knowledge about available resource discovery methods to locate qualitative data. One respondent remarked that,

"The main problem is finding appropriate qualitative data. Even locating short samples of data that is suitable for use in teaching research methods is difficult, since published reports rarely include long enough segments of data."

In a similar vein,

"This is the first time I have ever tried to find raw data, ever. I have spent two hours on a computer finding lots of journal articles on my topic, but no raw data. The charts, graphs and conclusions in these articles have to come from somewhere."

Respondents also expressed concern about other aspects of accessing data: Half saw problems with the organisation of data, half with access conditions and 44% with format

of data, for example, data not being available in digital format. A further third expressed difficulties with ethical constraints.

When asked about difficulties faced when using data, of the 22 respondents, 59% said they did not have time to consult sources thoroughly, 56% thought that data or project documentation were not adequate for their needs, and 22% admitted to having difficulties in reanalysing qualitative data. One respondent noted that "using service-based data identified inconsistencies and gaps in data recording." Another noted that there was a noticeable "lack of books on the subject."

Ninety-six percent agreed that thematic guides and samples of data would make it easier to access and use archived qualitative data in their respondents' own research (Table 3). Themes mentioned were: ethnic minorities, health issues; information society; consumer studies; UK classic community studies; marriage and the family; social stratification; oral history interviews; and research commissioned by agencies.

Would online access to data held at the UK Data Archive be considered useful? An overwhelmingly proportion of the sample (96%) thought that it would, with as many as two thirds agreeing that that it would *very useful*. One respondent felt that "samples of data online to aid selection of data sets for secondary study would be beneficial", as would having "the ability to access electronic and/or hard copies of e.g. transcripts away from archives and re-code/analyse using preferred software/manual methods."

Equally, a very high proportion (90%) considered that coded data, in addition to raw data, from the original research would be useful (with 31% believing it to be *very useful*). When asked about the benefits of being able to explore data or conduct new basic thematic analysis online, 94% believed that it would be beneficial, with just under half of those suggesting that they would find it a *very useful* facility.

Finally, respondents were less keen on the usefulness of having access to data in audio-visual format in addition to textual sources - while 64% felt this to be quite useful, only 15% believed this to be *very useful*. Formats suggested of most use were: audio-visual data in digitised and XML format (defined delimited format - DDI standard was cited); coded data in the appropriate file formats (e.g. Atlas-ti); audio files in MP3, MPEG; images in BMP, tiff, or JPEG2000 format. Finally access to tapes and transcripts of interviews was considered to be useful in order to check accuracy and aid interpretation of the material.

Table 3

Would any of the following be useful in using archived qualitative data in your research?	
	%
Thematic guides and samplers for data sources available	
Not useful	4
Of some use	67
Very useful	29
<i>N=all</i>	69
Online access to data held at the UK Data Archive	
Not useful	4
Of some use	24
Very useful	72
<i>N=all</i>	68
Coded data in addition to raw data from the original research	
Not useful	10
Of some use	58
Very useful	32
<i>N=all</i>	67
Ability to explore data/conduct new basic thematic analysis online	
Not useful	6
Of some use	49
Very useful	45
<i>N=all</i>	67
Access to archived data in audio-visual format in addition to textual sources	
Not useful	21
Of some use	64
Very useful	15
<i>N=all</i>	66

Next, respondents were asked how useful they would find access to data from mixed method studies. All but 4% were in favour of having these data accessible, with 60% believing that it would be very useful.

Finally, researchers were asked to select from a predefined list of themes, already proposed by Qualidata as potential future thematic resources, those of most interest. Table 4 indicates that the most popular theme was family and social networks (54%) followed closely by social class, work and employment (51%) and life stories and social change (49%).

Table 4

Thematic resources based on high quality data and classic studies. Which themes would be of most interest to you?	
Theme	%
Family and social networks	54
Social class, work and employment	51
Life stories and social change	49
Youth culture	36
Crime and social order	26
Mental health and institutionalisation	19
Other	26
<i>N=all (min)</i>	72

Other more specific themes called for largely reflected respondents' own areas of research (column 2, Table 5A). Reference was made explicitly to the UK classic community studies.

