



**Economic Impact Evaluation
of the
Economic and Social Data Service**

**Charles Beagrie Ltd
and
The Centre for Strategic Economic Studies (CSES)
University of Victoria**

March 2012

Contents

Executive Summary	4
1. Introduction.....	9
1.1. Background	9
1.2. Aims and Objectives.....	9
1.3. The ESDS	10
1.4. Conceptual Approaches in the Study.....	12
2. Methods Used to Collect and Analyse Data.....	14
2.1. Desk Research.....	14
2.2. Interviews	14
2.3. Online Surveys.....	15
3. Methods and Approaches used to Assess Economic Impact	17
4. Economic Analysis.....	21
4.1. Weighting the Data	21
4.2. Data Analysis.....	24
4.3. Investment and Use Value.....	24
4.4. Contingent Value	27
4.5. Consumer Surplus	29
4.6. Net Economic Value.....	29
4.7. Potential Efficiency Impact	30
4.8. Return on Investment in Social Science Data Infrastructure	31
4.9. Summary of Economic Analysis	33
5. Illustrating Wider Benefits.....	36
5.1. Benefits Identified in Desk Research	36
5.2. Benefits identified in the Interviews.....	37
5.3. Benefits Identified in the User Survey	42
5.4. Benefits Identified in the Depositor Survey	43
5.5. Summary of Benefits Analysis.....	43
6. The Case Studies	45
6.1. Case Study 1: E3MG: An Energy-Environment-Economy (E3) Model	45
6.2. Case Study 2: Knife Crime: A Review of Evidence and Policy.....	47
6.3. Case study 3: Trends in Obesity	49
7. Conclusions	51
7.1. Economic Impact and Benefits of the ESDS.....	51
7.2. Economic Impact and Benefits of Social Science Research Data Service Infrastructure in General	51
7.3. Effectiveness of the Evaluation Methodology and Approaches.....	52
8. Summary of Recommendations	56
8.1. Good Practice/Lessons for Maximising the Benefits of Research Data Service Infrastructure.....	56
8.2. Approaches to Future Economic Impact Evaluations	56
References.....	58
Appendix A: Findings from the Surveys.....	64
Appendix B: Summary of Interviews	87

Acknowledgements

We would like to thank: Peter Williams and Anna Palaiologk for their input as researchers to the study; Matthew Woollard, Lucy Bell and colleagues at the UK Data Archive for their collaboration and support, together with Keith Cole and colleagues at MIMAS and Vanessa Higgins, Angela Dale and colleagues and students at CCSR. Margaret Macadam at ESRC provided feedback and guidance throughout. Finally, our thanks go to all the interviewees and survey respondents who gave valuable time and input to the study.

Neil Beagrie (Charles Beagrie Ltd) and John Houghton (CSES)

March 2012

Executive Summary

Background

The Economic and Social Research Council (ESRC) evaluates the quality and impact of its research and data infrastructure investments. The Council's Evaluation Committee operates a well-developed system for evaluating academic quality and impact, and in recent years has developed qualitative and quantitative approaches to assessing policy and practice impacts of the research projects it funds and for establishing its broader contribution to the economic and social well-being of the UK.

Through this study, the ESRC is now looking to assess the economic benefits of its investment in the Economic and Social Data Service (ESDS). The ESDS is a distributed service based on a partnership between the Universities of Essex and Manchester, and is co-funded by ESRC and JISC. The annual operating budget for ESDS is around £3.3 million. There are around 23,000 active users of the ESDS, including users from outside the academic community. The overall aim of the ESDS is to promote wider and more informed use of data for research and teaching in the social sciences and to ensure that these data continue to be available over time.

This report provides an overview of methods and approaches explored in the study, issues encountered and our findings and recommendations.

Aims and terms of reference

The aims of this study are to: (i) evaluate the economic benefits and impact of ESDS; and (ii) contribute to the further development of impact evaluation methods that can provide ESRC with robust estimates of the economic benefits of its data service infrastructure investments. This includes ESDS's economic impact on the creators and users of the data it provides access to. The study does not attempt to quantify policy impacts but it does include case studies to illustrate the broader context and provide understanding of the breadth of both quantifiable and qualitative benefits including areas such as policy.

The terms of reference for this evaluation are to identify and assess the economic impact of the ESDS by:

- Applying economic valuation techniques where appropriate to derive a quantitative estimate of the economic benefits of the services and data provided by the ESDS;
- Presenting the results of the economic assessment within the context of a broader qualitative analysis of the benefits of the ESDS;
- Conducting case studies of the impact generated by research based on data accessed through the ESDS;
- Reflecting on the results and identifying best practice and lessons for:
 - maximising the benefits from research data infrastructure investments; and
 - developing approaches for future economic impact evaluations.

Economic impact of ESDS

Our economic analysis included a range of approaches, starting with the most immediate and direct measures of value that are likely to represent lower bound estimates of the value of ESDS data and services and moving outwards to estimates of the wider economic benefits. They included:

- *Investment and use value* - with the amount of time and money spent producing/obtaining ESDS data and services indicating the minimum value of ESDS;
- *Contingent value* - with the amount that users would be willing to pay to access ESDS data and services and/or willing to accept to forego access indicating the value of ESDS to them;
- *Consumer surplus* - with the total willingness to pay minus the cost of obtaining indicating the benefit they derive from ESDS;
- *Net economic value* - with the users' benefits derived minus the cost of providing ESDS data and services indicating the net economic value;
- *Efficiency gains* - with estimates of the value of research and teaching efficiency gains realised by ESDS users indicating the impacts of ESDS on the user community; and
- *Increases in returns on investment in data creation and infrastructure* - with estimates of the potential increases in returns to investment arising from the additional use facilitated by ESDS indicating the impacts of ESDS on the funder, data creator/depositor and user communities.

Data for the analysis were drawn from desk-research, interviews, and two online surveys of ESDS registered users and depositors respectively. For the user survey, with the agreement of ESRC, we excluded registered school students and under-graduates along with a small number of registered consenting users from non-Anglophone, non-Eurozone countries. The exclusion of school and under-graduate students from the survey has little impact on the economic analysis (see Section 4).

Results of the economic analysis are summarised in Figure 4 and briefly described below:

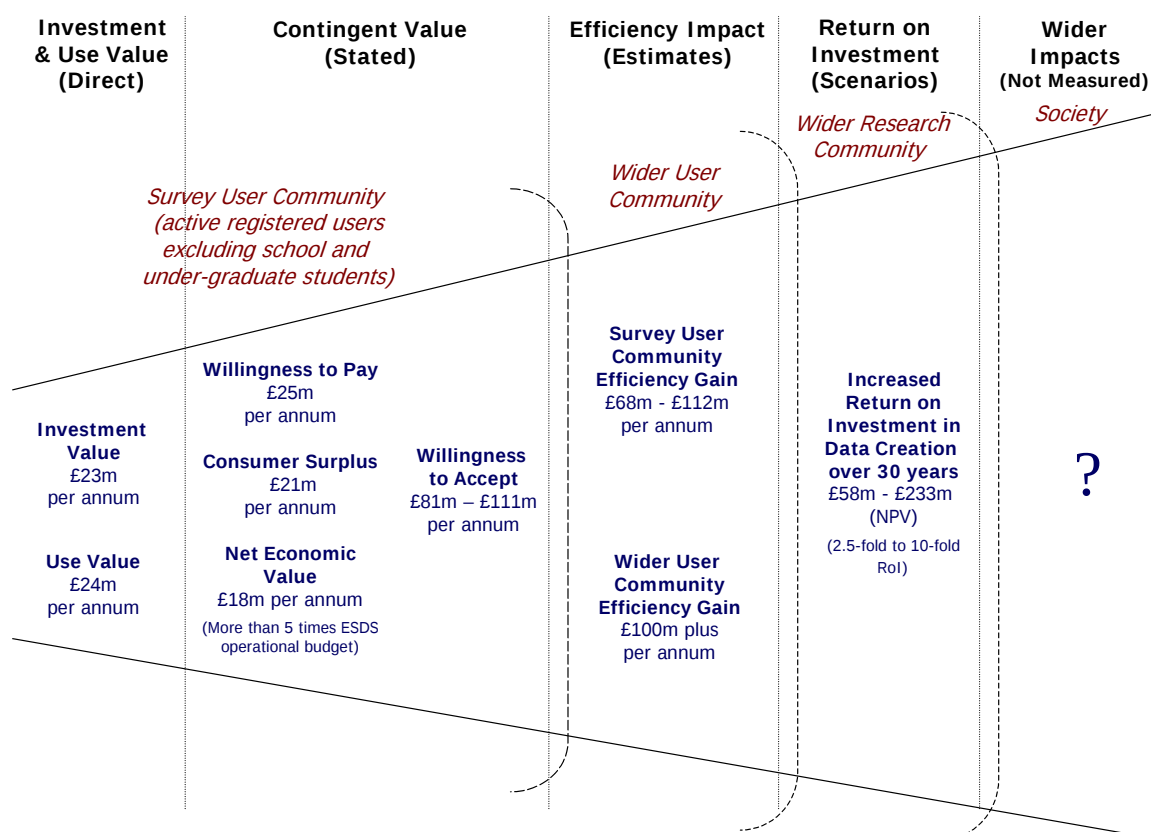


Figure 4: The value and impacts of ESDS research data infrastructure.

Source: Authors' analysis.

Looking at the lower bound, both the *investment and use values* of appear to be around £23 million per annum (excluding the value to school and under-graduate users). This is reflected in the *contingent valuations*, where users' willingness to pay amounts to around £25 million per annum. Hence, the *consumer surplus* is around £21 million per annum (after deducting user access costs) and the *net economic value* (net of operational costs) around £18 million per annum - more than five-times the ESDS operational budget. As is typically the case in contingent valuation techniques, what users would be willing to accept in return for foregoing access to ESDS data and services is much higher, at £111 million per annum.

Moving to the wider economic impacts and benefits of ESDS research data infrastructure, we estimate *efficiency impacts* for ESDS's active registered user community (excluding school and under-graduate students) of £68 million to £112 million per annum, which might translate to at least £100 million per annum or more for the wider user community.

Exploring scenarios in section 4 suggests that ESDS research data infrastructure services facilitate an *increase in the return on annual investment in the data and ESDS research data infrastructure services* of £58 million to £233 million over 30 years (Net Present Value). Given non-sunk data infrastructure costs of around £23 million per annum, this suggests a 2.5-fold to 10-fold return on investment.

Box 1 Return on Investment for data, library and information services

While individual studies focus on different information services and content and use different methods and measures, it is possible to explore Return on Investment (RoI) findings to give a sense of how ESDS data and services compare:

- British Library (2004) concluded that: "The British Library generates value around 4.4 times the level of its public funding."
- King (2010) summarized findings relating to library services and concluded that: special libraries exhibit an RoI of 2.9 to 1, academic libraries 3.4 to 1 (for staff), and public libraries 5.8 to 1.
- Imholz et al. (2007) summarized a number of studies, finding, inter alia, that Ohio public libraries showed a RoI of 3.8 to 1 and the Carnegie Library in Pittsburgh 3 to 1.
- Houghton (2011) estimated the benefit/cost ratio of the Australian Bureau of Statistics making data and publications freely available online and using Creative Commons licensing at 5.3 to 1.

Both the benefit/cost ratio of net economic value to ESDS operational costs (5.4 to 1) and the increase in returns on investment in data and related infrastructure arising from additional use facilitated by ESDS (up to 10 to 1) are at the higher end of return on investment for library and information services, national statistics and other information and data services reported over the last decade (See Box 1).

Wider Benefits of ESDS

An extensive set of benefits from ESDS were identified by ESDS users, depositors, and staff in the desk research, interviews and online surveys. These have been described and arranged in an ESDS benefits summary in section 5 of the report (see Box 2).

Three impact case studies were also undertaken and are described in section 6. The case studies represent a spread of use of the main ESDS data collections and focus on the major policy issues of climate control, knife crime, and obesity. All three are valuable in demonstrating how research based on ESDS has had significant impact in terms of debate and media coverage of these major social issues. However we found it difficult to identify case studies that can conclusively show direct impact on policy and practice. This is a challenging task because of the widely acknowledged difficulties associated with attribution and time-lags.

Box 2: Summary of ESDS Benefits arranged in a KRDS Benefits Framework

Benefits Summary for ESDS	
Direct Benefits	Indirect Benefits (Costs Avoided)
Time and resource savings for researchers and teachers	No re-creation of data -Trusted Digital Repository status eliminates re-ingest costs
Verification of research through increased data citation thanks to relevant citation information and	Lower future archiving costs increase likelihood of data being available, earlier in the lifecycle

tools Access to data provides new research opportunities by increasing use of data within collections thanks to proper Collections Development Policy Re-purposing and re-use of data maximises the value of data holdings when the data is available for third parties Increasing research productivity Assisting in provision of a skills base Customer service ethos of ESDS staff	Re-purposing data for new audiences Use by new audiences Protecting returns on earlier investments Reduces potential duplication of effort
Near Term Benefits	Long-Term Benefits
Value to current researcher and students Single point of access Increasing speed of access to data Ease of access for researchers and students Quality of documentation and contextual information, and resources for teaching No data lost from post-doc turnover thanks to a reliable preservation system Secure storage for data intensive research Availability of data underpinning journal articles Research data integrity since data is of high quality	Data preserved for the long-term Secures value of high quality data for future researchers and students Value added over time as collection grows and develops critical mass Input for future research by maximising use across data user community (ie including students) Impact on wider profession as a centre of excellence Promoting quantitative methods and skills Fostering innovation in research and data management practice
Private Benefits	Public Benefits
Benefits to sponsor of research Benefits to sponsor of data service provider Benefits to researcher Fulfil grant obligations Increased visibility/citation Aggregator of data for data providers – providing them with one point of access to UK customers Higher usage (and profile) of their data for depositors Removes user burden from depositors	Source of high-quality and often unique data Motivating new research Enables research that otherwise could not be undertaken Research integrity since others can check the outcomes of research Service targeted at academic community and supports their needs Fostering transferable skills in data analysis Articulating user needs to data providers

Introduction

Background

The Economic and Social Research Council (ESRC) evaluates the quality and impact of its research and data infrastructure investments. The Council's Evaluation Committee operates a well-developed system for evaluating academic quality and impact, and in recent years has developed qualitative and quantitative approaches to assessing policy and practice impacts of the research projects it funds and for establishing its broader contribution to the economic and social well-being of the UK¹.

The evidence from the Evaluation Committee's programme of work demonstrates the ESRC's significant contribution to economic and social well-being, and the Committee has recently been developing new approaches to valuing that contribution. The Evaluation Committee recognises that this is a challenging task because of the absence in most cases of recognisable markets for social science outputs, and the widely acknowledged difficulties associated with attribution and time-lags. The key aspect of the ESRC's approach is that while economic evaluation is important and necessary, it needs to be put into context. This context involves broader, more qualitative assessments that capture the full range of benefits from social science research and the research data infrastructure that supports it.

The ESRC decided next to commission an exploratory evaluation study to assess the economic benefits and impact of its main research data infrastructure, the Economic and Social Data Service, and to develop further its approach to economic impact evaluation.

The ESRC issued a call for proposals in May 2011 for a consultancy to provide an Economic Evaluation of Research Data Infrastructure: specifically its investment in the ESDS. Following open competition, a partnership of Charles Beagrie Ltd (lead partner) and the Centre for Strategic Economic Studies at Victoria University in Melbourne (Prof. John Houghton, senior economist and partner) was selected for the study. The study was conducted between July and December 2011.

Aims and Objectives

The aims of this study are to: (i) evaluate the economic benefits and impact of ESDS; and

(ii) contribute to the further development of impact evaluation methods that can provide ESRC with robust estimates of the economic benefits of its data service infrastructure investments. This includes ESDS's economic impact on the creators and users of the data it provides access to. The study does not attempt to quantify policy impacts but it does include case studies to illustrate the broader context and provide understanding of the breadth of both quantifiable and qualitative benefits including areas such as policy.

The terms of reference for this evaluation are to identify and assess the economic impact of the ESDS by:

¹ <http://www.esrc.ac.uk/impacts-and-findings/impact-assessment/analysis-and-scoping.aspx>

- Applying economic valuation techniques where appropriate to derive a quantitative estimate of the economic benefits of the services and data provided by the ESDS;
- Presenting the results of the economic assessment within the context of a broader qualitative analysis of the benefits of the ESDS;
- Conducting case studies of the impact generated by research based on data accessed through the ESDS;
- Reflecting on the results and identifying best practice and lessons for:
 - maximising the benefits from research data infrastructure investments; and
 - developing approaches for future economic impact evaluations.

The ESDS

Background

The ESDS is a distributed service based on a partnership between the Universities of Essex and Manchester, and is co-funded by ESRC and JISC. There are around 23,000 active users of the ESDS, including users from outside the academic community. The overall aim of the ESDS is to promote wider and more informed use of data for research and teaching in the social sciences and to ensure that these data continue to be available over time.

The ESDS selects, prepares for preservation, archives and provides access to a wide range of research data including **data collections** from major Government and ESRC funded surveys and longitudinal studies (for example, the General Household Survey, Understanding Society, and several Birth Cohort studies), data arising from ESRC funded research projects and doctoral theses, as well as data made available under licence from International Governmental Organisations, including the IMF and the OECD.

Organisational diversity

It is worth emphasising that the ESDS is a diverse organisation involving a range of partners in its service provision (Cathie Marsh Centre for Census and Survey Research, MIMAS and the UK Data Archive) and a range of very different data services (ESDS Government, ESDS International, ESDS Longitudinal, and ESDS Qualidata). Although the ESDS is a very broad set of services, several major research data infrastructure services also funded by ESRC such as the Secure Data Service or the Census services are not part of it (though some activities across these services are managed together). In addition, there are services within UKDA such as the History Data Service funded from other sources. Even users and depositors with very close links to ESDS do not always understand or can have an erroneous view of what is in or out of scope of the ESDS service.

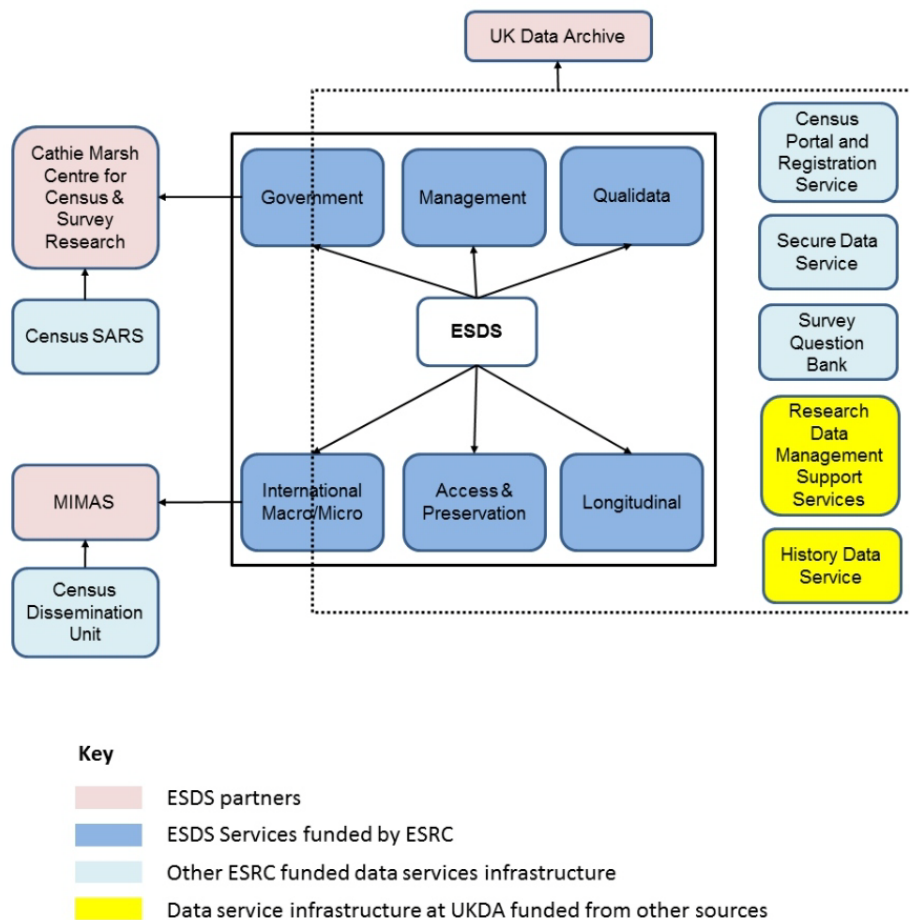


Figure 1: Organisational Diagram for ESDS and Other Services.

Range of data depositors/suppliers and users

One consequence of this breadth and diversity covered by ESDS is that there is a corresponding extensive range of depositors, data suppliers and users of the service. For example, data suppliers and depositors range from major international organisations that have their own access infrastructure and license data via ESDS to UK users, to researchers in small-scale research projects who might deposit with the service data created in an ESRC-funded research project once or twice over their career. The ranges are often quite extreme and have made the survey design and analysis much more difficult than it would otherwise have been.

Analysing and understanding use of ESDS

This diversity in organisation, data suppliers and user base has implications for the interpretation of access statistics for ESDS. An extract of an explanation of this by the ESDS in its annual report for 2009-2010 is reproduced below.

‘In order to accommodate the wide range of data needs ESDS provides a combination of data access tools. As a result of these multiple routes to, and forms of, data access it is not a straightforward task to measure data usage and certainly

not easy, or in some cases appropriate, to attempt to compare the various usage figures across the component parts of ESDS.' (ESDS 2010a)

The published statistics for ESDS data collections delivered by access method during 2009-2010 (after ESDS 2010a) were as shown in Tables 1 and 2.

Table 1: Data collections delivered, 2009-10

<i>2009-2010</i>	<i>Download</i>	<i>FTP</i>	<i>Via/access</i>	<i>Portable</i>	<i>Nesstar</i>	<i>Total</i>
ESDS Core	10,009	606	942	5	16	11,578
ESDS Government	28,936	2,387	2,577	1	84	33,985
ESDS International*	1,282	148	159	0	8	1,597
ESDS Longitudinal	7,599	352	455	6	18	8,430
ESDS Qualidata**	1,074	24	89	0	0	1,187
Total	48,900	3,1517	4,222	12	126	56,777

Notes: *International micro data collections only, macro data collections counted separately (see below). ** Nesstar does not include qualitative data

Table 2: Other access sessions and reports delivered, 2009-10

	<i>Total</i>
Data Report downloads for international macrodata via Beyond 20/20	121,356
Data analysis sessions for international macrodata via Beyond 20/20	6,691
Data analysis sessions for Nesstar online microdata browsing and analysis service	8,716

Within the study we found that this diversity in access is mirrored by complexity in terms of actual use (as distinct from just access) and potential impacts. Users of some resources have high browse and discard rates (perhaps using 50 per cent or less of what they access). Others download and access once but may use the data collections they download multiple times, for example in teaching assignments. Similarly the metadata, research data and documentation may all be accessed independently or in combination for different purposes. Each impact is subtly different and even negative data (such as browsed and discarded information) can contribute positively to research outcomes. Questions in the interviews and user survey simply asked what data types respondents had used and how many times they had "accessed/downloaded" data, which sought to cover the range of access methods from browsing through to downloading in as simple a way as possible.

Conceptual Approaches in the Study

The conceptual framework and starting points for the wider assessment of benefits in this study were the Keeping Research Data Safe (KRDS) activity cost model (Beagrie et al 2008) and KRDS Benefits Framework (Beagrie et al 2010, Charles Beagrie 2011).

The KRDS Benefits Framework is a tool for identifying, assessing, and communicating the benefits from investing resources in the curation/long-term preservation of research data.

Once potential benefits have been identified it can also assist in articulating them to a broad audience of stakeholders and in customising their expression to address different stakeholder audiences. The Framework organises benefits along three broad dimensions: the outcome achieved; when the outcome is achieved; and who benefits from the outcome. Each of these dimensions can be subdivided into two categories: direct and indirect benefits, near-term and long-term benefits and internal and external benefits respectively. This is summarised graphically below.

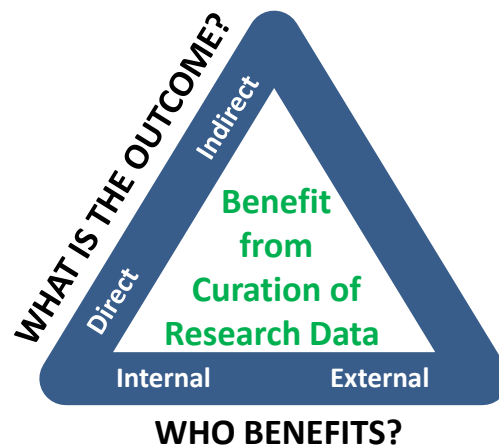


Figure 2: The KRDS Benefits Framework

The outcomes from a KRDS Benefits analysis can be presented in a number of simple tables and a set of generic benefits identified by KRDS edited and augmented for a specific case.

A number of approaches were used to explore the economic value and benefits of ESDS data and services, beginning with approaches that can be seen as estimating minimum values and moving progressively toward approaches that can be seen as measuring the wider value. These include investment and use value, contingent valuation using stated preference techniques, welfare approaches to estimating consumer surplus, and a macro-economic modelling approach that seeks to explore the returns to investments in data creation and hosting.

These are supplemented by use of the KRDS benefits framework and case studies, to build on the economic approaches and set them in a broader economic and policy context.

Methods Used to Collect and Analyse Data

In selecting conceptual approaches to the study, we took account of the practical limitations of collecting the necessary data through interview and survey techniques, and sought to maximize economy in data collection through commonality (ie the same data can be used to inform more than one of the approaches). We combined: (i) desk-based analysis of existing evaluation literature and reports, looking at both methods and findings; (ii) existing data from KRDS and other studies; (iii) existing management and internal data collected by ESRC and ESDS such as user statistics, internal reports, and the ESDS Mid-Term Review; and (iv) original data collection in the form of semi-structured interviews, case studies, and an online survey of ESDS users and depositors.

Desk Research

Desk-based research has provided an important source, allowing us to minimise the amount of data collection required through interviews and the online survey. ESDS and ESRC have been the primary sources for data relating to use of ESDS data and services (eg. user registrations, downloads and sessions, data collections, etc.), as well as operational and activity costs. The UK Office for National Statistics (ONS) and international agencies, such as OECD and Eurostat have also provided a range of input data.

We examined data and analyses drawn from a range of sources including:

- Relevant evaluation literature for ESDS;
- Existing research data service infrastructure evaluation reports;
- Documentation on ESDS data usage and users;
- Internal ESDS management information and other relevant internal reports;
- ESDS Annual Reports and Mid-Term Review reports;
- Relevant ESRC reports and documentation.

Interviews

Twenty-five structured individual or group interviews involving thirty-three individuals were carried out with key stakeholders. This was five more interviews than originally specified. Four of the interviews were with ESDS staff at different locations. Participants included:

- The ESDS Director and key staff across all ESDS sites;
- Researchers and analysts at academic and research organisations who deposit and/or use data through the ESDS, including leading professors, readers and lecturers in academic and research institutions;
- Users, depositors and others at government institutions;
- Policy makers and practitioners at key organisations such as RIN (Research Information Network), and JISC (Joint Information Systems Committee);
- Representatives of non-academic institutions and the private sector with an interest in research data service infrastructure;
- Lecturers, teachers and others involved in teaching quantitative skills.

Potential interviewees were identified by ESDS staff, based on their knowledge of users, depositors and others; and a scrutiny by the research team of ESDS literature (particularly annual reports and newsletters) which included details of research projects exploiting ESDS data collections. This list was then prioritised in consultation with ESRC and a set of reserves agreed should anyone on the initial list be unavailable or decline to be interviewed.

Potential interviewees were sent an explanatory letter and brief details of the project by email and invited to contribute by offering their views in person or by telephone. Those willing to do so were further asked to suggest a suitable time when they could be available for 30 – 60 minutes for interview. ESDS staff group interviews were of longer duration being scheduled for 2-3 hours each.

Of the twenty-five interviews, fifteen were by telephone (using Skype Voice over Internet Protocol in one case) and ten in person (with three group interviews having one interviewee joining via telephone). Only one person approached formally declined to participate citing time constraints and pressures of work rather than disinterest in the project.

User and depositor interviews each followed a formal structure although subjects were encouraged to expand on answers to closed questions. The interviews opened with a broad question asking respondents:

‘Can you please tell me very briefly about your work, and the importance of ESDS data and services in it (whether in your research, teaching, training PhD students, policy development or some other capacity)’

In some cases lengthy and wide-ranging responses to this question precluded the need to ask some of the following ones.

Some interviewees who occupied specialist positions answered questions more appropriate to their own circumstances.

Overall the high acceptance rate for interview was mirrored by the high level of engagement, participation and contribution during the interviews. Most interviewees were extremely positive in their contributions to the study contributing for the full 60 minutes and often more.

The interviews were conducted in parallel to the development of the online surveys due to the tight timescale for the study. The individual interviews were collated for further analysis. Findings from the interviews helped to inform development of the survey questionnaires and the analysis, and provided additional insights for the study that are discussed further in section 5 and Appendix B.

Online Surveys

Two online surveys were conducted, aimed respectively at depositors and ESDS Registered Users.

Development of the survey questionnaires

The survey questionnaires were developed iteratively by the project team with external review and input from ESDS staff, ESRC and others. The initial starting point was the structured questionnaire used for interviews. Given the nature of some of the economic valuation approaches being explored, and the range and complexity of the survey populations, substantial effort and experimentation was needed to (re)design this for an online survey. In particular, the number of questions was reduced to a maximum of 10 substantive questions for the depositor survey and 22 questions for the user survey.

Prior to the survey, there were concerns expressed about likely “survey fatigue” amongst potential recipients and the relatively low response rates to two recent surveys in the ESDS community (for example there were 289 responses from the ESDS user community to the survey for the RIN data centres study). Significant effort, therefore, was spent on trying to reduce the likely burden on recipients in terms of time to complete the questionnaires, wording of the invitations to participate, advance notice of the surveys via the ESDS website, and offering five £50 Amazon vouchers in a draw for participants. As a result, the surveys enjoyed high response rates (more than 15 per cent for the user survey and 30 per cent for the depositor survey) and good completion rates given the topics and number of non-mandatory questions.

The survey design used a range of standard survey approaches, including use of “critical instances” such as the last access/download (for users) or last deposit/data supply (for depositors). These questions were supplemented by questions asking for views on “average” or “typical” experiences in case low response rates required some estimation based on this. Responses to these “backstop” questions are likely to be less reliable than those to critical instances from a large number of respondents.

A small number of questions also sought specific financial information on the costs of creating the data collection, preparing it for deposit, or “contingent valuation” (ie willingness to pay or accept). Answers to these questions have been interpreted carefully in the context of additional text comments in the surveys and other findings from the interviews and desk research to ensure that protest answers are excluded.

The most novel and difficult approach in the survey was a set of questions asking for responses in terms of defining percentage changes (rather than a more qualitative 1-5 scale with values such as “low” or “high”). The original approach of asking for these to be stated by the user with positive or negative values underwent several modifications in light of user feedback requesting pre-defined scales.

Draft survey questionnaires benefited from feedback from ESRC and ESDS staff, two separate rounds of user testing conducted by post-graduate students at CCSR, and an independent review by an associate of the company. However, the timeframe for development and completion of the surveys was very tight and undoubtedly the development of the questionnaires would have benefited from a slightly longer timescale.

Methods and Approaches used to Assess Economic Impact

Much has been written about the costs and, to a lesser extent, the benefits of more open access to research publications. There is also a rapidly growing literature on the costs and benefits of providing open access to public sector information (PSI), including economic and social statistics, which are similar to some of the ESDS holdings. To date, much less attention has been given to the value of the open curation and sharing of research data, although a few pioneering studies exist. There is also a related and extensive literature on the value of library and information services. Each of these literatures suggests possible approaches to economic valuation, although some are more directly relevant than others. Our purpose here is to draw ideas from this literature and assess what might be most useful for this study.

Research publications have been the focus of much economic analysis, although the vast majority focuses on costs of creation, access and preservation and few look at the value and benefits of access. Those addressing the value and benefits have done so from the supply-side, in the form of macro-economic analysis based on estimates of the impacts of changes in accessibility and efficiency on returns to R&D expenditure (Houghton et al, 2009; Houghton and Sheehan 2009), and from the demand-side, in the form of estimates of the impact of research on innovation and the value of that innovation to firms (Mansfield 1991, 1998; Beise and Stahl 1998; Houghton et al. 2011). The latter approach is not readily applicable to ESDS as commercial users make up a relatively small share of total users, and the former is relatively data intensive - depending on robust estimates of levels of R&D spending, average returns to that spending and the extent of changes in accessibility and efficiency that result from the data being made available. Data limitations make this approach difficult.

Public sector information (PSI) has been the focus of a number of studies seeking to estimate its value and the benefits to be derived from making it freely available. PIRA (2000) combined measures of the investment cost (ie the amount spent on the collection/generation of the PSI) and expenditure on PSI by users and re-users, then for final users, estimated value as expenditure on PSI or, where the PSI was freely available, as the investment cost of its collection/generation. In the MEPSIR study of Dekkers et al. (2006), demand and economic performance were measured in an extensive survey by directly asking both public content holders and re-users for key economic data, such as total turnover against turnover related to PSI, total number of staff against the number of staff dedicated to handling PSI, and estimates of the domestic market size for a particular type of PSI. DotEcon (2006) and collaborator Pollock (2009) adopted a bottom-up approach to estimating the economic value of PSI products and services in the UK, seeing the net economic value of PSI as the willingness to pay for PSI minus the cost of supplying it. These studies suggest the potential of approaches based on investment and use value, and estimating net economic value from a survey approach to willingness to pay or accept (ie contingent valuation through a stated preference approach).

Research data repositories have been the topic of a number of cost studies, but few have tried to look at the value or benefits of the open curation and sharing of data. Beagrie et al. (2008, 2010, and 2011) investigated the medium to long-term costs to Higher Education Institutions (HEIs) of the preservation of research data, developed an activity cost model for research data archiving and also a Framework for assessing the mainly non-economic

potential benefits from preservation of research data. Fry *et al.* (2008) sought to identify benefits arising from the curation and open sharing of research data. Based on the work of co-authors Houghton and Rasmussen, the report presented a simple example of cost-benefit analysis applicable to an individual data collection or repository, based on costs and potential cost savings. It described the data requirements and walked the reader through the process step-by-step. The approach was then extended to explore the more diffuse benefits of data curation and sharing at the institutional and disciplinary levels, based on a macro-economic approach to measuring the impact of increased accessibility on returns to research expenditure.

Library and information services have been the focus of many studies estimating their value in quantitative and qualitative terms, with the former including numerous studies based on investment and use value and estimates of consumer surplus, and fewer based on contingent valuation leading to cost-benefit analyses (Imholz, *et al.* 2007; Oakleaf 2010). Most focus on the more direct economic impacts and calculate a Return on Investment (RoI), but some explore wider impacts and outcomes (eg. the relationship between library spending and successful grant applications (Tenopir *et al.* 2010)) and explore Social Return on Investment (SRoI) by combining the quantitative and qualitative measures using a balanced scorecard approach. One example, measuring the economic impact of the British Library, combined contingent valuation in the forms of willingness to pay and accept with investment value and estimates of the cost of alternatives (British Library 2004).

Our approach

What all of these studies suggest is that the field is relatively new and no single approach has dominated across these different but related fields. Consequently, we propose to draw on a number of approaches to explore the economic value and benefits of ESDS services and data, beginning with approaches that can be seen as estimating minimum values and moving progressively toward approaches that can be seen as measuring some of the wider value. These include investment and use value, contingent valuation using stated preference techniques, welfare approaches to estimating consumer surplus, and a macro-economic approach that seeks to explore the impacts of increased use on returns to investment in data creation/collection. In selecting these approaches, we have taken account of the practical limitations of collecting the necessary data through interview and survey techniques, and sought to maximize economy in data collection through commonality (ie the same data can be used to inform more than one of the approaches).

The most direct indicators of value are the *investment value* (ie the amount of resources spent on the production of the good or service) and *use value* (ie the amount of resources spent by users in obtaining the good or service). Measures of the investment in access suggest the minimum amount that the good or service is worth to the consumers. *Contingent valuation* involves the assignment of money values to non-market goods and services based on preferences (DTLR 2002). If a good or service contributes to human welfare, it has economic value, and whether something contributes to an individual's welfare is determined by whether or not it satisfies that individual's preferences. An individual's welfare is higher in situation A than situation B, if the individual prefers A to B. Preferences are revealed by what an individual is willing to pay for a good or service and/or by the amount of time and other resources spent obtaining the preferred good or service. Where preferences are not revealed in the market, individuals can be asked what they would be willing to pay or to accept in return for the good or service in a hypothetical market situation (ie stated preference).

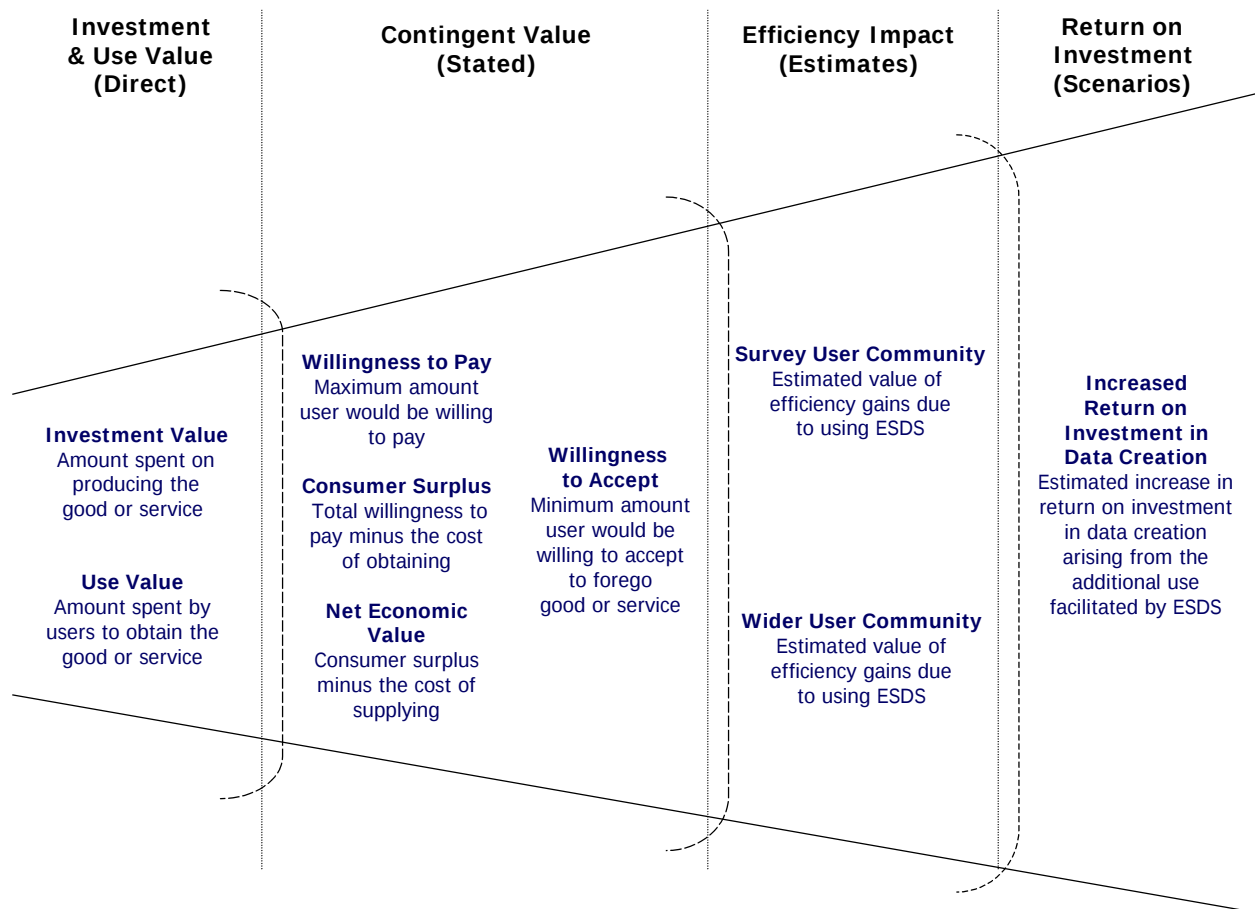


Figure 3: Methods exploring the value and impacts of ESDS research data infrastructure. Source: Authors' analysis.

The benefit or welfare impact of a good or service for a consumer is measured by the *consumer surplus*. In a market situation, willingness to pay is made up of what is actually paid and any excess willingness to pay over and above the price paid (ie consumer surplus). Hence, consumer surplus is the net gain derived by the consumer from the purchase of a marketed good or service. In a non-market context, all the willingness to pay is consumer surplus because there is no market price. In practice, however, some expenditure, be it in the form of time or money, will be incurred in obtaining the non-market good or service (eg. the time spent accessing ESDS data). In this case, the consumer surplus will be the net gain (ie willingness to pay minus the cost of obtaining).

Wider benefits and impacts can be explored by looking at the *efficiency gains* enjoyed by users and assigning an economic value to them (eg. activity cost savings), and by estimating the impacts of increased data use on *returns to investment* in the data collection/creation and the related data infrastructure services necessary for hosting and sharing the data. As these latter impacts are recurring during the useful life of the data it is necessary to use a simple Perpetual Inventory Method to estimate the overall value of the impacts over time.

Each of these methods is briefly explained in the context of its application in Section 4 (below).

Economic Analysis

The focus of the economic analysis is to try to shed light on the value of ESDS data and services to users and to the wider community. Self-evidently, our survey respondents are a self-selected sub-group of users and depositors: (i) because, for users, they opted to allow ESDS to contact them upon their registration with ESDS, and (ii) because they responded to the survey. The response rates to both user and depositor surveys were high and a brief analysis of responses by the major possible "cuts" of the data by affiliation, field of research, role and data type used/deposited suggests few major differences beyond the obvious (eg. staff are willing (and able) to pay more for such services than are students). Nevertheless, on average, one might reasonably expect that those taking the time to respond to an ESRC initiated survey are likely to use and value ESDS data and services more than those that did not respond. Moreover, we deliberately excluded school and under-graduate students from the user survey, as they were unlikely to be able to answer many of the cost and value questions. It is also important to note that few, if any, users use all of ESDS, but rather experience just part of it, and they can only express value relating to the part they use. Consequently, it is important to consider how best to weight the survey responses (described in detail in Annex A) to better reflect the wider user community and uses of ESDS data and services.

Weighting the Data

A good deal of effort was put into weighting the survey data prior to the economic analysis, in order that the values expressed by respondents reflect the broader user community, and not simply our respondents, and the totality of ESDS data and services, and not simply the part that our respondents used. As the focus of the study was on the value to users and in use, user survey weighting was done on the basis of data collections delivered (use). The depositor survey was weighted by data collections acquired and processed for online delivery. This involved a two-step process of disaggregation of responses by data type accessed/downloaded for users and deposited for depositors and re-aggregation by data collections delivered/acquired, followed by the calculation of weighted means from responses about costs, time, values, etc. (Table 3).