When asked about any other useful service or developments, specific comments pointed to the desire to have portable data formats and access to international qualitative data. Other comments included:

"Easily downloadable data files in XML or RTF format for use in CAQDAS packages."

"Making a search engine that connects the various datasets in several countries. These datasets or databases should provide a standardized (perhaps XML type) output. So the researcher could work in a cross-national database in the given topics."

"Resources in other languages."

"Linkage with existing international projects."

7.5 Using data in teaching: barriers and requirements

Thirty percent of respondents have used qualitative data in their teaching, and a further 18% who have not used it for teaching might do so. Of these, 70% have used or might use data for post-graduate teaching, while 49% do or might do so for undergraduate teaching (Table 7A). Fewer use data for professional courses (27%) or for further education (15%).

The most common specific elements of courses taught by respondents using qualitative data were: qualitative data analysis (76%) followed by interview guide construction or interviewing techniques (67%). Half use/will use data for teaching about data quality issues (49%); and under one in four for addressing triangulation or mixed methods research (39%) and participant observation or field note keeping (36%). Other specific aspects of researcher training mentioned were: CAQDAS packages, theoretical and epistemological bases, grounded theory, discourse analysis, narrative analysis, conversation analysis and focus group training.

The substantive areas in which qualitative methods teaching was focused are shown in Table 5A (column 3). Social stratification (e.g. marriage and the family, gender), health and psychology featured as the most popular areas. When asked about the sources of data respondents use/might use in teaching, the majority said they use/would use their own data (94%), compared to 73% who had consulted/would consult colleagues data and fewer, 67% who said they had obtained/would obtain data from a dissemination service such as the UK Data Archive. Other sources mentioned were: the CHILDES database; samplers from CAQDAS; and the ethnographic literature.

A reasonably high proportion of the 22 respondents who had used data in their teaching had experienced difficulties with accessing qualitative data (Table 5).

Table 5

Have you experienced any difficulties accessing or using archived qualitative data in your teaching?	
Accessing data	%
Locating appropriate data	70
Gaining permission to use data	48
Format of data e.g. paper	48
Archival organisation of data	50
Availability of tailor-made teaching datasets	64
<i>N=all who use for teaching (min)</i>	22
Using data	
Teaching with 'real' qualitative data	41
Time to consult sources thoroughly	80
Adequate data or project documentation	62
<i>N=all who use for research (min)</i>	22

The biggest problem was found to be with locating appropriate data (70%), followed by the lack of availability of tailor-made teaching datasets (64%). Half also noted problems in gaining permission to use data, with the format of data, e.g. only available in paper format, and in the archival organisation of data collections.

When it came to applying data in a teaching context, respondents noted that the biggest problem was in not having the time to consult existing sources thoroughly (80%). Four out of ten said they saw problems in teaching with 'real' qualitative data sources, and just under two thirds thought that data or project documentation were not adequate for the purposes of employing archived data sources in their teaching data.

A number of open comments supported the fact that there exists difficulties in using data for teaching:

"It is hard to find suitable data quickly."

"Time is the greatest challenge."

That said, certain products or services that helped to overcome these barriers would make utilising archived data sources a more attractive prospect.

"It would be excellent if qualitative data was available to use in teaching. Data from my own research is often not suitable because of ethical issues."

"Easy access and availability."

"Prepared formats available online to manipulate."

"Resources in other languages."

Overall, a substantial proportion of respondents who were teaching with qualitative data liked the idea of having themed teaching datasets available to use. Just under half (45%) of respondents thought that they would be very helpful in their teaching and a further 48% believed that they would be a little helpful. The most popular topics for themed datasets were identified as health, marriage and the family, and gender.