*Table 3: Data deposited and delivered by data type, 2009-10.
Source: ESDS Annual Report 2009-2010, Authors' analysis.*

<i>Data Collection</i>	<i>Deposited</i>	<i>Deposit Weighting</i>	<i>Delivered</i>	<i>Use Weighting</i>
Longitudinal	60	8 per cent	8,430	15 per cent
International	180	23 per cent	1,597	3 per cent
Government	317	41 per cent	33,985	60 per cent
Qualidata	19	2 per cent	1,187	2 per cent
Core / Other	202	26 per cent	11,578	20 per cent
Total	778	100 per cent	56,777	100 per cent

For the *user survey* we weight the survey data by data collections delivered (accessed/downloaded) to reflect the wider user community (not just respondents). User survey respondents reported the data types they use, with 894 respondents giving us 1,410 data type responses (Q5). For each of the five data types we create a cut, so that every respondent in that cut uses that data type. However, they also use other data types, so the same respondent can appear in multiple cuts. Indeed, these cuts gave a total of 2,896 data type responses (Table 4 'Counts'). The mean values for costs, willingness to pay or accept, etc. in each of the five data type cuts is then distributed according to the percentage share of data type responses within that cut, and reassembled by data type by adding the distributed values for each of the cost/value variables from each of the five data type cuts (Table 4 'Shares').² These data type mean values are then weighted by the percentage share of ESDS reported data collections delivered 2009-10 by data type (Table 3).

This weighting does not include 20/20 and Nesstar accesses, which creates a much lower weighting for international (multi-nation) data than would be the case if the more than 120,000 data report downloads through 20/20 were included. However, sensitivity testing suggests that the impact of such a change in weighting on the overall results is relatively small - creating a slightly lower willingness to pay/accept, which is likely to be because most of these international data are available elsewhere (eg. directly from the OECD, etc.).

Table 4: Data type responses (User Survey Question 5)

	Long	Multi	Govt	Quali	Other
Counts:					
Long	422	79	206	65	80
Multi	79	208	61	36	36
Govt	206	61	404	50	86
Quali	65	36	50	167	44
Other	80	36	86	44	209
Shares:					
Long	50 per cent	9 per cent	24 per cent	8 per cent	9 per cent
Multi	19 per cent	50 per cent	15 per cent	9 per cent	9 per cent
Govt	26 per cent	8 per cent	50 per cent	6 per cent	11 per cent
Quali	18 per cent	10 per cent	14 per cent	46 per cent	12 per cent
Other	18 per cent	8 per cent	19 per cent	10 per cent	46 per cent

We deliberately excluded under-graduate and school students from the sample, along with a very small number of registered and consenting users from non-Anglophone non-Eurozone countries. Under-graduate and school students were excluded because we felt that they would have found it difficult to answer questions about time and costs, and we had no basis

² For example, the mean values reported by respondents in the Longitudinal data type cut are disaggregated to data types according to the shares by data type - Long 50 per cent, Multi 9 per cent, Govt 24 per cent, Quali 8 per cent and Other 9 per cent, and so on for each of the data types (Table 4 'Shares').

for attributing a cost to their time (eg. average annual salary) and would most likely have zero-costed their time. Consequently, for the user-related estimations, we take the user community to be the total reported active registered users minus the students we did not survey, averaged over the last three years (ie around 18,100). As a result, we are missing the value that school and under-graduate students get from their use of ESDS data and services, but we are likely to be getting some upward bias from the fact that people who bothered to respond may be more likely to value ESDS. Of course, we are also missing the value of ESDS to the many non-registered users who download material that is freely available from the ESDS website (eg. guides) without registering.

For the *depositor survey*, we weight by data collections acquired by data type, reflecting the need to explore depositor costs. Again, the data types deposited were redistributed to data type by responses, as multiple data types could be deposited by each respondent. The process is the same as the use weighting, described above, with mean values from cuts by data type and reported data type deposited distributed and reassembled (Table 5), then weighted by data collections acquired and processed for online delivery by data type during 2009-10 (Table 3).

Table 5: Data types responses (Depositor Survey Question 3)

	Long	Multi	Govt	Quali	Other
Counts:					
Long	26	3	7	4	5
Multi	3	9	1	1	1
Govt	7	1	40	1	5
Quali	4	1	1	80	9
Other	5	1	5	9	71
Shares:					
Long	58 per cent	7 per cent	16 per cent	9 per cent	11 per cent
Multi	20 per cent	60 per cent	7 per cent	7 per cent	7 per cent
Govt	13 per cent	2 per cent	74 per cent	2 per cent	9 per cent
Quali	4 per cent	1 per cent	1 per cent	84 per cent	9 per cent
Other	5 per cent	1 per cent	5 per cent	10 per cent	78 per cent

A further key parameter is the number of data collections acquired (ie should it be new data collection additions or should it include updates and new editions?). On balance, we felt that data creation costs should be for new data collections only (approx. 450 per annum), while preparation and deposit costs should include updates and new editions as well (approx. 750 per annum in total).

Data Analysis

This section explores the economic methods, data and data requirements, and presents an analysis based upon them. As noted in section 3, a number of approaches are being used to explore the economic value and benefits of ESDS data and services, beginning with approaches that can be seen as estimating minimum values and moving progressively toward approaches that can be seen as measuring some of the wider value. These include investment and use value, contingent valuation using stated preference techniques, welfare approaches to estimating consumer surplus, and an approach that seeks to explore the impacts of increased data use on returns to investment in data creation/collection. In selecting these approaches, we took account of the practical limitations of collecting the necessary data through interview and survey techniques.

Investment and Use Value

The most direct indicators of value are the investment value (ie the amount of resources spent on the production of the good or service) and use value (ie the amount of resources spent by users in obtaining the good or service). Measures of the investment in access suggest the minimum amount that the good or service is worth to the consumers.

Investment value includes annual ESDS funding, the costs that depositors face in preparing data for submission/deposit and in making that submission/deposit. For simplicity, each can be annualised (ie expressed as an annual cost in current prices and at current levels of activity). Activity times have been converted to costs using the annual average salaries for

academic staff and graduates reported in the most recent Times Higher Education Salary Surveys and the TRAC fEC method of full economic costing (See Annex A for details).³

Scaled to the annual average number of data collections acquired and processed for online delivery over the three years 2007-08 through 2009-10 and weighted by type of data collections acquired by ESDS during 2009-10, depositor preparation and deposit costs amount to almost £20 million per annum (ie 750 data collections deposited at a weighted mean cost of £26,570 each). The annual operating budget for ESDS is around £3.3 million. Hence, treating data collection costs as sunk costs (ie assuming that the data would be collected whether or not ESDS existed), *investment value* amounts to £23 million per annum (ie annual deposit costs plus operation costs) (Table 6).

³ THES reported average academic staff salary 2009-10 £46,998 plus on-costs at full economic costing is £105,746 per annum, or £64 per hour at 220 working days and 7.5 hours per day. Post-graduate students were costed in the same way on the basis of graduate salaries of £25,500 per annum, or £35 per hour.

Table 6: Investment and use value for non-school and under-graduate student users, per annum.
Source: Authors' analysis.

<i>Parameter</i>	<i>per annum</i>
Investment Value:	
ESDS costs (ie funding) (ESRC operational data)	£ 3,332,104
Depositor preparation and deposit costs (Depositor Survey Q8)	£ 19,918,365
• Data collections acquired and processed	750
• Deposit-weighted mean preparation and deposit costs	£ 26,570
Data creation costs (Depositor Survey Q5 and Q7)	£ 794,327,744
• New data collections acquired	453
• Deposit-weighted mean data creation costs	£ 1,754,774
Total investment value (excl. sunk data creation costs)	£ 23,250,469
Use Value:	
Total non-student user access costs (User Survey Q8, Q9 and Q10)	£ 4,048,931
• Non-school and under-graduate student users	18,098
• Use-weighted mean user access costs	£224
Estimated annual depositor preparation costs (Depositor Survey Q8)	£19,918,365
Total use value	£23,967,296
Upper bound value to users (proxy indicator):	
Non-school and under-graduate student users who could re-create (23 per cent)	4,163
Mean respondent frequency of access/download per annum	3.9
Data creation costs per use	£13,509
Upper bound total use value	£219,703,723

Use value includes ESDS user access and download costs, as well as the use of ESDS by depositors (ie preparation and deposit). Again costing is done on the basis of average annual academic and graduate salaries and full economic costing using the TRAC fEC method, thus ensuring that user-side Internet communications and data processing infrastructure costs are included. Multiplying the use-weighted mean of user access costs (£224 per annum) by the number of active registered non-under-graduate and school student users (18,098) suggests annual user access costs of £4 million for the survey user community (ie excluding school and under-graduate students). With depositor preparation and deposit costs of £20 million per annum, *use value* amounts to almost £24 million per annum (ie the sum of user and depositor costs) (Table 6). Of course, treating data creation costs as sunk costs and including depositors as users means that data preparation and deposit costs are a major element in both investment and use value.

Indicators of upper bound value to users include the cost of collection/creation multiplied by the number of additional uses, which suggests the upper bound of what the good or service is worth to the consumers as they may alternatively forego use. Indeed, for many of the data collections held by ESDS re-creation is simply not plausible (eg. large-scale international and government collections). Nevertheless, just to explore a possible upper bound use value: 548 user respondents said they could not have obtained the data they downloaded in any other way (Q11), of which 330 said they could not have re-created the data (Q12), so 218 (23 per cent of total respondents) perhaps could have re-created the data. If this were characteristic, and 23 per cent of all registered non-school and under-graduate student users were able to re-create the data and had done so instead of accessing ESDS during the last year at the weighted average of depositor reported data creation costs, less ESDS, access and deposit costs, it would have cost them around £219 million during the last year (ie cost of creation, minus the costs of access, deposit and hosting, times additional uses) or 4,163 users times £13,509 creation costs times 3.9additional uses) (Table 6). Of course, some of the accesses are repeat accesses to the same data, it is more likely that they could have re-created the smaller and cheaper to create data collections than the larger and more expensive data collections, and some would have foregone the data if faced with re-creation costs, so this is likely to be very much an upper bound value.

Contingent Value

The contingent value of a non-market good or service is the amount users are "willing to pay" for it and/or "willing to accept" in return for it. For a public good, the value is the sum of "willingnesses", as consumption is non-rivalrous (ie the same information can be consumed many times). The key difference between willingness to pay and willingness to accept is that the former is constrained by the person's ability to pay (typically, by disposable income), whereas the latter is not. Hence, willingness to pay directly measures the demand curve with a budgetary constraint and willingness to accept measures the demand curve without a budgetary constraint (British Library 2004).

In the case of ESDS, where many users expect institutional support through their library, willingness to accept might be the better indicator of the value of ESDS to them. Although, a number of respondents expressed a willingness to accept nothing because they believe that data should be freely available - not because they do not value it.

Where there is a bundle of different goods and services, these can be treated in the aggregate, or dis-aggregated and re-aggregated in a way that reflects the bundling and/or use (eg. multiplying the average willingness to pay expressed by users of specific types of ESDS data collections by uses of those data), thus weighting by the structure of the bundle and its use. This is the approach adopted herein (see Section 4.1).

In the user survey, respondents were asked to express their *willingness to pay* in terms of an annual (subscription) fee and on a pay-per-access (PPV) basis. Ignoring the protest answers that are typical of this technique and weighting as described above (Table 4), the use-weighted mean of the individual willingnesses to pay was £1,374 per annum. The use-weighted mean on a pay-per-access basis was £285, which when multiplied by the use-weighted mean frequency of access would amount to £1,435 per annum (ie very similar to, and tending to confirm, the annual willingness to pay). Multiplying these use-weighted means by the three-year annual average number of active registered non-school and under-

graduate student users (18,098) suggests a willingness to pay (WTP) of around £25 million per annum among the survey user community (ie excluding school and under-graduate students (Table 7).

Table 7: Contingent value for non-school and under-graduate student users, per annum

Parameter	Weighted mean and per annum
Non-school and under-graduate student users (three-year average)	18,098
Use-weighted mean individual willingness to pay (per annum) (User Survey Q21a)	£ 1,374
Total users' WTP (per annum), excluding school and under-graduate students	£ 24,872,220
Use-weighted mean individual willingness to pay (per access) (User Survey Q21b)	£ 285
Use-weighted mean frequency of access (per annum) (User Survey Q8)	5.0
Total users' WTP (per annum), excluding school and under-graduate students	£ 25,974,286
Use-weighted mean individual willingness to accept (excl. zero responses) (User Survey Q20)	£ 6,154
Total users' WTA (per annum), excluding school and under-graduate students	£ 111,372,492

Source: Authors' analysis.

Willingness to accept is not constrained by capacity to pay or by the expectation among academic respondents that the payment should/would be made at the institutional rather than individual level. The method depends on individual preferences, so institutional responses are not useful and were excluded. There was also the issue of some respondents expressing a willingness to accept nothing because they believe that the data should free, rather than it being of no value to them. Including the non-protest zero responses, the use-weighted mean of the individual willingness to accept was £5,333 per annum, and excluding the 23 non-protest zero responses £6,154 per annum (Table 7). Multiplying these means by the three-year annual average number of active registered non-school and under-graduate student users (18,098) suggests a willingness to accept of £111 million per annum among the survey user community (ie the amount that staff and post-graduate student users would accept in return for giving up all their rights to access ESDS data and services for one year).

However, there was one regular user of large government data collections who suggested he/she would be willing to accept a minimum of £1 million to give up all access to ESDS for a year. This answer was 10 times higher than the second highest answer, although there was nothing else to suggest that it was a protest answer. Indeed, the respondent also rated ESDS highly, with responses including "extremely important", lack of access would have a "severe impact", etc. It should also be noted that this respondent's answers to all other questions were unremarkable. On balance, therefore, we are inclined to include the response. However, if this response were excluded, then willingness to accept (WTA) would fall from £111 million per annum to around £81 million per annum.

Consumer Surplus

The welfare impact or *consumer surplus* for a non-market good or service with public good characteristics is indicated by the total willingness to pay minus the cost of obtaining or accessing the good or service. From the data above, that is a willingness to pay of £25 million per annum minus the cost of non-school and under-graduate user access of £4 million per annum (see Table 6), suggesting a net consumer surplus of £21 million per annum among the survey user community (ie excluding school and under-graduate student users).⁴

Net Economic Value

The *net economic value* of a good or service is indicated by the consumer surplus (ie the willingness to pay minus the users' cost of obtaining access = £21 million) minus the cost of supplying the good or service (ie ESDS funding of approximately £3 million per annum). As we did not survey non-users, we are not able to formally estimate the total economic value. Nevertheless, the implied net economic value of ESDS data and services to its non-school and under-graduate student user community is around £18 million per annum - more than 5-times the ESDS operational budget (ie a return on investment of 5.4 to 1). Hence, every £1 spent on ESDS realises £5.40 value.

⁴ Data limitations make it difficult to take account of any implied deadweight loss associated with funding ESDS through taxation (ie take account of the costs of the machinery of taxation and distribution), although a common assumption is that a tax-payer pound costs £1.25. That would suggest a net consumer surplus closer to £17 million per annum.

Box I Return on Investment for data, library and information services

While individual studies focus on different information services and content and use different methods and measures, it is possible to explore Return on Investment (RoI) findings to give a sense of how ESDS data and services compare:

- British Library (2004) concluded that: "The British Library generates value around 4.4 times the level of its public funding."
- King (2010) summarised findings relating to library services and concluded that: special libraries exhibit an RoI of 2.9 to 1, academic libraries 3.4 to 1 (for staff), and public libraries 5.8 to 1.
- Imholz et al. (2007) summarized a number of studies, finding, inter alia, that Ohio public libraries showed a RoI of 3.8 to 1 and the Carnegie Library in Pittsburgh 3 to 1.
- Houghton (2011) estimated the benefit/cost ratio of the Australian Bureau of Statistics making data and publications freely available online and using Creative Commons licensing at 5.3 to 1.

Potential Efficiency Impact

The user survey asked if respondents had any sense of the extent to which their use of data and services from ESDS had changed their *research efficiency*. Respondents were offered percentage scales following feedback during pilot testing of the questionnaire and there were a number of limitations and problems with the scales used for these questions (see Annex A for details). Moreover the responses seemed to the authors to be high, raising some doubt as to whether respondents had interpreted the question to refer: (i) to research on the specific project for which they had used the last data accessed through ESDS; (ii) to research using ESDS-sourced data; or (iii) all their research (which was the intended meaning). In view of this, we explore the impact of these alternatives.

Weighting the user survey data as described above (Table 4), the use-weighted mean response suggested a 46 per cent increase in research efficiency (Q17). For the purposes of estimation, we take all of the post-graduate students and half of the academic staff to be research active (on an FTE basis),⁵ and cost them at full economic costing using the TRAC fEC method based on average graduate and academic salaries, respectively:

- If 50 per cent of researchers' time is spent on research and data occupied a use-weighted average of 51 per cent of their research time (Q15), then a 46 per cent increase in efficiency would be worth around £85 million per annum. (ie 8,548 post-graduate students at £57,375 per annum full economic cost plus 2,234 academic staff at £105,746 per annum full economic cost, times 46 per cent efficiency impact, times 50 per cent time share, times 51 per cent time share spent working with data).
- If respondents had referred to the impacts on research specifically using ESDS data, then we can adjust this to the use-weighted mean 25 per cent of research time

⁵ Reflecting the approximate share of research active staff in higher education institutions.

reportedly spent with ESDS data (Q15), and it would be worth around £41 million per annum (ie as above, but with a 25 per cent time share spend working with ESDS data).

These estimates exclude the potential efficiency impacts on school and undergraduate students, independent and retired scholars, and staff in government, non-profit and commercial organizations. As such, they are likely to be very conservative (under)estimates of the efficiency impacts of ESDS on its direct user community, let alone on the wider community.

To those respondents whose roles included teaching, we posed the same question about the impacts of their use of ESDS data and services on their *teaching efficiency*. Weighting user survey responses as described above (Table 4), the use-weighted mean response suggested a 23 per cent increase in teaching efficiency (Q19). If we assume that half of the active registered ESDS user academic staff are teaching (on an FTE basis), cost them on the basis of average academic salaries with full economic costing using the TRAC fEC method, and assume that they spend 50 per cent of their time teaching, then the implied teaching efficiency impacts of their use of ESDS data and services would be worth around £27 million per annum (ie 2,234 academic staff at £105,746 per annum full cost, times 50 per cent time share, times 23 per cent efficiency impact).

This is, of course, in addition to the research efficiency impacts noted above, suggesting that the *direct efficiency impacts of ESDS data and services* may be in the range of £68 million to £112 million per annum among the immediate user community - excluding the potential efficiency impacts on school and under-graduate students, independent and retired scholars, staff in government, non-profit and commercial organizations, and unregistered users of ESDS website material. On this basis, one could speculate that the direct efficiency impacts of ESDS might be worth at least £100 million per annum or more to the wider user community.

Return on Investment in Social Science Data Infrastructure

There is an extensive literature on returns to R&D and a range of studies have explored possible impacts of increases in accessibility and/or efficiency on returns to research expenditure (Houghton et al. 2009; Houghton and Sheehan 2009; Houghton, Rasmussen and Sheehan 2010). We had originally hoped that this approach might be possible for this study. However, we found that relatively little is known about the economic and social returns to investment in social science research or the total level of social science research funding in the UK.⁶ Moreover, it is likely that a good deal of social science research is not directly funded (eg. being done as a part of organizational activities and without specific project funding) and/or not funded at its full economic cost. As a result of these limitations it has not been possible to explore estimates of possible impacts of increases in accessibility and/or efficiency on returns to research expenditure.

⁶ We examined a range of sources including Higher Education Statistics Agency (HESA) and reports from other social science funders. However it was quickly apparent that there would be no consistency with ESRC definitions of social science disciplines.

Nevertheless, it is possible to explore some scenarios relating to the potential impacts of ESDS research data infrastructure on returns to investment in the data. There are a number of data elements required for such an analysis, including: annual investment in the research data, services and infrastructure; average returns to that investment; and the increase in access and use resulting from it. The user and depositor surveys and ESDS operational data and annual reports provide a foundation for estimates:

- Multiplying the acquisition-weighted mean of the reported annual data creation costs (£1,754,774) by the average annual number of new data collections acquired and processed for online delivery (453), suggests annual data creation costs of £794 million. Acquisition-weighted mean depositor preparation and deposit costs amount to around £20 million per annum, and ESDS operation costs to some £3.3 million per annum, suggesting total data infrastructure and services investment costs of around £818 million per annum (ie the sum of these elements).
- There is an extensive literature on returns to R&D, which, while varied, suggests that returns are high - typically in the region of 20 per cent to 60 per cent per annum (Bernstein and Nadiri 1991; Griliches 1995; Industry Commission 1995; Salter and Martin 2001; Scott et al. 2002; Dowrick 2003; Shanks and Zheng 2006; Martin and Tang 2007; Sveikauskas 2007; Hall et al. 2009). Much of this literature relates to the natural sciences and one might imagine that average returns to social science may be lower. Nevertheless, there are a number of points to consider:
 - Social science plays a key role in innovation and the application and commercialization of science (eg. pharmaceuticals and prescription pharmaceutical benefits schemes);
 - Social science provides key concepts and tools that underpin decision making (eg. the system of national accounts); and
 - Social science makes a direct contribution to allocative efficiency (eg. in the design of radio-frequency spectrum auctions).

Hence, it is reasonable to explore a range of returns at and below the lower bound average returns to R&D expenditure reported in the literature (eg. 20 per cent and 5 per cent). Returns to R&D are recurring annual gains to a single year's investment and should be discounted over the useful life of the data.

- The other issue is what impacts ESDS data and services have on access and use of the data hosted and delivered. Some 548 respondents to the user survey indicated that they could not have obtained the data in any other way if ESDS had not existed (Q11). While all of these represent additional use, responses to question 12 suggested that, of these, 330 could not have created the data themselves. So we assume that the remainder (23 per cent of respondents) could have re-created the data and that this is typical of ESDS users, suggesting that as much as 23 per cent of ESDS use may be additional use that could have saved data (re)creation costs.

We proceed on the basis of these data, first estimating the annual increase in return on investment. At the lower-bound average 20 per cent return on the £818 million investment, if 23 per cent of ESDS use is additional use, the implied increase in returns would be £38

million per annum, and at the lower 5 per cent return the implied increase in returns would be £9.4 million per annum (ie £818 million invested for 56,777 uses at 20 per cent gives a return per use of £2,880, then 23 per cent additional uses (13,059) at £2,880 is £38 million. A 5 per cent return results in per use return of £720, which in the above calculation results in an increase in returns of £9.4 million).

As these increases in returns are recurring during the useful life of the data we use a simple Perpetual Inventory Method to estimate the overall value of the impacts. Drawing on preliminary work on the UK R&D Satellite Account (Evans et al. 2008) we depreciate publicly-funded research data at 5 per cent per annum, and following the lead of the US R&D Satellite Account (Sveikauskas 2007) we set the useful life of the data/knowledge created each year at 30 years and apply a lag of three years between data creation and preparation expenditure and the first realisation of returns. Applying a 3.5 per cent discount rate to estimate net present value (HM Treasury 2011), we model the recurring returns outlined above as follows:

- At the lower-bound average 20 per cent return on data investment, if 23 per cent of ESDS use is additional use the implied overall increase in returns on one-year's data infrastructure investment would be £233 million (NPV); and
- At the lower 5 per cent average return on data investment the implied overall increase in returns would be £58 million (NPV).

Hence, we are likely to be looking at an increase in return on annual investment in the data and ESDS research data infrastructure services of £58 million to as much as £233 million over 30 years (Net Present Value) due to the additional use facilitated by ESDS. Given non-sunk data infrastructure costs of around £23 million per annum (see Section 4.3), this suggests a 2.5-fold to 10-fold return on investment.

Summary of Economic Analysis

The various approaches explored give a variety of partial pictures or different pieces of the puzzle, with some overlaps and some gaps. We start with the most immediate and direct measures of value that are likely to represent lower bound estimates of the value of ESDS data and services to its active registered non-school and under-graduate student user community, and move outwards to more uncertain estimates of the wider benefits (Figure 4). What is notable is that the various approaches produce comparable estimates and exhibit the kind of progression from direct minimum values to the survey community to higher values to the wider user and research communities that one would expect to see, enhancing confidence in the results.

Looking at the lower bound, the investment and use values of ESDS to this sub-set of users (ie the survey user community) both appear to be around £24 million per annum. Of course, treating data creation costs as sunk costs and depositors as users means that data preparation and deposit costs are a major element in both investment and use value.

These investment and use values are reflected in the contingent valuations, where users' willingness to pay per annum or on a pay-per-access basis both amount to around £25

million per annum. Hence, the consumer surplus is around £21 million per annum (WTP net of user access costs), and the net economic value around £18 million per annum (consumer surplus net of ESDS operational costs) - more than five-times the ESDS operational budget.

As is typically the case in contingent valuation techniques, what users are willing to accept in return for foregoing access to ESDS data and services is higher, at £111 million per annum, than what they are willing to pay - which is constrained by what they can afford. This is a significant constraint, as there are a large number of post-graduate student users with very limited capacity to pay, as well as many social science researchers with limited funding among the survey community.

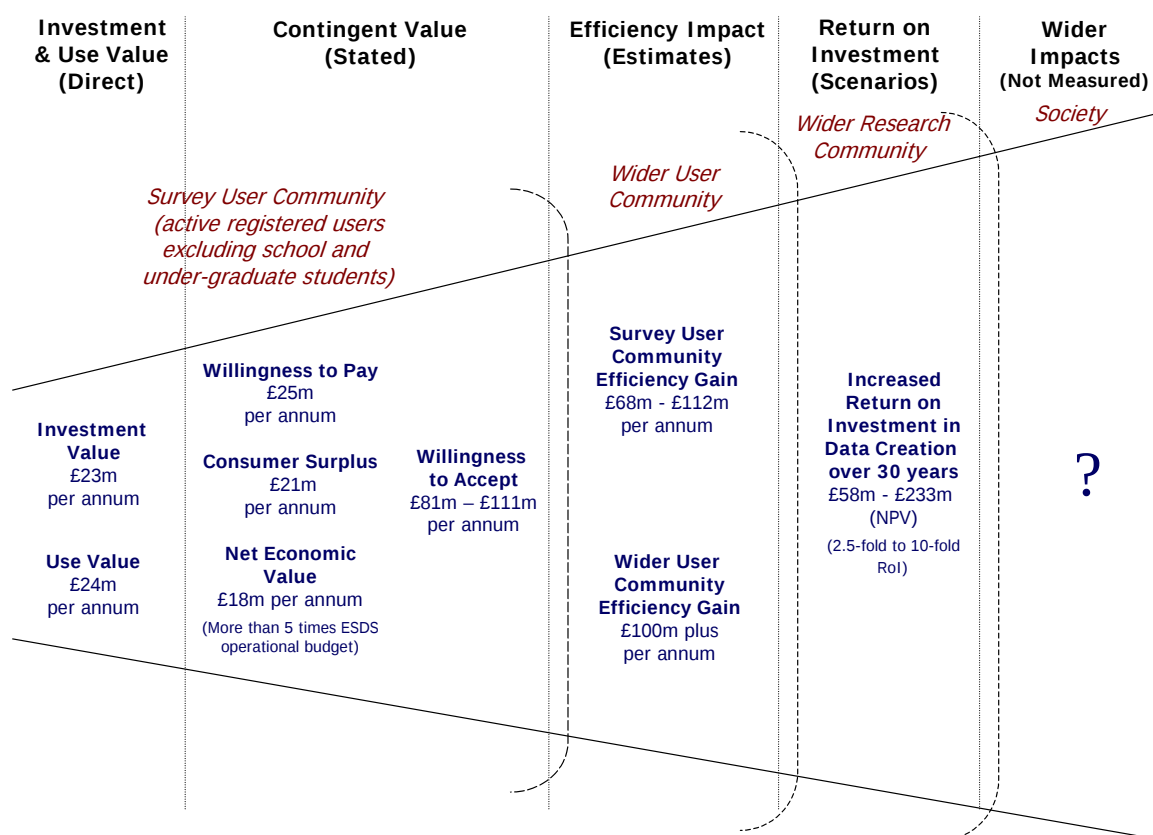


Figure 4: The value and impacts of ESDS research data infrastructure.
Source: Authors' analysis.

Moving to the wider impacts and benefits of ESDS research data infrastructure, we estimate research and teaching efficiency impacts for ESDS's active registered user community (excluding school and under-graduate students) at some £68 million to £112 million per annum. This might translate to at least £100 million per annum or more for the wider user community.

Exploring scenarios based on survey and reported data suggests that ESDS research data infrastructure services facilitate an increase in the return on annual investment in the data and ESDS research data infrastructure services of £58 million to as much as £233 million over 30 years (Net Present Value). Given non-sunk data infrastructure costs of around £23 million per annum, this suggests a 2.5-fold to 10-fold return on investment.

Such a return on investment is at the higher end of RoI for library and information services, national statistics and other information and data services reported over the last decade (See Box 1, above), suggesting that ESDS data and services provide comparable though slightly higher value for money than academic, special and public library services.

Illustrating Wider Benefits

The terms of reference for the study include presenting the results of the economic assessment within the context of a broader qualitative analysis of the benefits of the ESDS. This is important if the wider social benefits of ESDS which cannot be captured in the economic impact analysis are to be illustrated. It is also important in understanding some of the finer details of the activities and attributes of the ESDS that contribute to its economic impact.

We used the KRDS Benefits Framework (see section 1.4) as an underlying conceptual framework for approaching and illustrating the benefits arising more generally from the ESDS. To do this we have drawn on the desk research, interviews, relevant parts of the user and depositor surveys, and their observations on the benefits of ESDS. Relevant findings on wider benefits from these sources are described in sections 5.1 - 5.4 and then finally synthesised and presented in a KRDS Benefits Framework in section 5.5.

Benefits Identified in Desk Research

There are still only a relatively small number of socio-economic studies focussing specifically on the impact or wider benefits of data services or research data infrastructure. The UKDA and/or ESDS have featured directly in two of these: the KRDS projects funded by JISC; and the RIN/JISC funded study on *Data centres: their use, value and impact* (Simmonds et al, 2011).

For assessing benefits, KRDS created a “benefits taxonomy” tool and illustrated this with two benefits case studies one of which was the UK Data Archive (Beagrie et al 2010). This work was extended and tested further in the Digital Preservation Benefits Analysis Tools project and the UKDA case study implemented in a more developed form by the UK Data Archive in a KRDS Benefits Value-Chain Worksheet (Charles Beagrie 2011, Woollard 2011). The KRDS Benefits Framework has also been implemented subsequently by ESDS as a handout summarising benefits from ESDS in its recent mid-term review (ESDS unpublished). These sources present a wider range of individual benefits identified by UKDA staff for ESDS, arranged under the KRDS dimension of:

- direct and indirect benefits;
- short and long-term benefits; and
- private and public benefits.

We have integrated this work on ESDS benefits using the KRDS methodology to provide the main elements of the benefits summary presented in section 5.5.

The Research Information Network (RIN) and JISC Data Centres report was an analysis of the usage and impact of five UK based research data centres, including ESDS. The report found the social and, particularly, economic impacts of data-centres proved more difficult to identify, principally because neither researchers nor data service infrastructure are easily able to keep track of how their outputs (research and data) are used post-delivery. The report included a number of statements on generic benefits in its survey and response rates on level of agreement from ESDS users. Those with highest levels of agreement (>60 per cent) included:

- It has reduced the time for data acquisition/processing;
- It has improved the efficiency of research;
- It has helped reduce the financial cost of data acquisition/processing;
- It has reduced duplication of effort ie unnecessary re-creation of data;
- It has enabled research to go ahead that otherwise might not have done.

Benefits identified in the Interviews

Twenty-five structured individual or group interviews involving thirty-three individuals were carried out with key stakeholders. Findings from the interviews helped to inform development of the survey questionnaires and the analysis. In addition they provide in-depth insights for the study or “mini-case studies” that can illustrate some of the findings in the survey and economic analysis. These insights have been selected from interview notes and grouped under a set of relevant headings in Appendix B.

Discussion of wider benefits from ESDS described in the interviews are summarised here and illustrated by a small number of extracts from interview notes.

Value for research

‘It would be impossible to do his job without ESDS – he made the point (quite forcefully) that ESDS was a monopoly provider of the data collections he uses, and thus his whole research output was reliant on ESDS (minus the data which he himself deposits).’

Academic research user and depositor (Economics/Social Policy)

‘Datasets are cleaner and better with ESDS. There is good quality control. It is also a one-stop-shop. All this saves time and enables more concentration on actual research.’

Academic research and teaching user (Economics)

Interviewees highlighted the following major benefits for researchers:

- time savings from having a single point of access plus a usable and familiar interface;
- high quality data;
- large time and cost savings in acquisition of data and obtaining permissions for use;
- reducing delays in research by increasing the speed of access to data; and
- no viable alternative supplier for data collections held uniquely by ESDS.

Value for teaching and training

‘... he has 300 students on an economics introductory course, and is in charge of Under-Graduate and some Masters dissertations. The data they tend to access is from IMF, OECD, ONS statistics etc...He felt that the time taken to access data from other sources would be detrimental to his teaching in all these areas. Also, the documentation in ESDS is very good, and adds to the novelty of his teaching ... He estimated that the time taken to access what they do on ESDS was less than 10 per cent of what it would otherwise have been...it is

benefitting master's students ... and will have positive impacts in the future. This is because their understanding of the issues inherent in economics is now more connected with data, and so the approaches are more data driven. This will have important impacts when these students begin their professional work – being able to approach and solve problems on evidence...'

Academic teaching user (Economics)

Teaching staff reported that after mediation (eg. simplification and removal of some variables) a downloaded data collection could be reused many times for class work, allowing many people to benefit. Not having to register with and login to multiple providers has improved the efficiency and quality of teaching and assessment. It has allowed more time for students to learn how to understand the structure of data, what is available and how to interrogate it, giving them transferable skills; their approach to issues is more data driven and empirical. They valued:

- the quality of the documentation, teaching sets, search tools and contextual information;
- the quality of the data;
- the international range of data collections available from a single point of access;
- ease of access, including licensing issues and simple login process for the students themselves;
- the intrinsic value of the data collections as part of the history of social science research; and
- the transferable skills in data analysis it fostered in the students.

Recommendation 1: Review how ESDS usage statistics reflect teaching use. Interview evidence shows substantial local use per individual download.

Value for service provision

'...ESDS are better at publicising it than his group would be, and they are also better at data security – they can invest more. He guessed that usage would fall if ESDS wasn't there by maybe 20-25 per cent...By depositing with ESDS there is heavier usage of the data, which helps publicise his work...More people will find the data if it is released through ESDS, because it is a single point for such survey data ...In other words we would have to do more promotional work as well as direct work on data dissemination and it is less easy for us to target the whole of the audience.'

Academic research depositor (Economics and Sociology - Longitudinal Data)

'ESDS is basically an aggregator of data which it provides on to its clients. This arrangement is good for us as it means that we can negotiate and work with only one customer but reach several clients...The main advantage is that data is not lost. Much research funding requires researchers to deposit data with ESDS. The UK is doing a good job in respect of data

gathering and archiving. There is no equivalent to ESDS in other countries, so lots of data gets lost or obscured in universities [elsewhere]...'

Non-academic depositor (International Data)

Interviewees valued:

- the higher usage of their data within ESDS where it is more likely to be found;
- ESDS providing better publicity and data security than they could do individually;
- ESDS removing the burden of interacting with individual users from them;
- ESDS dealing with user registration, licensing and correspondence as well as curation and access;
- ESDS as an aggregator of data providing one point access to many UK clients.

Value as a consumer champion

'In a government department people are concentrated on the demands of the department, not helping external parties to understand the data. At ESDS their function is to help people use data which encompasses customer service skills.'

Research user (UK Government)

'He also noted that the relationship between ESDS and data providers such as the ONS; government etc. helped make things clearer as to needs of academic community – another advantage of having the ESDS.'

Academic research user and depositor (Sociology)

Interviewees valued:

- the ESDS focus on making it easy for customers to use the data;
- the ESDS articulating the needs of academic users to data providers; and
- The customer service ethos of ESDS staff.

Value as a centre of excellence

ESDS staff were also concerned we did not overlook the impact from their status in the profession in terms of professional visitors to the archive and ESDS seeking to learn how to set-up a data archive, leap-frogging early mistakes/set-backs etc. The professional network that ESDS maintains internationally helps to develop international infrastructure and access for UK users. Knowledge transfer to users from this work is also done via their newsletters etc. ESDS receives significant benefits from University of Essex and UK Data Archive R&D project income out with the core ESRC grant. This research also has an impact.

Innovation in practice

‘Can’t put a figure on it but the change in use has been quite rapid. There’s a strong sense of innovation. The change is accelerated by cutbacks in the rest of the public sector ... The speed of deposit from major data suppliers to ESDS is getting much faster- change to the good. Now depositors feel incentive to deposit at earlier point to stop people ringing them up.’

Research user (UK Government)

‘She gave an opinion on how much she felt research has changed. Data is readily accessible now, and with strategic direction from ESRC, people more willing to use and learn about secondary qualitative data – there has been a huge change.’

Academic research depositor (Sociology and Social Policy - Qualitative Data)

Interviewees suggested ESDS is fostering changes in research practice by:

- making data more readily available to researchers; and
- working with data providers so that they are able to deposit much earlier.

Promoting quantitative methods

‘Another aspect is having a place academics can go and conquer their fear of quantitative analysis. So few people believe it is possible to use the data, ESDS makes it available in a psychological sense for students to learn quantitative methods from the data and documentation.’

Research user (UK Government)

‘The problem is that so few people have the skills to use the data collections (although they are easy to access and retrieve data from)... they need training.’

Academic research user (Medical Sociology)

Interviewees suggested ESDS is helping to break down psychological barriers towards applying quantitative methods and provides resources and training for understanding the correct application of quantitative methods.

Recommendation 2: The absence of sufficient skills in quantitative methods remains a significant barrier to maximising use of ESDS and other data. ESRC and ESDS should continue to encourage re-use of data, use of quantitative methods, and use of ESDS data collections in teaching.

Value to Commercial and Government Users

‘He felt that the nature of the resources available via ESDS is not what many commercial organisations require... a commercial organisation is far more likely to want quick information, say on crime etc. and that as such would be unlikely to spend time going

through the process of obtaining the source data and analysing it themselves. They require a digested version which summarises findings ...’

Non-user (Commercial Sector)

‘The crucial thing is time: the typical deadline is a day or two to come up with answers – for that reason they tend to have to work from existing analyses.’

Research user (UK Government)

Although not insubstantial, commercial and government users of ESDS are significantly lower than users from the academic sector. Interviewees suggested a number of reasons for this and changes that might address it. Time pressures and deadlines for both government and commercial customers can limit their use of the raw data held by ESDS. Provision of more digests and simpler summaries of the findings from the data collection might promote greater usage by these sectors. Commercial users are also often looking for more current data.

Recommendation 3: If ESRC wishes to extend use of ESDS by the commercial sector or Government it could provide more digests and simpler summaries of the findings from the data collection that might promote greater usage by a wider range of users.

Defining Research Data Service Infrastructure

During interviews, ESDS staff and users emphasised that ESDS (or research data service infrastructure generally) is not solely about data: there is a broad spectrum of value-added activities that provide wider benefits to ESDS stakeholders.

It is worth considering in the study therefore how research data service infrastructure might be defined and how its components contribute to economic impact and wider benefits. On the basis of the interviews, we believe the broad four-part division of intellectual capital/human capital/ organisational capital/relationship capital proposed (Hunter 2006) for intangible assets covered by digital preservation or digital curation can provide, with some adaptation, a suitable definition for ESDS or ESRC funded research data service infrastructure. The components of that research data service infrastructure would be:

- Research data packages [Intellectual Capital]
- Staff knowledge and skills [Human Capital]
- Technical and organisational environment [Organisational Capital].
- Disciplinary and professional networks [Relationship Capital].

Although the data held by the ESDS has a social science focus, some interviewees also emphasised how ESDS, in its roles as a centre of excellence and a “consumer champion” often has benefits extending to other disciplines. The benefits from ESDS staff knowledge and skills and its disciplinary and professional networks in fostering innovation in research

data management more generally is likely to be of particular interest to JISC as a co-funder of the ESDS.

Recommendation 4: The JISC as a co-funder of the ESDS alongside the ESRC may wish to note our findings on the economic impact of ESDS and the wider benefits identified from its fostering of best practice and innovation in research data management.

Benefits Identified in the User Survey

Results from the user survey are presented in full in Appendix A. Questions in the survey were primarily focussed on obtaining data for the economic analysis, but a number focussed on exploring broader benefits and impacts. The responses largely echo findings from the interviews. They provide supporting quantitative data from a large and broad range of respondents that enhances findings on benefits reported from the interviews.

Question 7 of the user survey asked what impact not having access to ESDS data and services would have on their work. The majority of respondents (61 per cent) suggested that it would have a major (33 per cent) or severe (28 per cent) impact. The various types of respondents gave slightly different answers, with academics suggesting that not having access would have a major to severe impact, government and non-profit organization users a severe to moderate impact, and corporate and other users a moderate impact. From the academics, staff suggested that they would be affected slightly more than students. Users of big data collections (ie Longitudinal Survey Data, Multi-nation databanks and International survey data, and Large-scale government surveys) suggested that they would be much more affected than users of small data collections (ie qualitative and mixed methods data, and other survey data).

To explore possible valuations through counterfactual means and establish the extent to which ESDS use is additional use, respondents were asked if they could have obtained the data they used in another way had ESDS not existed (Q11). Some 70 per cent said they could not have obtained the data in any other way. The question elicited 210 comments (27 per cent of responses, but more than 90 per cent of those saying they could have obtained the data in another way), of which the majority commented on the extra time and costs that would be involved in obtaining the data elsewhere, a lower confidence in its reliability and uncertainty as to the completeness of their search.

Question 12 asked if they thought they had saved time or money in a number of areas of activity as a result of using ESDS data and/or services. More than 80 per cent of respondents nominated the ability to find data from a single point of access as the biggest area of saving, followed by the quality of data (ie the level of preparation, validation and documentation associated with it) (66 per cent), and the fact that it was beyond their ability to create or collect the data for themselves (58 per cent). Note that the responses to this question reflect primarily the researcher perspective of benefits to them personally and their own research. It will not necessarily reflect relative benefits to other stakeholders, such as funders or institutions, from activities, such as reduced hosting or licensing costs.

Question 13 asked to what extent they benefited from ESDS in a number of ways. It was clear that users saw methods and documentation as a major benefit, followed by user support (eg. guides, helpdesk, etc.) and best practices (eg. case studies and standards).

Benefits Identified in the Depositor Survey

Results from the depositor survey are presented in full in Appendix A. Questions in the survey were primarily focussed on obtaining data for the economic analysis, but a number focussed on exploring broader benefits and impacts. The responses largely echo findings from the interviews. They provide supporting quantitative data from a large and broad range of respondents that enhances findings on benefits reported from the interviews by depositors.

Question 9 in the depositors survey asked depositors to state the level of benefits they receive from depositing their data with ESDS. Most of the depositors reported very high to medium benefit from the following features: data preserved long-term (78 per cent), dissemination targeted to academic community (83 per cent), wider exposure and data more discoverable (69 per cent), single deposit and licence provides access to many users (67 per cent), and fulfilling grant obligations (74 per cent). They reported benefiting somewhat less from fulfilling organizational mandates (56 per cent).

When asked what impact it would have on their work if they could not deposit/provide data to ESDS (Q10), half of the depositor respondents reported that their work would be severely to moderately affected, 22 per cent reported a slight impact and a quarter of the respondents stated that the absence of ESDS would not harm their work at all. Depositors of big data collections suggested that they would be more affected than depositors of small data collections.