7.6 Online training and support requirements

Thirty percent of respondents said that they would find courses to raise awareness of the datasets and their research potential very useful (Table 6). A further 61% said they would find these of some use. Those who thought that introductory forums would be

beneficial, proposed courses in a variety of broad substantive areas, like health, employment and social policy (Table 5A, column 4), but also more introductory methodologically oriented courses. The latter incorporated, for example, collecting and analysing data, grounded theory, the Internet as a field for researching, CAQDAS, conversation analysis and oral history methods, and more generally, "how to implement data".

One respondent observed that:

"there is a need for such courses in psychology, particularly within professional areas such as clinical and educational psychology. I do some such training myself and would find it helpful if more courses were available that I could send my students."

and another noted, "the need to get qualitative methods taken seriously in medicine."

While eighty-eight percent of respondents said that they would find courses on specific aspects of particular datasets, or kinds of data useful only one in five of these respondents would find this idea very beneficial. Specific data types mentioned most frequently were life stories, mixed methods data and focus groups. One respondent remarked on the need to gain:

"acceptance and raised awareness of a multiple methods approach to fit specific research projects."

Table 6

Which of the following forms of training might you find useful?	
%	
Courses to raise awareness of qualitative data and their research	
Not useful	9
Of some use	61
Very useful	30
<i>N=all respondents</i>	<i>66</i>
Courses on specific aspects of particular datasets, or kinds of data (life stories, focus groups, mixed methods data or classic studies)	
Not useful	12
Of some use	70
Very useful	18
<i>N=all respondents</i>	<i>66</i>
Topic-based courses drawing on a range of data	
Not useful	10
Of some use	55
Very useful	35
<i>N=all respondents</i>	<i>67</i>
Online training and support resources	
Not useful	8
Of some use	52
Very useful	40
<i>N=all respondents</i>	<i>67</i>

Classic community studies of the 1960s, the CLAN/CHILDES and TALKBANK, the use and analysis of participant observation and field notes, observational recordings, interview summaries, and documents were also suggested as possible areas of training.

Just over one-third (23) of respondents said that they would find topic -based courses drawing on a range of data sources *very useful*. A further 55% said they would find these of some use. Those who thought they would be *very useful* were particularly

interested first in criminology (a quarter specifically mentioned this topic), followed by health, social change and social class, and family and social networks (see Table 5A, column 4).

Finally, respondents were asked their opinion on the usefulness of online training and support resources, including course materials and 'how to' guides. An overwhelming proportion were in favour of this kind of support provision, with 40% saying that they would such resources *very useful*. A further 52% said they would find these of some use. Specific reference to the attractiveness of 'how-to' guides, and exemplars and case studies of re-use, shown in Table 8A.

Finally, respondents were asked to add any other comments about qualitative data resources. The majority of responses were positive and looking forward to a new service focussed on providing enhanced access to data, support and training. Two-thirds were happy to take part in further user consultation exercises:

"I am a new lecturer and am very interested in how I could use this resource in preparing new lecture material. I look forward to learning more about it".

"The main issue for me is time to develop my knowledge, whether this is through short courses or online."

"I'm really pleased that you are undertaking this consultation and look forward to improved qualitative data archives as a result - this will be tremendously helpful to me for both teaching and research."

7.7 Observations and recommendations

The results of the user survey confirm what we already suspected, that existing qualitative data sources are not that well used. This is due to a number of reasons: the lack of sources available in areas of greatest demand; and perhaps more crucially, in the relative inaccessibility (in terms of physical format and archival location) of many data sources currently archived. The initial ESRC driven hub and spokes model of qualitative data provision (clearing house) that focused on data being dispersed across the UK in traditional (paper based) repositories, is clearly not satisfactory in terms of meeting users' desire for quick and easy access to data.

As anticipated, there is an overwhelming call 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. While data is used to a greater extent for research than teaching, across the board, there is an eagerness to use archived qualitative data in the future.