Summary of Benefits Analysis

The outcomes from a KRDS benefits analysis can be presented in a simple table structured around the dimensions of the KRDS Framework. . It provides a simple visual summary of the key benefits from a research data service or activity in an easily digestible form.

We have integrated previous KRDS based work on ESDS benefits to provide the main elements of the benefits summary for ESDS presented below. We have also reviewed and where necessary added the benefits of ESDS mentioned above in other desk research, the interviews, and relevant parts of the user and depositor surveys to enhance the summary.

Box 2: Summary of ESDS Benefits arranged in a KRDS Benefits Framework

Benefits Summary for ESDS	
Direct Benefits	Indirect Benefits (Costs Avoided)
Time and resource savings for researchers and teachers	No re-creation of data -Trusted Digital Repository status eliminates re-ingest costs
Verification of research through increased data citation thanks to relevant citation information and tools	Lower future archiving costs increase likelihood of data being available, earlier in the lifecycle

Access to data provides new research opportunities by increasing use of data within collections thanks to proper Collections Development Policy Re-purposing and re-use of data maximises the value of data holdings when the data is available for third parties Increasing research productivity Assisting in provision of a skills base Customer service ethos of ESDS staff	Re-purposing data for new audiences Use by new audiences Protecting returns on earlier investments Reduces potential duplication of effort
Near Term Benefits	Long-Term Benefits
Value to current researcher and students Single point of access Increasing speed of access to data Ease of access for researchers and students Quality of documentation and contextual information, and resources for teaching No data lost from post-doc turnover thanks to a reliable preservation system Secure storage for data intensive research Availability of data underpinning journal articles Research data integrity since data is of high quality	Data preserved for the long-term Secures value of high quality data for future researchers and students Value added over time as collection grows and develops critical mass Input for future research by maximising use across data user community (ie including students) Impact on wider profession as a centre of excellence Promoting quantitative methods and skills Fostering innovation in research and data management practice
Private Benefits	Public Benefits
Benefits to sponsor of research Benefits to sponsor of data service provider Benefits to researcher Fulfil grant obligations Increased visibility/citation Aggregator of data for data providers – providing them with one point of access to UK customers Higher usage (and profile) of their data for depositors Removes user burden from depositors	Source of high-quality and often unique data Motivating new research Enables research that otherwise could not be undertaken Research integrity since others can check the outcomes of research Service targeted at academic community and supports their needs Fostering transferable skills in data analysis Articulating user needs to data providers

The Case Studies

The evidence from the ESRC's Evaluation Committee's programme of work demonstrates the ESRC's significant contribution to economic and social well-being, and the Committee has recently been developing new approaches to valuing that contribution.

The key aspect of the ESRC's approach is that while economic evaluation is important and necessary, it needs to be put into context. This context involves broader, more qualitative assessments that capture the full range of benefits from social science research and the research data infrastructure that supports it. As a result the terms of reference for this study included undertaking case studies to assess the policy and practice impact of research based on data accessed via the ESDS (including an analysis of why and how impact has been generated).

Three potential case studies were identified jointly by the research team and the ESRC, based on the obvious importance of the topic, impact, and data collections used. Regarding the latter, case studies were chosen to represent a spread of use of the main ESDS data collections: one using Government data (knife crime); one International (climate control) and one using Longitudinal data (obesity) respectively.

The case studies were reported in terms of the research undertaken, impacts the research occasioned, and pathways and processes which facilitated the impacts elicited. The first two of these elements were undertaken by desk research. We then interviewed key researchers involved in each research project or area to obtain information about the pathways to impact, although interviewees also gave us extra information on their research and its impacts too. In fact it was often difficult to disentangle impacts from the pathways.

The interviews were informal and open. Interviewees were invited to talk through the three aspects to be addressed in the case study but asked to elucidate in particular the pathways to the impacts. Interviewees were then each sent a draft version of the case study for approval.

Case Study 1: E3MG: An Energy-Environment-Economy (E3) Model at the Global Level

Researcher: Terry Barker, Department of Applied Economics, University of Cambridge and Director of the Cambridge Centre for Climate Change Mitigation Research, Cambridge Econometrics and colleagues

Data collections used: The OECD's Structural Analysis, World Bank's World Development Indicators, Eurostat, and International Energy Agency data collections (ESDS International)

The research

Cambridge University economist Terry Barker and colleagues explored the costs of reducing carbon dioxide emissions. The researchers constructed a sophisticated 'econometric' model (the 'E3MG', standing for 'Energy Environment Economic Model at Global level) gathered from international governmental organisation (IGO) databanks,

(Barker et al, 2010) and used it to find support for the argument that there are economic benefits to stabilising carbon dioxide concentrations and that such benefits increase the higher the target concentration used (See ESDS, 2006)

A key aspect of Barker's work is that the models allow an examination of how the economic system changes over time, providing year-by-year projections of these changes into the future without assuming the existence of economic equilibrium. As a result, E3MG and the other models that the group develop do not fall into the category of General Equilibrium Models, which show only how real resources are re-allocated between economic activities after the economy has 'equilibrated' following introduction of a policy or a 'shock' to the system such as a price change in oil. This allows a modelling of the uptake of new technologies such as low carbon energy sources, providing a better understanding of the time it will take for the economy to accommodate these technologies and reduce the national emissions of greenhouse gases and other pollutants (Cambridge Centre for Climate Change Mitigation Research, 2011).

In an interview for this report, Dr Barker suggested that, for him, the main advantage of ESDS was that the data is grouped together and accessed from one source. This makes the efficiency much greater. This, in turn leads to better quality research, because time savings mean that one can concentrate more on thinking about the research questions than on how to get to certain data required. An ESDS report points out also that 'constructing an empirical test of this magnitude, and using a multitude of data collections would be prohibitively expensive without the UK-wide academic redistribution agreements negotiated between the ESRC and the IGOs. It provides an example of how ESDS services deliver savings to the UK academic community as a whole' (ESDS 2006: p19).

The impact

The model is being used to assess the impacts of binding global climate agreements made at the Copenhagen UN Climate Change conference in December 2009 and was used in the paper presented by Dr Terry Barker at the Avoiding Dangerous Climate Change international symposium in 2005 (see Barker et al, 2006). A major outcome from that symposium was the contribution to the discussion of the economic benefits to carbon emission stabilization. The change in limit on the 'safe' atmospheric concentrations of greenhouse gases that followed this event also fed into a subsequent change in greenhouse gas emissions targets in the UK Climate Change Bill (which became an Act in Nov 2008), from 60 per cent to 80 per cent below 1990 levels by 2050 (ESDS, 2006: p19). Clearly, the change in 'targets' does not necessarily imply that real change is guaranteed. However, the annual targets in the Act equate to falls of 23 per cent to 2012. In fact, excluding emissions trading, emissions on a carbon budgets basis were provisionally estimated to be 577.9 MtCO₂e in 2010, which is 26.2 per cent below the baseline and so ahead of target. 'Accounting for the effect of emissions trading increases the UK's emissions in 2010 on a carbon budgets basis to 585.6 MtCO₂e'. This is 25.2 per cent below the baseline (DECC 2012).

Pathways to impact

The key to creating the impact described above has been the dissemination of the results of Dr Barker's E3MG model by the Intergovernmental Panel on Climate Change (IPCC). This

organisation is the leading international scientific body for the assessment of climate change, established by the United Nations and the World Meteorological Organization. It reviews and assesses the most recent scientific, technical and socio-economic information produced worldwide relevant to the understanding of climate change. Governments participate in the review process and the panel's plenary sessions, where main decisions about the IPCC work programme are taken and reports are accepted, adopted and approved. The E3MG model has been used in a series of projects and papers reported in the IPCC's 4th assessment report on the mitigation of climate change (Metz et al, 2007). The model is featured in Table 11.15 on p654. The report was submitted to government for consideration. Following this, the model went into comparing the macro-economic costs of mitigation to those of 2030. The study showed that there could be a benefit in mitigation rather than a cost.

Case Study 2: Knife Crime: A Review of Evidence and Policy

Researcher: Chris Eades, Roger Grimshaw, Arianna Silvestri, Enver Solomon, Centre for Crime and Justice Studies, formerly based at Kings College London

Data collections used: The British Crime Survey; police recorded crime figures; and, for children and young people, the Offending, Crime and Justice Survey and the Youth Surveys commissioned by the Youth Justice Board and carried out by MORI (ESDS Government).

The research

The research, 'Knife crime: a review of evidence and policy' was carried out and published in 2007 by the independent Centre for Crime and Justice Studies (CCJS). This report reviewed data from several sources to examine the nature and extent of knife crime, the reasons why knives were carried, and who might be most affected by the problem of knife crime. It also examined potentially effective ways to reduce knife crime.

According to official statistics the number of violent crimes involving knives in England and Wales has remained stable in recent years. Within particular offence categories there is no substantial evidence of a significant change in the proportions of knife use. As a proportion of all homicides the use of sharp instruments has fallen over the past decade, accounting for less than thirty per cent of homicides in 2005/2006 compared to nearly forty percent in 1995.

There a paucity of literature on the motivations for knife carrying (the report called for more research into this) but children who have been a victim of crime are more likely to carry knives. There appears to be a link between knife carrying and whether or not young people feel safe from crime and victimisation. It appears that young people, those living in poor areas and members of black and minority ethnic communities are more likely to be the victims of knife offences.

The impacts

The main impacts of the research have been in stimulating debate, notably in the House of Commons (HoC), but also in the media. Regarding the former, it has been cited in a HoC briefing paper (Berman, 2011), and the latter, in both the national and local press. The Daily Telegraph (Steel, 2007) quotes from the report in an article about the vagaries of crime statistics, and The Sunderland Echo carried an article about a father of a murdered victim of

knife crime, who considers the (first edition of the) report to be 'too soft' on offenders. 'Think Tanks' and charities have also mentioned the report. Policy Exchange, for example, an independent, 'non-partisan' educational charity working with academics and policy makers, published a paper on gang and knife crime (Golding and McClory 2008) in which it was cited extensively. An article in the Times newspaper (Leppard, 2007) was also mentioned in the paper which quoted one of the Knife Crime authors, Enver Solomon.

The research has been extremely widely circulated and discussed. It is, for example, accessible via the Local Government (LG) Improvement and Development website (<http://www.idea.gov.uk/idk/core/page.do?pageld=1>), which supports innovation in local government and works with local authorities and their partners to develop and share good practice. It is also cited extensively on the 'Policy Hub' website (see: http://www.nationalschool.gov.uk/policyhub/news_item/crime_knife07.asp) This is a resource developed by the Government Social Research Unit which aims to improve the way public policy is shaped and delivered.

One of the report's authors, Roger Grimshaw, pointed out in written communication with the research team that the police were familiar with the report. The Police Federation representative who read the report was David Pellatt. The national police spokesman on knife crime was DAC Alf Hitchcock, who was 'certainly aware of the report'. Roger met both when invited to speak to the Police Federation conference 12.05.09. He feels that these contacts suggested that the police took the findings very seriously.

Impact has also been in the form of more research interest and enquiry. In 2008 the Centre was asked by the children's commissioner for England and Wales to undertake a literature review around the factors behind knife crime and the methods used to reduce it. The organisation was also approached to do further work, for example by the Street Weapons Commission (Channel 4), chaired by Cherie Booth, and asked to collate evidence about street weapons, knives and guns in five cities the Commission was visiting. It is worth noting, finally, that downloads to the report from the CCJS site have been heavy – nearly 7,500 (7453) as of 28.11.11, and the fact that other major databases have been storing or referring to the report must add to the chances of it being downloaded.

It is worth emphasising to conclude this case study that the impacts that the team have been able to elicit – from the literature and from the researchers using ESDS to undertake their studies – centre on stimulating debate and interest in further research. Thus, it is not possible to say that the research has led to a change in policy, much less in a reduction in knife crime. Indeed, figures produced as the final draft of this report was being prepared show that 'knifepoint' robberies rose by 10 per cent in the year to September 2011⁷.

Pathways to impact

Roger Grimshaw feels that securing maximum impact for research outputs hinges on the wide dissemination of results. There is a need to invest in more than producing a report, such as engaging with the press and submission to various non-academic bodies. For the

⁷ BBC Online (2012) Knifepoint robberies rise by 10 per cent <http://www.bbc.co.uk/news/uk-16626558>

knife crime study the Centre invested in producing a note of evidence to House of Commons committees. The study was cited in a House of Commons debate on 05.06.08 and on 22.10.07. As mentioned above, Roger was also actively engaged with the police, meeting key figures and speaking at the Police Federation conference.

The CCJS also made a submission to the Home Affairs committee, which responded (see Home Affairs Committee, 2009), referring to the evidence submitted by CCJS in connection with policy on amnesties and sentencing – thus giving more publicity to the report amongst politicians. Roger noted that ‘the aim of the report was to introduce evidence to a discussion that lacked it; the impacts were therefore about informing the public and decision-makers. In terms of changing official policy on sentencing, for example, there was no material impact. However, as an independent body dedicated to the public interest our business is to educate the public about social harms. We have every reason to think that in many key respects we succeeded’.

Case study 3: Trends in Obesity

Researchers: Leah Li, Christine Power (Institute of Child Health, University College London)

Data collections used: 1946 and 1958 Cohort (ESDS longitudinal)

The research

In England obesity is responsible for 9,000 premature deaths each year and reduces life expectancy by nine years. The cost to the health service of obesity is estimated to be in excess of £3.3 billion per year (HMG, 2004). Evidence from the cohort studies has been used to identify how quickly obesity can develop in young children and the types of backgrounds that can make some people more at risk from obesity than others.

Studies by Leah Li and other researchers at the MRC Centre of Epidemiology for Child Health at the Institute of Child Health, University College London (Li et al, 2008) studied Body Mass Index over time, using 1946 and 1958 British birth cohort data. Findings showed that BMI trajectories diverged from early adulthood, with a faster rate of BMI gain in the 1958 cohort than in the 1946 cohort. By mid-adulthood, the 1958 cohort had on average a greater BMI, larger waist and hip circumferences, and a higher prevalence of obesity. High BMI and excessive weight gain at different life stages are associated with cardiovascular disease, diabetes and some cancers.

In another study, Li and other colleagues (Pinot de Moira et al, 2010) examined members of the 1958 British birth cohort and offspring born to mothers under age 30 years to establish whether risk factors for childhood obesity have changed over time (1965–1991). The authors found that the prevalence of overweight/obesity had increased by more than 50 per cent between generations. There was evidence of a widening social gap in childhood obesity: Indicators of lower socioeconomic position were associated with increased BMI in offspring. Prevalences of parental obesity and maternal employment had increased. Socioeconomic factors had improved across generations. Parental obesity, maternal employment, and socioeconomic factors may play an increasing role in the childhood obesity epidemic.

Impact

These studies, by studying 'the health of the nation' and the causes of current health problems, were of both major public interest and of great interest to the general public. One impact, therefore, has been to stimulate discussion in the popular press. The Daily Mail (MaCrae, 2010) and Daily Telegraph (Collins, 2010) reported on the research, with both majoring on the finding that children of working mothers (described by the former as 'latchkey kids') are more likely to be obese.

The research has had possible policy impacts too (although it is difficult to isolate the effect of one study or paper or, indeed, of one research group, in government policy development).

Some work by the researchers on the 1958 cohort also contributed to a Department of Health report 'The changing social patterning of obesity' (Adamson et al, 2007), presented to the Policy Research Programme.

It is highly likely that research by these researchers – particularly as many contributed to that Department of Health (DH) report - contributed to the commissioning of a White Paper 'Healthy Lives, Healthy People: Our Strategy for Public Health in England', (DH, 2010) which sets out how the Government plans to improve public health. The DH has declared that 'reducing obesity is a priority for the Government' (DH, 2011: unpaginated), and has funded 'Change4Life', a national social marketing campaign to promote healthy weight. It aims to prevent people from becoming overweight by encouraging them to eat well, move more and, thus, live longer. These are all concrete examples of a fight against obesity that was shown to be necessary, at least in part, by the pioneering research undertaken by Leah Li and other researchers at the MRC Centre of Epidemiology for Child Health, using cohort study data available via ESDS longitudinal collections.

Pathways to impact

One key way in which the impacts described above came about was, as with the other case studies, by disseminating research findings. In this case, little needed to be undertaken by the researchers, as the interest in the subject was so high. Journals usually select papers for press release before they are published, so that media can read them first, and thus it is highly likely this is what occurred with regard to newspaper take-up of the findings.

Some work on the 1958 British birth cohort contributed to the 'The changing social patterning of obesity' report, which was funded by the DOH Public Health Research Consortium. This group brings together researchers from 11 UK institutions in an integrated programme of research, with the aim of strengthening the evidence base for interventions to improve health, with a strong emphasis on tackling socioeconomic inequalities in health. Impacts of the cohort data by this team are facilitated by links with the Policy Research Programme, a national research-funding programme within the Department of Health's Research and Development Directorate which commissions research.

Conclusions

Economic Impact and Benefits of the ESDS

The various approaches used to explore the economic impacts and benefits of ESDS data and services are tried-and-trusted approaches to valuing library, information and similar services, and they show that ESDS is at the higher end of the range of return on investment (RoI) reported for such services over the last decade.

The direct value of ESDS to the survey community of active registered users (excluding school and under-graduate students) is of the order of £25 million per annum at 2010 prices and levels of activity and use. Net of costs, that amounts to around £20 million per annum. The value to the wider user community, as measured in terms of efficiency gains and cost savings, is of the order of £100 million per annum or more, and we estimate that the increased use of the hosted data facilitated by ESDS delivers an increase in returns on annual investment in the creation/collection of the data of up to £233 million over 30 years (expressed in Net Present Value).

In addition to this, there are many wider economic and social impacts arising from the use of ESDS data (eg. in influencing policy decisions and driving social outcomes) that are not being captured in this analysis. Nevertheless, the forms of value that we are measuring reveal a net economic value that is more than five times ESDS operational costs, and an increase in returns on investment in hosted data that is up to ten times the costs. As such, we are able to demonstrate that ESDS generates considerable value at little direct operational cost - a benefit/cost and return on investment somewhat higher than is typical of academic, special and public libraries.

Economic Impact and Benefits of Social Science Research Data Service Infrastructure in General

While our analysis has focused on ESDS data and services, there is no obvious reason to think that ESDS is unusual. It may be more cost-effective and provide better supporting material and infrastructure than other social science data services, thereby generating more value for users, but it is unlikely to be very different. While quantitative studies of the impacts and value of research data infrastructure have not been common to date, those that have been conducted (Beagrie, et al. 2008, 2010; Fry et al. 2008; Houghton 2011) together with our analysis of ESDS suggest that there is likely to be value in social science research data infrastructure.

Our desk research, interviews, surveys (section 5) and case studies (section 6) also demonstrated some of the wider benefits arising from ESDS and explored some of the potential impacts on society generally in policy areas such as climate control, knife crime, and obesity from research based on ESDS data collections. The approaches applied to analysing wider benefits and impacts on social policy and practice are clearly applicable to other research data infrastructure.

Effectiveness of the Evaluation Methodology and Approaches

We have been asked as part of the study to undertake a critical review of the approach, methodology and results of the evaluation.

Interviews

The interviews had a high participation rate and provide rich context and information for the study. However they are relatively time-consuming and therefore expensive and only a relatively small (21) number of users, data suppliers and depositors could be approached by this method.

Online surveys

The online surveys are a (relatively) cheaper and faster method of approaching a much larger number of users and data suppliers and depositors. However they have imposed a number of limitations and trade-offs. Participants are often less willing to give as much time to completing an online questionnaire as they will to an interview so the format and number of questions needs to be more constrained to achieve reasonable levels of response. This was a persistent feature of feedback from user testing of the draft questionnaires which was reflected in the design. We have achieved high participation and completion rates, but have been constrained in the depth of questioning we can follow whilst at the same time exploring a range of questions needed for different economic methods.

“Economic” data and methods

Asking for economic data (costs, time expended, efficiency changes, etc.) at all was challenging for some users who were unaccustomed to thinking in that way or uncertain about the degree of confidence in their replies. However, others were clearly more confident and in some cases able to provide quite detailed explanations and illustrations in their responses - particularly in interviews, but also in the open-ended survey questions. Designing questions necessary to collect the range of economic data required in such a way that users could answer was a significant challenge and a number of compromises had to be made (eg. offering ranges to choose from rather than open-ended questions). Nevertheless, the high response and completion rates, and relatively small number of non or protest answers, gives us some confidence in the results and analysis. It is notable that, despite difficulties in framing contingent valuation questions for a survey community that typically does not pay for access to publications and data on an individual basis, we were able to elicit responses to the "willingness to pay" and "willingness to accept" questions that were plausible and (internally) confirmatory.

Similarly, the range of economic methods used provided an opportunity to use each to confirm the others, and the extent to which the results from the various approaches gel is heartening. While each of the approaches is partial and all conservative, both in the sense of being tried-and-trusted methods and of being applied conservatively to the data, they all produced results in which we can have some confidence - because of the high survey response and completion rates and the internal consistency of the answers.

A good deal of effort was put into weighting the survey data prior to the economic analysis, in order that the values expressed by respondents reflect the broader user community, and

not simply our respondents, and the totality of ESDS data and services, and not simply the part that our respondents used. As the focus of the study was on the value to users and in use, user survey weighting was done on the basis of data collections delivered (use). The depositor survey was weighted by data collections acquired and processed for online delivery. This involved a two-step process of disaggregation of responses by data type accessed/downloaded for users and deposited for depositors and re-aggregation by data collections delivered/acquired, followed by the calculation of weighted means from responses about costs, time, values, etc. While this seems to have worked reasonably well, it would no doubt be easier and more accurate to focus analysis on particular parts of ESDS data and services. This, together with further mining of the rich survey data captured in this study, might be worthwhile follow-up activities.

Recommendation 5: Consider further mining the survey data at lower levels of aggregation, using "cuts" of the survey responses of interest. This would enable a more detailed view of how users of particular elements of ESDS data and services value and benefit from those data and services.

Finally, we feel further research could be undertaken to try to quantify the wider economic and social impacts. For example, as noted in section 4.8 there is an extensive literature on returns to R&D and a range of studies have explored possible impacts of increases in accessibility and/or efficiency on returns to research expenditure in other disciplines. We had originally hoped that this approach might be possible for this study. However, we found that relatively little is known about the economic and social returns to investment in social science research. As a result it was not possible to explore estimates of possible impacts of increases in accessibility and/or efficiency on returns to research expenditure.

While extremely challenging, future research might involve further development of macro modelling approaches and collection and refinement of the necessary supporting data and/or exploring means of integrating quantitative and qualitative valuations into a single framework. It might also involve further development of approaches exploring the impacts of research data infrastructure services on the accessibility of research (data) outputs and efficiency of research data use, and/or integrating semi-quantified qualitative assessments through, for example, the use of a balanced scorecard approach.

Recommendation 6: Consider further research to quantify the wider economic and social impacts of social science research infrastructure.

Counter-factual data

Counter-factual analysis has a number of practical constraints. The most notable of these is that users in the online survey in most cases (approx. 70 per cent) think that If ESDS had not existed they would not have been able to obtain the data in any other way. Moreover, 58 per cent said that it was beyond their capacity to (re)create the data themselves. Hence the extent to which ESDS is saving data (re)creation costs may be rather limited, with the main impact being facilitating research that could not otherwise have been done.

It is also often difficult to compare “like with like” when analysing data for comparison, particularly if the comparisons are drawn from different jurisdictions or sectors. Pricing of services is rarely the same as the cost of those services (ie full cost recovery may not apply to prices) and there is often considerable disparity in the pricing models used by data producing agencies.

In addition, much of the data hosted by ESDS would have been collected whether or not ESDS existed or hosted the data. As such, data creation costs are, effectively, sunk costs. Hence, we have preferred to explore increases in returns on investment in data creation/collection, focusing on ESDS facilitation of additional uses.

Use of KRDS to assess the wider benefits of ESDS

We found the KRDS approach and framework to be an effective way of summarising the wide-range of benefits from ESDS identified in the desk research, interviews and surveys.

Impact Case Studies

As with previous ESRC research on impact, we found it difficult to identify case studies that can conclusively show direct impact on policy and practice. This is a challenging task because of the widely acknowledged difficulties associated with attribution and time-lags. In addition two of the best examples of policy impact mentioned by interviewees proved to be based on data from the Census Service and therefore ineligible for this study. The three impact case studies we did explore represent a spread of use of the main ESDS data collections: Case Study 1 using International data (climate control); Case Study 2 using Government data (knife crime); and Case Study 3 using Longitudinal data (Obesity). All three are valuable in demonstrating how research based on ESDS has had significant impact in terms of debate and media coverage of these major social issues.

The terms of reference for this study included undertaking the case studies to assess the policy and practice impact of research based on data accessed via the ESDS. The concern was that while economic evaluation is important and necessary, it needs to be put into context. In practice, we have found a thorough approach to economic impact is not easily accommodated in a single report with policy impact via case studies. Each is a major topic.

Scope of the Study

The scope of the study was very wide and future studies would benefit from a narrower scope. This breadth of scope has a number of different aspects: the range of activities carried out by the ESDS and the focus of this study on the “macro” level; the range of different economic approaches applied; and the inclusion of policy impacts via case studies alongside the economic analysis. This was an ambitious scope for the timescale and resources available.

Timescale of the Study

The effective timescale for and timing of the study was problematic. The timescale for the study to complete its data analysis and draft final report was originally 5 ½ months (beginning of July to mid-December) but this included the summer months when many interviewees and staff would be unavailable. Although scheduled to commence at the

beginning of July there was a few weeks delay so the study commenced in mid-July and timescales became very tight.

Recommendation 7: We would recommend longer timescales (or narrower scope) for future studies and allowing elapsed time for major holiday periods wherever possible.

Overall Results

Taken together, the interviews, case studies and online surveys provide a rich source of information about ESDS data and services, their users and uses. The economic analysis suggests considerable value for users by ESDS data and services at relatively little cost, with a benefit/cost ratio and return on investment at the higher end or above what is typical of library and information services. The nature of the wider benefits is drawn out in interviews and case studies, which demonstrate the pathways to realising the value and benefits. Many detailed responses in the surveys and interviews suggest ways in which ESDS might improve its services and enhance its value proposition.

Summary of Recommendations

A summary of the recommendations that have been made in the report is provided below under two main headings. Each recommendation is reproduced with a note of where it first appears in context of the report text in brackets eg. [section 5.2].

Good Practice/Lessons for Maximising the Benefits of Research Data Service Infrastructure

Recommendation 1: Review how ESDS usage statistics reflect teaching use. Interview evidence shows substantial local use per individual download. [section 5.2]

Recommendation 2: The absence of sufficient skills in quantitative methods remains a significant barrier to maximising use of ESDS and other data. ESRC and ESDS should continue to encourage re-use of data, use of quantitative methods, and use of ESDS data collections in teaching. [section 5.2]

Recommendation 3: If ESRC wishes to extend use of ESDS by the commercial sector or Government, it could provide more digests and simpler summaries of the findings from the data collection that might promote greater usage by a wider range of users. [section 5.2]

Recommendation 4: The JISC as a co-funder of the ESDS alongside the ESRC may wish to note our findings on the economic impact of ESDS and the wider benefits identified from its fostering of best practice and innovation in research data management. [section 5.2]

Approaches to Future Economic Impact Evaluations

Recommendation 5: Consider further mining the survey data at lower levels of aggregation, using "cuts" of the survey responses of interest. This would enable a more detailed view of how users of particular elements of ESDS data and services value and benefit from those data and services. [section 7.3]

Recommendation 6: Consider further research to quantify the wider economic and social impacts of social science research infrastructure. [section 7.3]

Recommendation 7: We would recommend longer timescales (or narrower scope) for future studies and allowing elapsed time for major holiday periods wherever possible. [section 7.3]

References

All URLs checked 16 February 2012

Adamson, A., Chadwick, T., Dezateux, C., Griffiths, L., Howel, D., Kelly, S., et al. (2007). *The changing social patterning of obesity: an analysis to inform practice and policy development*. Peer reviewed report to the Policy Research Programme, DOH. Available http://phrc.lshtm.ac.uk/papers/PHRC_BI-06_Final_Report.pdf

Barker, T., Pan H., Köhler, J., Warren, R., & Winne, S. (2006). Avoiding dangerous climate change by inducing technological progress: Scenarios using a large-scale econometric model. In, Schellnhuber, H.J., Cramer, W., Nakicenovic, N., Wigley, T., & Yohe, G. (Eds.) *Avoiding Dangerous Climate Change* (pp. 361-373). Cambridge: Cambridge University Press

Barker, T., Anger, A., Dessens, O., Pollitt, H., Rogers, H., Scricciu, S., et al. (2010). Integrated modelling of climate control and air pollution: Methodology and results from one-way coupling of an energy–environment–economy (E3MG) and atmospheric chemistry model (p-TOMCAT) in decarbonising scenarios for Mexico to 2050, *Environmental Science and Policy*, 13(8), 661-670.

Beagrie, N., Chruszcz, J., & Lavoie, B. (2008). *Keeping Research Data Safe: a cost model and guidance for UK Universities*, Joint Information Systems Committee. Available <http://www.jisc.ac.uk/publications/publications/keepingresearchdatasafe.aspx>

Beagrie, N., Lavoie, B., & Woollard, M. (2010). *Keeping Research Data Safe 2 Final Report* London: JISC. Available <http://www.jisc.ac.uk/publications/reports/2010/keepingresearchdatasafe2.aspx#downloads>

Beagrie, N., Chruszcz, J., & Lavoie, B. (2008). *Keeping Research Data Safe*, London: JISC. Available <http://www.beagrie.com/krds.php>

Beagrie, N., Lavoie, B., & Woollard, M. (2010). *Keeping Research Data Safe 2*, London: JISC. Available <http://beagrie.com/krds-i2s2.php>

Beise, M., & Stahl, H. (1998). *Public Research and Industrial Innovations in Germany*, ZEW Discussion Paper No. 98-37. Available <http://ideas.repec.org/p/zbw/zewdip/5209.html>

Berman, G. (2011). *Standard note: Knife crime statistics*, London: House of Commons Library. Available <http://www.parliament.uk/briefing-papers/SN04304.pdf>

British Library. (2004). *Measuring the Economic Impact of the British Library*, [note: a powerpoint slide presentation based on unpublished report]. Available <http://www.bl.uk/aboutus/stratpolprog/increasingvalue/publicvalue/confpres/pungelwesmarks.pdf>

Cambridge Centre for Climate Change Mitigation Research. (2011). E3MG Cambridge Centre for Climate Change Mitigation Research website. Available <http://www.4cmr.group.cam.ac.uk/the-model>

Charles Beagrie Ltd. (2011). *KRDS/I2S2 Digital Preservation Benefits Analysis Toolkit*. Available <http://beagrie.com/krds-i2s2.php>

Collins, C. (2010, May 10). Children of working mothers 'more likely to be overweight' *Daily Telegraph Online*. Available http://www.telegraph.co.uk/health/children_health/7767858/Children-of-working-mothers-more-likely-to-be-overweight.html

DECC (Department of Energy and Climate Change) (2012). *UK Greenhouse Gas Emissions: Performance Against Emissions Reduction Targets – 2010 Provisional Figures* Available: <http://www.decc.gov.uk/media/viewfile.ashx?filetype=4&filepath=11/stats/climate-change/2351-uk-greenhouse-gas-emissions-performance.pdf&minwidth=true>

Dekkers, M., Polman, F., te Velde, R., & de Vries, M. (2006). *MEPSIR: Measuring European Public Sector Information Resources*, Brussels: European Commission. Available http://www.epsipus.net/psi_library/reports/mepsir_measuring_european_public_sector_resources_report

Demographic Decisions Ltd. (2009). *A business case for microdata files from the 2011 Census*, London: Demographic Decisions Ltd. Available <http://www.ccsr.ac.uk/sars/2011/documents/businesscase.pdf>

DH (Department of Health). (2010). *Healthy lives, healthy people: our strategy for public health in England*. Available http://www.dh.gov.uk/prod_consum_dh/groups/dh_digitalassets/documents/digitalasset/dh_127424.pdf

DH (Department of Health). (2011). *Obesity*. Available <http://www.dh.gov.uk/en/PublicHealth/Obesity/index.htm>

DotEcon. (2006). *The commercial use of public information (CUPI)*, Report oft861, London: Office of Fair Trading. Available <http://www.oft.gov.uk/OFTwork/publications/publication-categories/reports/consumer-protection/oft861>

DTLR. (2002). *Economic Valuation with Stated Preference Techniques*, London: Department of Transport, Local Government and the Regions, for a summary guide. Available <http://www.communities.gov.uk/documents/corporate/pdf/146871.pdf>

ESDS. (2006). *Economic and Social Data Service Annual Report August 2005 – July 2006* Available <http://www.esds.ac.uk/news/publications/annrep2006.pdf>

ESDS. (2007). *Economic and Social Data Service Annual Report August 2006 – July 2007* Available <http://www.esds.ac.uk/news/publications/annrep2007.pdf>

ESDS. (2008). *Economic and Social Data Service Annual Report August 2007 – July 2008* Available <http://www.esds.ac.uk/news/publications/annrep2008.pdf>

- ESDS. (2009). *Economic and Social Data Service Annual Report August 2008 – July 2009*. Available <http://www.esds.ac.uk/news/publications/annrep2009.pdf>
- ESDS. (2010a). *Economic and Social Data Service Annual Report August 2009 – July 2010*. Available <http://www.esds.ac.uk/news/publications/ESDSAnnualReport2009-2010.pdf>
- ESDS. (2010b). *ESRC/JISC Economic and Social Data Service: Director's Evaluation Report* (unpublished confidential document)
- ESRC. (2011). *Impact Assessment*, Swindon: ESRC. Available <http://www.esrc.ac.uk/impacts-and-findings/impact-assessment/>
- Evans, P., Hatcher, M., & Whittard, D. (2008). The preliminary satellite account for the UK: a sensitivity analysis, *Economic & Labour Market Review*, 2(9), September 2008, 37-43.
- Fry, J., Lockyer, S., Oppenheim, C., Houghton, J.W., & Rasmussen, B. (2008). *Identifying benefits arising from the curation and open sharing of research data produced by UK Higher Education and research institutes*, London and Bristol: JISC. Available <http://www.jisc.ac.uk/media/documents/programmes/digitalrepositories/jiscdataproposal-public.pdf>
- Golding, B., & McClory, J. (2008). *Going Ballistic: Dealing with guns, gangs and knives*, London: Policy Exchange. Available http://www.policyexchange.org.uk/images/publications/pdfs/Going_Ballistic.pdf
- Griliches, Z. (1995). R&D and productivity: Econometric Results and Measurement Issues, In Stoneman, P. (Ed.) *Handbook of The Economics of Innovation and Technological Change*. Oxford: Blackwell, 52–89.
- Hall, B.H., Mairesse, J., & Mohnen, P. (2009). *Measuring the returns to R&D*, NBER Working Paper 15622, NBER, Cambridge MA.
- Hall, B.H., Mairesse, J., & Mohnen, P. (2010). Measuring the returns to R&D, in Hall, B.H., & Rosenberg, N. (Eds.) *Handbook of the Economics of Innovation*, North Holland.
- HMG. (2004). *Obesity: Third Report of the Session 2003-04, Volume I, House of Commons Health Committee*. Available <http://www.publications.parliament.uk/pa/cm200304/cmselect/cmhealth/23/23.pdf>
- HM Treasury. (2011). *The Green Book*, London: HM Treasury. Available http://www.hm-treasury.gov.uk/d/green_book_complete.pdf
- Home Affairs Committee. (2009). *The Home Affairs Committee Seventh Report: Knife Crime 2008/9* London: HMSO. Available <http://www.publications.parliament.uk/pa/cm200809/cmselect/cmhaff/112/11202.htm>
- Houghton, J.W. (2009). Exploring the impacts of enhanced access to publicly funded research, in Uhler, P. (Ed.) *The socioeconomic effects of public sector information on digital*

networks: *Towards a better understanding of different access and reuse policies*, Washington DC: National Academies Press. Available

http://www.nap.edu/openbook.php?record_id=12687&page=51

Houghton, J.W., Rasmussen, B., Sheehan, P.J., Oppenheim, C., Morris, A., Creaser, C., et al. (2009). *Economic Implications of Alternative Scholarly Publishing Models: Exploring the Costs and Benefits*, Report to The Joint Information Systems Committee (JISC) by Victoria University & Loughborough University. Available <http://www.cfses.com/EI-ASPM/>

Houghton, J.W., Rasmussen, B., & Sheehan, P. (2010) *Economic and Social Returns on Investment in Open Archiving Publicly Funded Research Outputs*, Washington DC: The Scholarly Publishing and Academic Resources Coalition (SPARC). Available <http://www.arl.org/sparc/publications/papers/vuFRPAA/index.shtml> (August 2011).

Houghton, J.W. (2011). *Costs and benefits of data provision*, Report to The Australian National Data Service, Canberra. Available <http://ands.org.au/resource/cost-benefit.html>

Houghton, J.W., & Sheehan, P. (2009). Estimating the potential impacts of open access to research findings, *Economic Analysis and Policy* 39(1). Available http://www.eap-journal.com/vol_39_iss_1.php

Houghton, J.W., Swan, A., & Brown, S. (2011). *Access to Research and Technical Information in Denmark*, Forsknings- og Innovationsstyrelsen and Denmark's Electronic Research Library (April 2011). Available <http://www.fi.dk/publikationer/2011/adgang-til-forskningsresultater-og-teknisk-information-i-danmark>

Hunter, L. (2006). *Investment in an Intangible Asset*, Instalment in DCC Digital Curation Manual (Digital Curation Centre 2006). Available <http://www.dcc.ac.uk/resources/curation-reference-manual/completed-chapters/investment-intangible-asset>

Imholz, S., & Arns, J. (2007). *Worth Their Weight: An Assessment of the Evolving Field of Library Valuation*, New York: Americans for Libraries Council. Available

<http://www.ala.org/ala/research/librarystats/worththeirweight.pdf>

ISSR. (2011). *Evidence-based Policy and Research Utilisation*, Brisbane: University of

Queensland. See <http://www.issr.uq.edu.au/policy-analysis-evaluation-projects#Evidence>

King, D.W. (2010). Presentation on Library Value. Available <http://libvalue.cci.utk.edu/sites/default/files/King2010June.pptx>

Leppard, D. (2007, August 19). Knife crime doubles in two years, *The Times*, p2

Li, L., Hardy, R., Kuh, D., Lo Conte, R., & Power, C. (2008). Child-to-Adult Body Mass Index and Height Trajectories: a comparison of 2 British Birth Cohorts. *American Journal of Epidemiology*. 168(9), 1008-1015.

Mansfield, E. (1991). Academic research and industrial innovation. *Research Policy* 20(1), 1991, 1-12.

MaCrea, F. (2010, May 10). Children of working mothers are telly tubbies: Obesity concerns for latchkey kids. *Daily Mail Online*. Available <http://www.dailymail.co.uk/news/article-1281444/Why-children-working-mothers-telly-tubbies-Latchkey-kids.html>

Mansfield, E. (1998). Academic research and industrial innovation: an update of empirical findings. *Research Policy* 26(7/8), 1998, 773-776.

Martin, B.R., & Tang, P. (2007). The benefits from publicly funded research. *SWEPS Paper No. 161*, Science Policy Research Unit, Brighton: University of Sussex. Available <http://www.erawatch-network.eu/reports/sewpl61.pdf>

Metz, B., Davidson, O.R., Bosch, D.R., & Meyer L.A. (Eds.). (2007). *Climate Change 2007: Mitigation of Climate Change: Contribution of Working Group III to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change*, Cambridge and New York: Cambridge University Press

Oakleaf, M. (2010). *Value of Academic Libraries: A Comprehensive Research Review and Report*, Chicago: Association of College and Research Libraries (ACRL). Available http://www.ala.org/acrl/sites/ala.org/acrl/files/content/issues/value/val_bibliography.pdf

Pinot de Moira, A., Power, C., & Li, L. (2010). Changing influences on childhood obesity: a study of 2 generations of the 1958 British birth cohort. *American Journal of Epidemiology*, 171(12), 1289-98

PIRA. (2000). *Commercial exploitation of Europe's public sector information*, Brussels: European Commission. Available http://www.epsipius.net/psi_library/reports/commercial_exploitation_of_europe_s_public_sector_information_pira_study

Pollack, R. (2009). *The Economics of Public Sector Information*, CWPE 0920, Cambridge: University of Cambridge. Available <http://www.econ.cam.ac.uk/dae/repec/cam/pdf/cwpe0920.pdf>

Salter, A.J. & Martin, B.R. (2001). The economic benefits of publicly funded basic research: a critical review, *Research Policy* 30(3), 509-532.

Scott, A., Steyn, G., Geuna, A., Brusoni, S., & Steinmueller, E. (2002). *The Economic Returns to Basic Research and the Benefits of University-Industry Relationships*. Report to the Office of Science and Technology, London.

Shanks, S. & Zheng, S. (2006). *Econometric modeling of R&D and Australia's productivity*, Staff Working Paper, Canberra: Productivity Commission

Simmonds, P., Stroyan, J., Brown, N., & Parker-Rhodes, L. (2011). *Data centres: their use, value and impact*, Research Information Network. Available <http://www.rin.ac.uk/our-work/data-management-and-curation/benefits-research-data-centres>

Steel, J. (2007, March 20). The vagaries of UK knife crime statistics. *The Daily Telegraph* Available <http://www.telegraph.co.uk/news/uknews/1546085/The-vagaries-of-UK-knife-crime-statistics.html>

Sveikauskas, L. (2007). *R&D and Productivity Growth: A Review of the Literature*, US Bureau of Labor Statistics Working Paper 408, Washington, DC: BLS.

Tenopir, C. & King, D.W. (2000). *Towards Electronic Journals: Realities for Scientists, Librarians and Publishers*, Washington D.C: Special Libraries Association

Tenopir, C., Love, A., Park, J., Wu, L., Baer, A., & Mays, R. (2010). *University Investment in the Library, Phase II: An International Study of the Library's Value to the Grants Process*, Elsevier Connect White Paper. Available <http://libraryconnect.elsevier.com/whitepapers/roi2/lcwp021001.html>

TRAC fEC Guidelines (2011). Available <http://www.jcpsg.ac.uk/guidance/index.htm>

THES (2008, March 13) *How much are you worth?*. Available <http://www.timeshighereducation.co.uk/story.asp?sectioncode=26&storycode=401025&c=1>

THES (2011, June 20). *Graduate salaries to rise as competition heats up*. Available <http://www.timeshighereducation.co.uk/story.asp?storycode=416639>

Woollard, M. (2011). KRDS Benefits Framework, Value- Chain and Benefit Analysis Tools: UK Data Archive Case Study. Paper presented at the JISC Digital Preservation Benefits Tools Project Workshop, London, 12 July 2011. Available http://www.esds.ac.uk/news/eventsdocs/JISC_digitalpreservation12jul2011.pdf

Appendix A: Findings from the Surveys

Survey Populations and Response Rates

Depositor Survey: a list of depositors and data suppliers since 1st January 2006 was drawn up by ESDS staff and individuals contacted for interview removed to prevent duplication of responses. A total of 638 were successfully emailed and invited to participate in the depositor/data supplier survey, and 193 individuals (30 per cent) responded: a very high response rate compared to other recent surveys of the community.