7.7.1 Data types

- The most popular data types, on which a data acquisition strategy and data provision should focus are:
 - In-depth/unstructured interviews
 - Semi-structured interviews
 - Interview summaries or notes
 - Case study notes.
- Data acquired should primarily be text-based and in digital format. There is little call by users for audio-visual sources, unless for linguistically-oriented analysis.

- Data should be organised in a systematic and familiar way. All digital data acquired the UKDA under ESDS Qualidata will be subject to standard processing and metadata creation procedures in line with relevant international standards.
- Provision should be made for enabling users to import data and coded data from completed research projects into CAQDAS packages. Coded data can provide a navigable entry point into voluminous amounts of data, allowing users to explore and browse data quickly, even if the schema used is just one researcher's way of classifying data. It is recommended that work, currently under development by Qualidata, on a DTD for defining a standard and exchangeable format for data and associated encoding should be actively pursued.

7.7.2 Data topics

- Health, family and social networks, social class, work and employment, social change, and crime and social order were the most frequently mentioned topics throughout the survey. It is recommended that the themes selected for the construction of thematic guides and samplers of data for the first four years of ESDS Qualidata are:
 - 1st year: Social class and social change
 - 2nd year: Health
 - 3rd year: Family and social networks
 - 4th year: Crime and social order
- Access to international data is of interest to users. It is recommended that a small scoping study is done to establish what are the key cross-national and comparable sources of qualitative data. This is not high priority and could be undertaken in year 3 or thereafter.

7.7.3 Data services

The ability to find data efficiently is seen as a priority. Equally, the need to access data quickly and efficiently is also seen as essential in enhancing the usability of archived data.

- It is recommended that a search facility for qualitative data is available via a more general data portal for UK-based social science data. This has already been established as one of ESDS's major deliverables. Browsing and introductory descriptions of popular datasets, and an overview of resources, by theme are also advocated.
- Online access to qualitative data via the instant download service is a fairly new service that has been offered by the UKDA. The ESDS should ensure that it is adequately publicised to wider audiences.
- Enabling interactive online access to data, ie by browsing and exploring content of data via web browsers, would be welcomed by users and should be considered as a priority for research and development work.
- Data should be free, where possible, from restrictive access conditions or 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.
- Since few users request access to audio-visual materials, acquisition and digitisation of these sources should be seen as secondary to textual and image sources

7.7.4 Training and teaching

- The idea of tailor-made and thematic training datasets was popular. These could be produced on the same yearly themes as identified under 'data topics' above.
- The teaching of qualitative data analysis *per se* is seen as a problem area. Priority should be given to creating web based samplers, simple overviews of data analysis together with exemplars and case studies. Ideally these would link up with other training initiatives and resources for qualitative data, such as the ESRC Research Methods Programme.

7.7.5 Courses and workshops

The following types of courses and workshops are proposed:

- Workshops to raise awareness of the datasets and their research potential prioritising the substantive areas of health, employment, social policy, and oral history.
- Introductory workshops on 'How do I get started?' to demonstrate locating data (searching the ESDS site) and exploring documentation, downloading and carrying out data analysis.
- Hands-on workshops on specific aspects of qualitative data analysis, particularly methodologically-oriented courses on grounded theory, CAQDAS, mixed methods, focus groups, life story and oral history methods, and conversation analysis.
- Topic-based courses drawing on a range of data, prioritising criminology, health, social change and social class, and family and social networks.
- Workshops that combine training in the use of CAQDAS packages and qualitative analysis would be useful, suggesting joint workshops with the ESRC CAQDAS Networking project would be a fruitful collaboration.
- It is further recommended that course materials are made available online.

7.7.6 Online training and support requirements

- It is recommended that online topic based user guides are produced on the same yearly themes as above in order to get people started on using datasets relevant to the topic, and directing users in the right direction of data and documentation.
- 'How-to' guides and case studies of secondary data analysis were found to be particularly popular with respondents.