User Survey: users registering with ESDS can give (or withhold) consent to be contacted via email by ESDS. An initial survey population of all registered users who had given consent was prepared and narrowed down to exclude school students, under-graduates (effects on teaching were assessed via questions to their academic tutors as it was felt that the students themselves would be unable to respond well to the topics covered), and international users outside the Eurozone, North America, and Australia (to limit the number of currencies in the survey and possible language difficulties). A total of 6,102 registered users were successfully emailed and invited to participate in the user survey, of which 5,534 were from the UK and 568 were international. A total of 952 individuals (16 per cent) responded. While some questions were optional, and some were indeed skipped, the overall completion rate was high.

The composition of the respondent population was found to be very similar to the user survey population in terms of user affiliation (HEI, Government, etc.) and role (academic staff, post-graduate, etc.). The composition of the wider ESDS user community was also similar with the major exception of under-graduates (excluded from the survey). It should be noted that under-graduates form between 11 per cent and 33 per cent of the users for the different main ESDS data collections (ESDS 2010a p. 28). The exclusion of school and undergraduate students from the survey has little impact on the economic analysis, as they have also been excluded from estimates based on user counts. Moreover, their time would have been attributed zero cost and their willingness to pay highly constrained by their ability to pay. Thus, while they do realise value for ESDS in educational terms, their impact would be small in direct economic cost/value terms. As the economic analysis is based on data weighted by use, any differences between the pattern of respondents' use and overall use of ESDS will have little effect (See Section 4).

Although five Amazon vouchers were offered to participants in the surveys we could see no evidence that this distorted responses to the surveys (either in terms of response rates from different groups or “opportunistic” replies): as might be expected given the relatively small number of vouchers (1 in 229 respondents received a voucher) and their relatively low value.

Survey Analysis

The user and depositor surveys contained both qualitative and quantitative elements. This Annex presents a brief description of responses. As noted above, during pilot testing of the questionnaires, some of the quantitative questions proved difficult for respondents to answer. As a result, a number of compromises were made to ensure good response and completion rates. These included the use of drop-down lists offering ranges instead of open

ended value questions. In these cases, we interpret the responses as the mid-point of the ranges, and those opting for the upper bound responses (eg. > 10) we interpret, conservatively, as one greater than the number specified. These instances are noted in the following description of results.

User Survey

The first five questions established the nature and affiliation of the respondent, their use of ESDS data and services and their location (local currency).

Q1: Reflecting the ESDS user community, 81 per cent of respondents were in higher and further education, 15 per cent in government and non-profit organizations and just over 4 per cent in commercial organizations and others (eg. independent scholars).

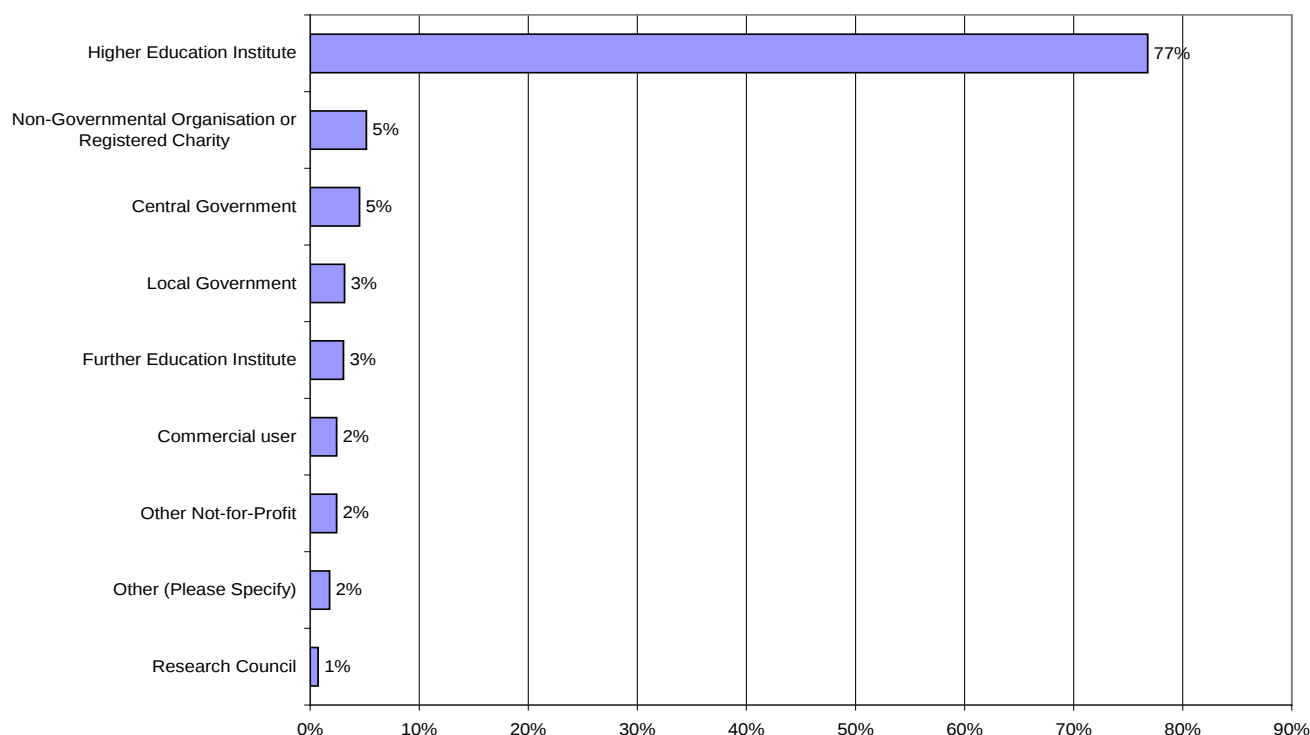


Figure 5: Main affiliation of respondents (N=952).
Source: User Survey, Authors' analysis.

Q2: Around 40 per cent of user respondents were staff, primarily academic and research staff, and 40 per cent post-graduate students and research assistants. The remaining 20 per cent reported a wide range of roles, such as general management, librarians, research managers, policy analysts and advisors, consultants and retired independent scholars.

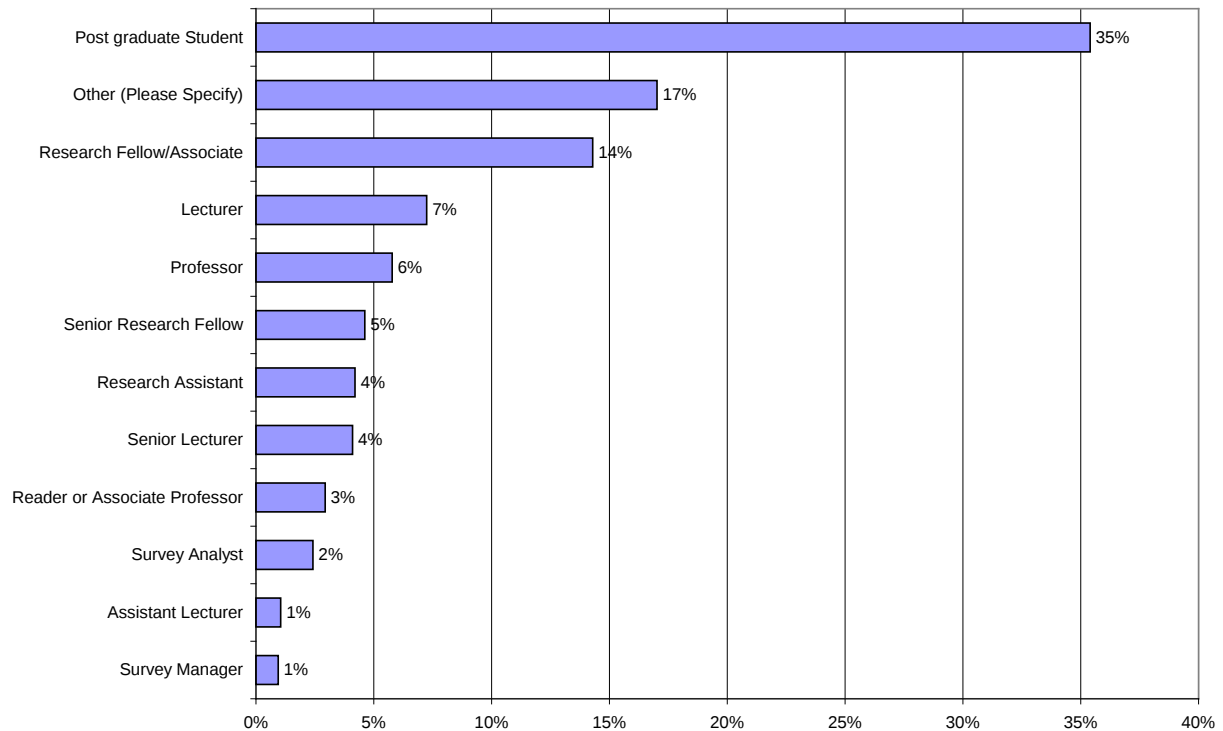


Figure 6: Main role of respondents (N=952).
Source: User Survey, Authors' analysis.

Q3: By research field or discipline, more than 40 per cent were in economics and business related fields, and it is not surprising that many of the others were in what might be considered more quantitative fields. Health and medicine, geography and areas studies were also well represented. Among those specifying other fields, many could perhaps have selected one of the alternatives offered, but others reported a wide range of fields (eg. engineering, urban planning, etc.), suggesting that there are many users of ESDS data and services who are not social scientists per se. This is indicative of ESDS's wider economic and social impacts.

These disciplinary shares are broadly similar to the those of reported ESDS data downloads by discipline during 2009-10, which were: Business/Accounting/Finance 11 per cent; Economics/Econometrics/Labour/Employment 45 per cent; Geography/Area Studies 4 per cent; Health/Medicine 4 per cent; History/Humanities 1 per cent; Politics and International Studies 6 per cent; Psychology 2 per cent; Social Policy/Administration 4 per cent; Sociology 8 per cent; Statistics/Methodology/Computing 4 per cent; and Other 11 per cent (ESDS 2010a) .

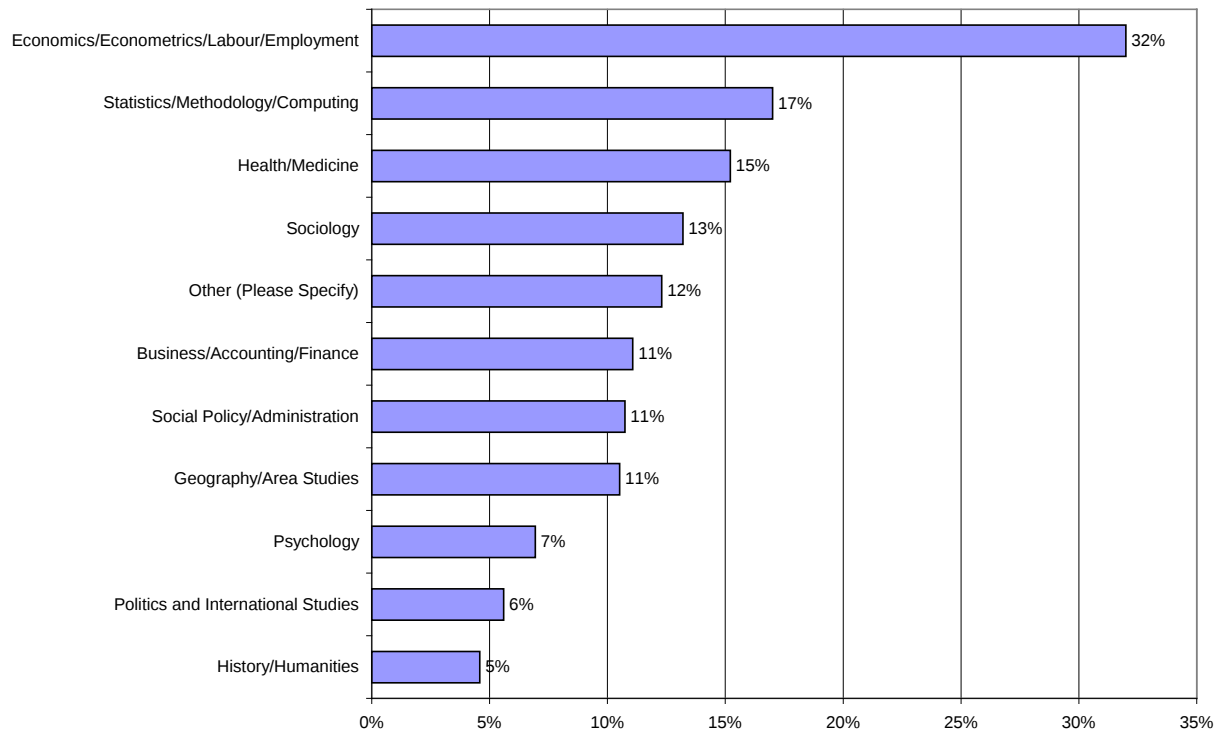


Figure 7: Main discipline or field of respondents (N=894).
Source: User Survey, Authors' analysis.

Q4: Asked for their local currency, for use in interpretation of the costs reported in later questions, 84 per cent nominated GBP (United Kingdom), 9 per cent the Euro zone, 5 per cent the United States, and 1 per cent each Australia and Canada. Non-UK currencies were converted to British pounds at an exchange rate current at the time the survey closed (ie EUR 1.138, USD 1.611, CAN 1.601 and AUD 1.504).

The next section of the user survey (questions 5 through 13) explored the types of data used, frequency of use, the time and costs involved in that use, and what impacts that use has on the users' work.

Q5: Asked which of the major ESDS data types they had used, respondents could tick multiple options. Indeed, 894 respondents gave 1,410 responses. More than 45 per cent reported having used Longitudinal Survey Data (eg. British Household Panel Survey) and Large-scale government surveys, (eg. the Labour Force Survey and the British Crime Survey). Around 23 per cent reported having used Other survey data (eg. British Election Studies, Farm Business Survey or smaller ESRC-funded academic research projects) and Multi-nation databanks and International survey data (eg. European Quality of Life Survey or World Bank's World Development Indicators), and 19 per cent Qualitative and mixed methods data.

While not directly comparable and open to respondent interpretation, these response shares reflect somewhat lower use of government data and higher use of other categories than reported data collections delivered by ESDS during 2009-10 (ESDS 2010a).

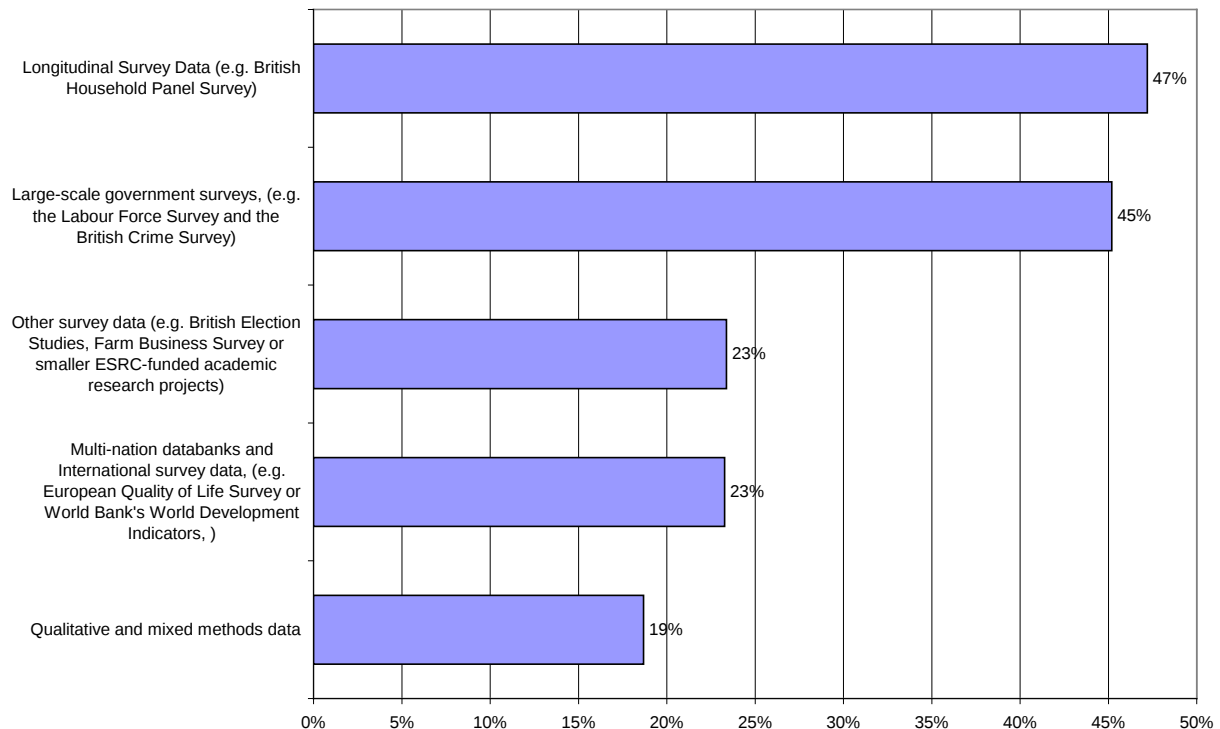


Figure 8: ESDS data types used (N=894).
Source: User Survey, Authors' analysis.

Q6: Asked to rate the importance of ESDS data and services for various aspects of their work on a five-point scale from "not at all important" to "extremely important", respondents rated the importance for academic research and policy development more highly than teaching and commercial research or consultancy. Not surprisingly, those in higher education rate its importance for research and teaching more highly, while those in government and non-profit organizations rated its importance in policy development and consultancy more highly.

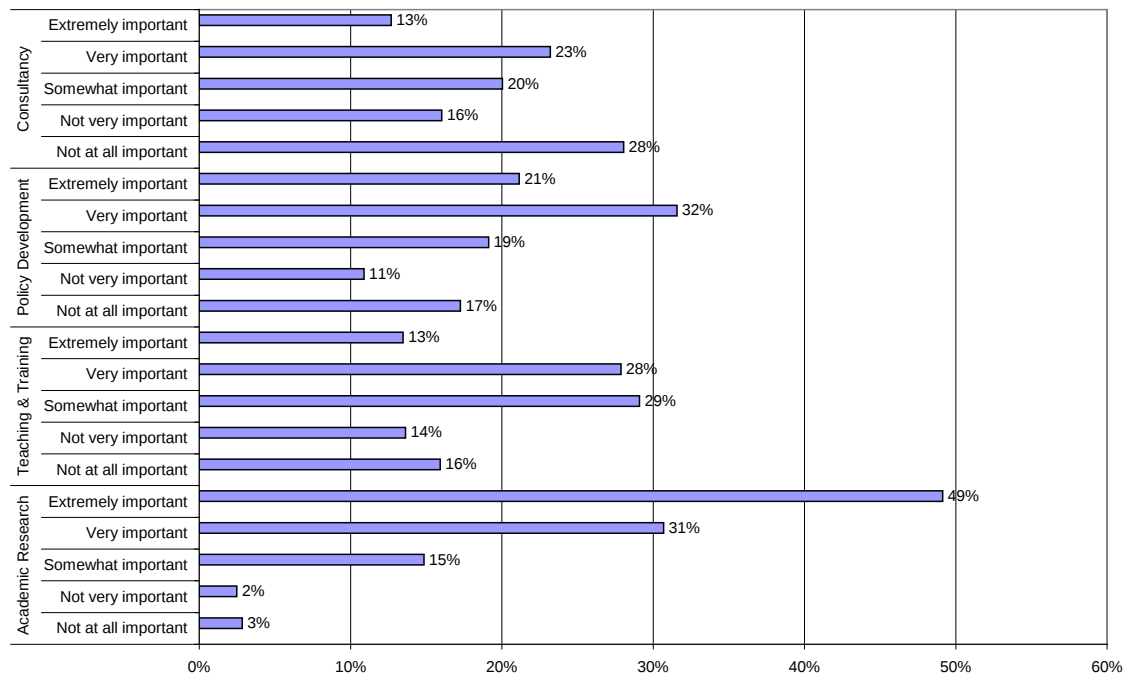


Figure 9: Importance of ESDS data and services in respondents' work activities (N=854). Source: User Survey, Authors' analysis.

Q7: Asked what impact not having access to ESDS data and services would have on their work, the majority of respondents (61 per cent) suggested that it would have a major (33 per cent) or severe (28 per cent) impact. The various types of respondents gave slightly different answers, with academics suggesting that not having access would have a major to severe impact, government and non-profit organization users a severe to moderate impact, and corporate and other users a moderate impact. From the academics, staff suggested that they would be affected slightly more than students. Users of big data collections (ie Longitudinal Survey Data, Multi-nation databanks and International survey data, and Large-scale government surveys) suggested that they would be much more affected than users of small data collections (ie Qualitative and mixed methods data, and Other survey data).