7.8 Conclusion

These recommendations appear to confirm many of the ideas that were formulated in the original ESDS Qualidata bid, and thereby tie in nicely with the milestones and deliverables already set out in the contract. The survey has enabled us to prioritise thematic areas and pinpoint the most useful kinds of support resources. Finally, and importantly, it has mandated the need to focus acquisition activities on digital textual materials, and to continue work on creating data standards and tools for online qualitative data exploration.

7.9 Appendix: ESDS Qualidata User Survey Tables

Table 1A

Affiliation/Institution of respondents	
	%
UK Higher Education	61
Non-UK	19
UK Further Education	10
UK Public Organisation	8
Other	2
<i>N=all respondents</i>	72

Table 2A

Status of respondents	
	%
Member of HE/FE staff	32
Undergraduate	26
Contract researcher	18
Research student	17
Information Professional/Policy maker	4
Visiting academic	3
<i>N=all respondents</i>	72

Table 3A

Discipline of HE/FE based respondents	
	%
Sociology/social studies	21
Health studies and medicine	11
Psychology	10
Social Policy	10
Management and Business Studies	8
Economics	8
Statistics and Computing	8
Economic and Social History	7
Human Geography	3
Social work studies	3
Education	3
Political Science & International Relations	3
Environmental Planning	3
Linguistics	1
Other	1
<i>N=all in UK FE, UK HE or non-UK academics</i>	61

Table 4A

When did you last consult qualitative data sources for secondary analysis?	
	%
More than 3 years ago	4
More than 1 year ago, but no more than 3 years ago	18
More than 3 months, but no more than 1 year ago	7
In the last 3 months	71
<i>N=all who have undertaken secondary analysis</i>	28

Table 5A

Research areas/themes	Being researched	Might use data for research*	Areas taught/might use data teaching-	Training areas requested
	N	N	N	N
Economics	2	2	2	1
Trade, industry and markets	2	2	1	
Labour and employment	4	4	2	1
Politics				
Law, crime and legal systems	1	3	1	4
Education	4	1	3	1
Information and communication	2	1	4	2
Health	9	6	5	4
Housing and land use planning		2		1
Social stratification and groupings	8	10	13	4
Society and culture	3	5	4	1
Social welfare policy and systems	3	2	1	
Science and technology			1	
Psychology	2	1	5	
History		1		
Reference and instructional resources		2		
<i>N (min)</i>	27	72	33	66

* includes thematic guides and samplers

Table 6A

Which data analysis software packages do you use on a regular basis?	
	%*
none	44
Atlas-ti	11
N4/NUD*IST	10
NVIVO	11
Winmax	8
Qualrus	3
Ethnograph	1
SPSS	4
Other	7
<i>N=all respondents</i>	72

*Percentages may sum to more than 100 because respondents could give more than one answer

Table 7A

In which context have you used the datasets in your teaching?	
	%*
Postgraduate	70
Undergraduate	50
Further education	15
Professional courses	27
Other	36
<i>N=all who use/might use for teaching</i>	33

*Percentages may sum to more than 100 because respondents could give more than one answer

Table 8A Training and support resources of most use

"the 'how-to' guides and support sources would be useful"

"'how-to' guides, worked examples through from data collection to analysis and interpretation"

"course materials, guides 'how to', case studies (e.g. CAQDAS, Internet Qualitative Research)"

"case studies online would be especially useful in terms of ease of access and 'how to' guides would help also"

"case studies, and definitely 'how to' guides including 'how to find' data sets"

"examples of exemplary secondary studies using different types of qualitative data"

"methodological/ethical guides on aspects of doing secondary analysis"

"tutorials on how to get the most from using qualitative data sources"

"brief introductions on a wider variety of applications of qualitative data analysis with hints on methods and procedures, including evaluation with CAQDAS packages"

"provide rather for up-to-date info (frequent updates) than comprehensive in-depth coverage: i.e. provide for good and viable starting points"

"particularly useful for introductory methods teaching at both undergraduate and postgraduate levels"