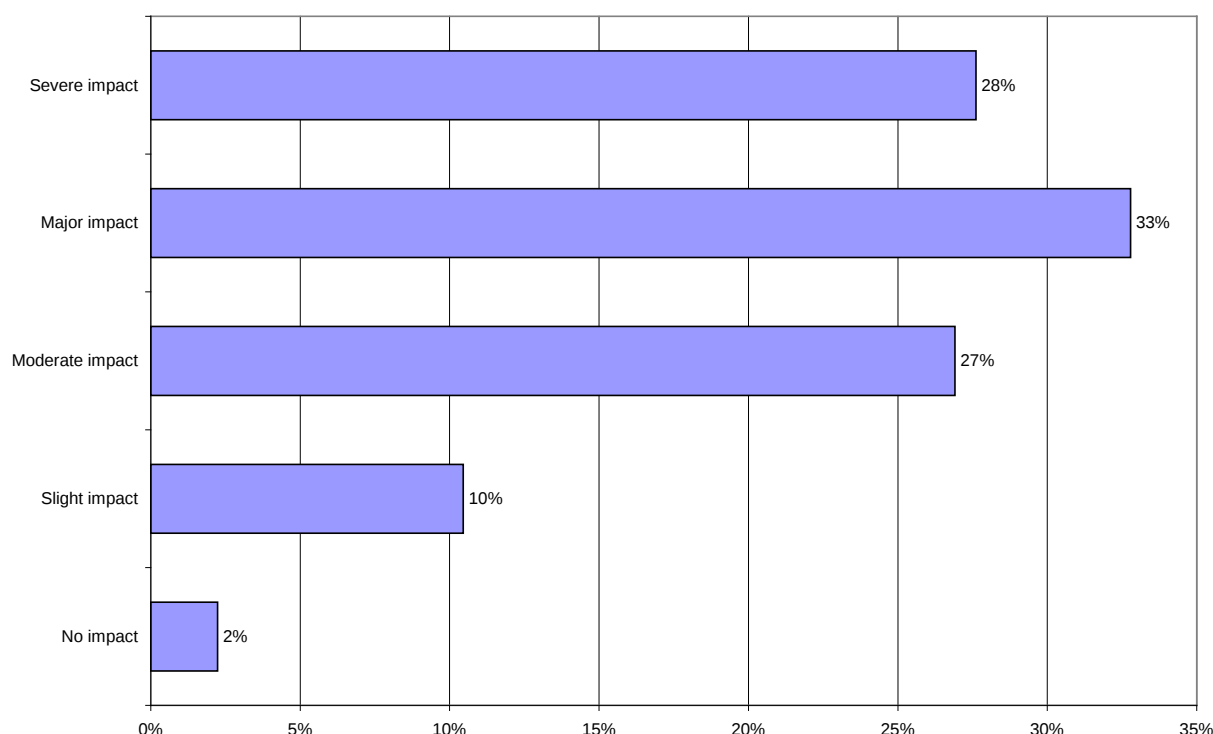


Figure 10: Impact of not accessing ESDS data and services on their work (N=851).

Source: User Survey, Authors' analysis.

This question elicited almost 100 comments (11 per cent of responses), of which the majority were positive about the data and related guides and support services. Some of the comments are personal, others are generic. Not surprisingly, those suggesting that no access to ESDS would have a severe or major impact were more positive than those suggesting that no access would have a slight impact. Comments included:

‘Acquiring all the information I needed for my research and projects would have been much more time consuming and the process more difficult and daunting. In addition, the accuracy of the information from various other sources is questionable.’

‘As a student, I have found data, services, and resources provided by ESDS invaluable in designing my own MA dissertation, learning about research methods (eg., different types of interviews); seeing examples of the methods used in successful research projects (eg., interview schedules), and thinking through and drafting a proposal for future PhD research involving secondary analysis of existing data.’

‘It would make teaching under-graduate quantitative methods very difficult and result in less quantitatively competent graduates. It would mean that I could not easily access data that is essential in doing my research - especially in generalising from qualitative case studies.’

‘Many studies, both large and small, simply would not take place if ESDS/UKDA did not exist.’

'To be able to access the data at all is vital to my work; to be able to access from a central portal is valued greatly as it is reliable, staff are helpful and as such it is a very efficient way of accessing the data sources.'

'Access to high quality data provided by ESDS International is essential for quality research, the support that is given saves so much time and effort.'

'The ready availability of high quality data for academic and policy-related research has transformed the way in which organizations like mine work.'

Q8: Respondents were offered pre-defined ranges to determine how often they access/download ESDS data. For quantitative analysis, mid-point values were used, with 'once in 3 years or less' interpreted as 0.33 times a year and more than 10 times a year interpreted as 11 times. The mean frequency reported was 3.9 times per annum.

Users in higher and further education reported using ESDS data more often than government and non-profit users, and both groups used the data more often than corporate and other users. Frequency of use between academic staff and post-graduate students did not vary much. Users of big data collections access/download the data much more often than users of small data collections - probably reflecting the types of use.

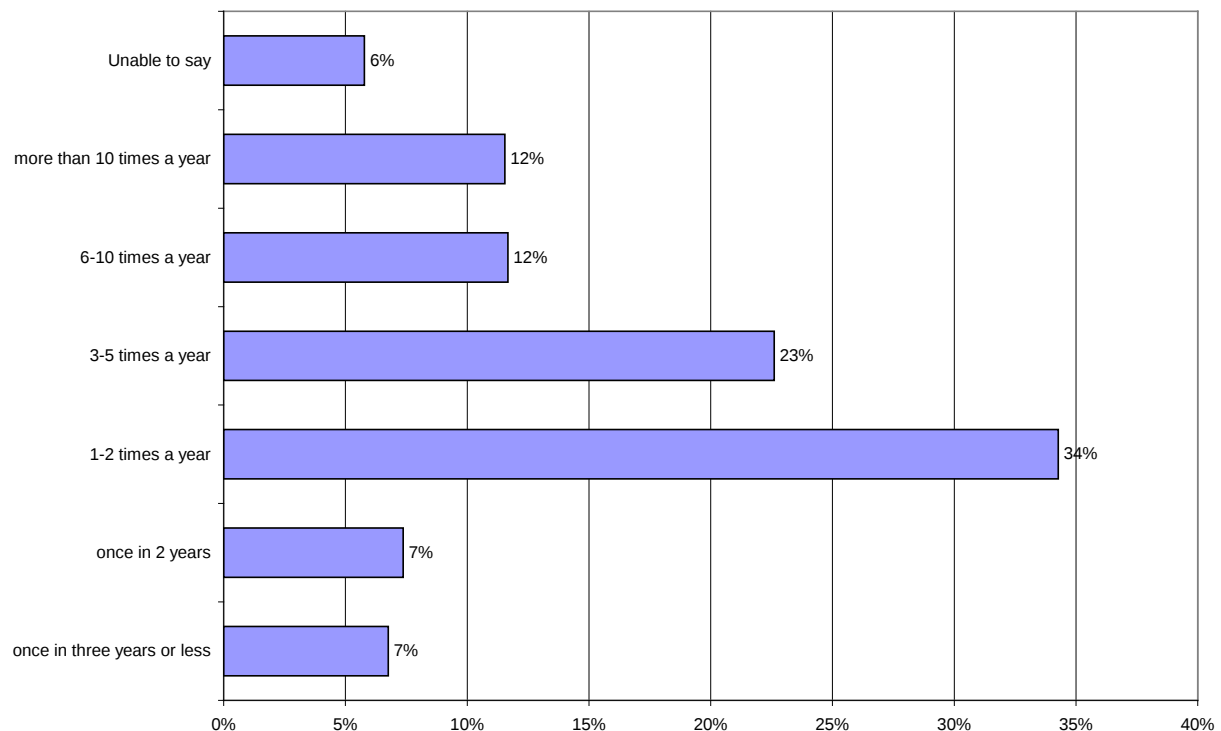


Figure 11: Frequency of access/download during the last 12 months (N=814).

Source: User Survey, Authors' analysis.

Q9: Using a "critical incident" approach, respondents were asked how long it took them to access the last data they accessed/downloaded. Interpreting 10 or more hours as 10 hours, respondents reported a mean of 1.23 hours (ie 1 hour 14 minutes) (N=765). Combining the answers to questions 8 and 9, we estimate that respondents spent an average 3.87 hours per annum accessing ESDS data.

To convert this to a cost, we estimated full cost using the TRAC fEC method for full economic costing, based on the reported averages from Times Higher Education Salary Surveys of UK-based academics (for staff £64 per hour) and graduates (for post-graduate students and research assistants £35 per hour). That suggested a mean access cost of £187 per annum across the respondents (when the additional costs mentioned by a few respondents in response to question 10 were factored in).

The question elicited a range of comments, with some praising the ease of access and others finding greater difficulty. On balance, of the 147 comments (19 per cent of responses), there were more positive comments than negative. They included:

'I found it quite easy to locate pertinent data.'

'I cannot find the relevant data from ESDS. It includes a lot of information and I don't know how to find it from mountains of categories, even [when] I use the search engine on the website.'

'Once you know your way round the system it works very well. There was a bit of a learning curve but this has been rewarded.'

'No real problems once I knew the jargon you use - would be helpful if more "public friendly" data tags could be added.'

'I found it extremely difficult to navigate the site and find data. More importantly, it was very difficult to find the survey instruments and methodological reports.'

'Very straightforward with good support once I knew what was needed.'

'Very easy to find when you know what you're looking for, not as easy to browse.'

'Excellent service - although I tend to browse - using search has not been as fruitful.'

Q11: To explore possible valuations through counterfactual means and establish the extent to which ESDS use is additional use, respondents were asked if they could have obtained the data they used in another way had ESDS not existed. Some 70 per cent said they could not have obtained the data in any other way (N=780).

The question elicited 210 comments (27 per cent of responses, but more than 90 per cent of those saying they could have obtained the data in another way), of which the majority commented on the extra time and costs that would be involved in obtaining the data elsewhere, a lower confidence in its reliability and uncertainty as to the completeness of their search. Comments included:

'Without ESDS, the research would simply not have been done. It is one of those rare assets which really are invaluable.'

'Difficult to know whether we could have found all of the data. It would certainly have been harder. We would probably [have] asked for data that we use regularly directly from government departments and the ONS (which would have taken some time, and might have had delays). It is likely that some data would not have been requested and used - reducing the amount of analysis that we can do...'

'Some of ESDS data can be accessed from other websites - some are free and some at a cost. At any rate, the convenience and ease of use of a central database provided by ESDS is a huge advantage.'

Q12: Asked if they thought they had saved time or money in a number of areas of activity as a result of using ESDS data and/or services, more than 80 per cent of respondents nominated the ability to find data from a single point of access as the biggest area of saving, followed by the quality of data (ie the level of preparation, validation and documentation associated with it) (66 per cent), and the fact that it was beyond their ability to create or collect the data for themselves (58 per cent).

Note that the responses to this question reflect primarily the researcher perspective of benefits to them personally and their own research. It will not necessarily reflect relative

benefits to other stakeholders, such as funders or institutions, from activities, such as reduced hosting or licensing costs.

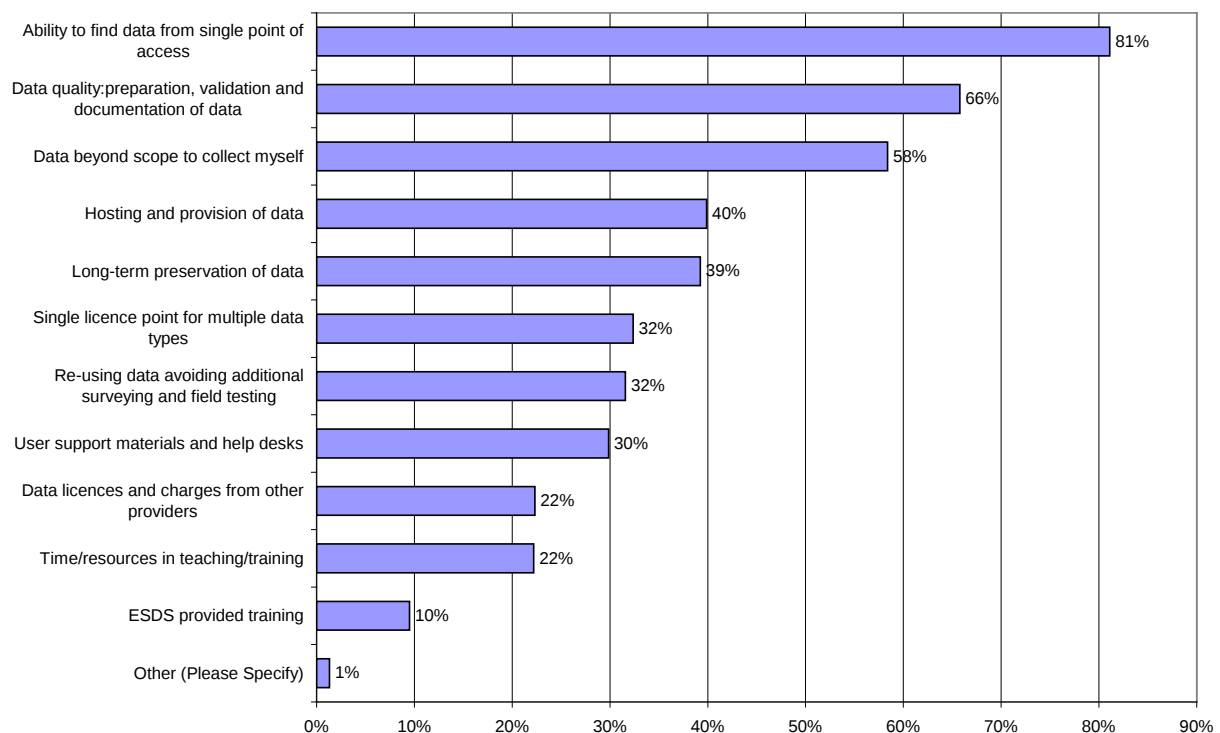


Figure 12: Areas of time/money savings as a result of using ESDS data and services (N=757).

Source: User Survey, Authors' analysis.

Q13: Asked to what extent they benefited from ESDS in a number of ways, it was clear that users saw methods and documentation as a major benefit, followed by user support (eg. guides, helpdesk, etc.) and best practices (eg. case studies and standards). A relatively high number of respondents suggested that they had not used the various supporting sources. Users of small data collections appear to make much less use of user support (49 per cent said they had not used it) than users of big data collections (30 per cent said they had not used it).

In general, we feel "haven't used" responses to this question may need to be interpreted with particular care and use could be under-reported. Many of the services listed are likely to be most heavily used when first accessing the service and less so as use becomes established. Under-graduate users (excluded from the survey population) may also be significant users. Understanding of terminology such as "tools" may also have been an issue, even when examples are provided. Moreover, some of the items may not be of use to certain users (eg. there is no reason why a researcher should use an interface tool, and no reason why someone accessing macrodata would use documentation). Hence, "haven't used" includes an implicit "not applicable" that could well be accounting for a significant share of the "haven't used" responses.

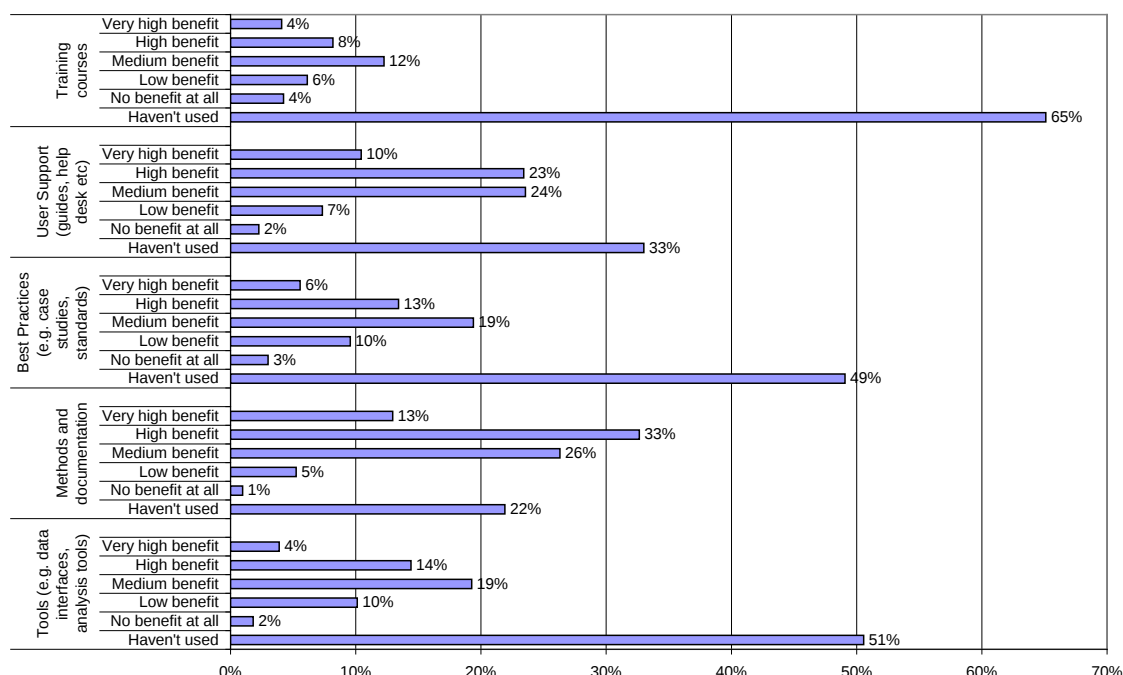


Figure 13: Extent of benefit derived from ESDS supporting sources (N=758).

Source: User Survey, Authors' analysis.

The next section of the user survey (questions 14 through 19) explored the importance of research data in the respondents' research and teaching activities, and the perceived impact of ESDS data and services on their research and teaching efficiency.

Q15: The 672 respondents (85 per cent) who said that their duties included research, were asked to estimate the share of their total research time spent creating, manipulating and analysing data in general, and ESDS data in particular, during the last twelve months.

Respondents reported a mean of 44 per cent of their research time spent working with data, with 30 per cent of respondents selecting >60 per cent of their time (N=625). The upper-bound response >60 per cent is interpreted (conservatively) as 61 per cent. The high number of >60 per cent responses may have been due to the high proportion of post-graduate students in the sample, many of who may spend their entire first 2 or 3 years analysing data prior to write-up, and a considerable number of non-academic staff survey analysts - although, the shares are not very different - with 12 per cent of academic staff selecting >60 per cent, 15 per cent of post-graduate students, and 20 per cent of non-academic staff doing so. Not surprisingly, the subset of that time spent with ESDS data was smaller, with a median response of 5 per cent and a mean of 20 per cent (N=605).⁸

⁸ During pilot testing of the questionnaire this and subsequent questions proved to be among the most difficult for people to answer and feedback had requested that they were offered ranges to choose from rather

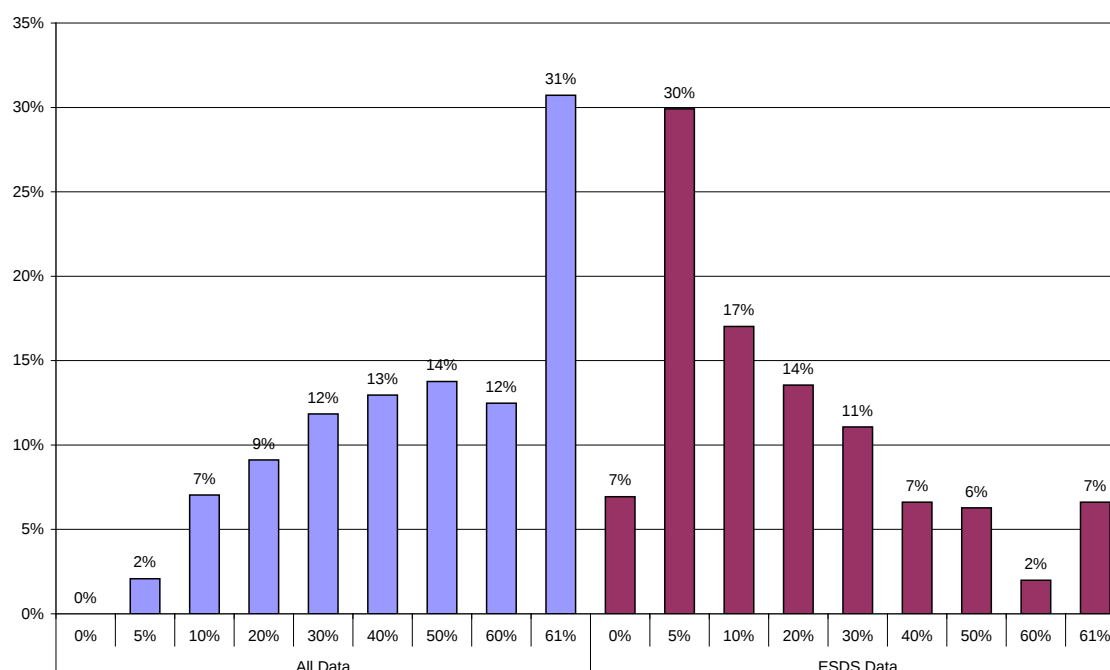


Figure 14: Share of respondents' research time spent creating, manipulating and analysing data (All Data N=625 & ESDS Data N=605).

Source: User Survey, Authors' analysis.

Q16: For the purposes of subsequent economic analysis respondents were asked for their impression of the share of time spent with data that might be typical of others in their field. They reported a mean 41 per cent of research time spent with data, in general, and 20 per cent with ESDS data.

Comments on questions 15 and 16 were distributed across commentary on the difficulty of answering such questions, and the high degree of variability of the percentage time depending on the phase of the research.

Q17: Asked if they could estimate the impacts of their use of ESDS data and services on their research efficiency in percentage terms, the mean response was a surprisingly high 38 per cent efficiency increase, with no fewer than 17 per cent of respondents selecting >75

than a blank box to record their own numeric values. As noted earlier, timescales were very constrained and interviews had to be run in parallel with development of the questionnaire. Ranges had to be developed based on the interview responses available and expectations based on them. At the expected upper limit a >60 per cent figure was used as the final cut-off range and as can be seen this proved to be too low.

per cent. Just 8 per cent of respondents reported that their use of ESDS data and services had no impact on their research efficiency.⁹

It is difficult to interpret this, as it seems unrealistically high unless some respondents interpreted the question to refer only to the project for which they were using the ESDS data. Even in that case, however, if they spend an average of 20 per cent of their research time with ESDS data and it contributes an average 38 per cent efficiency increase, the efficiency impact would be equivalent to 7.5 per cent of their total research time - itself a substantial efficiency gain.

Respondents to this question were offered a choice of percentage ranges as a result of the difficulties that we had experienced here with open-ended questions during pilot testing of the questionnaire (see also Q15). This necessitated interpretation, and we took mid-ranges as the values and interpreted >75 per cent (conservatively) to be 76 per cent. There was a minor error in the wording of one of the scales (1-10 per cent rather than 1-9 per cent).

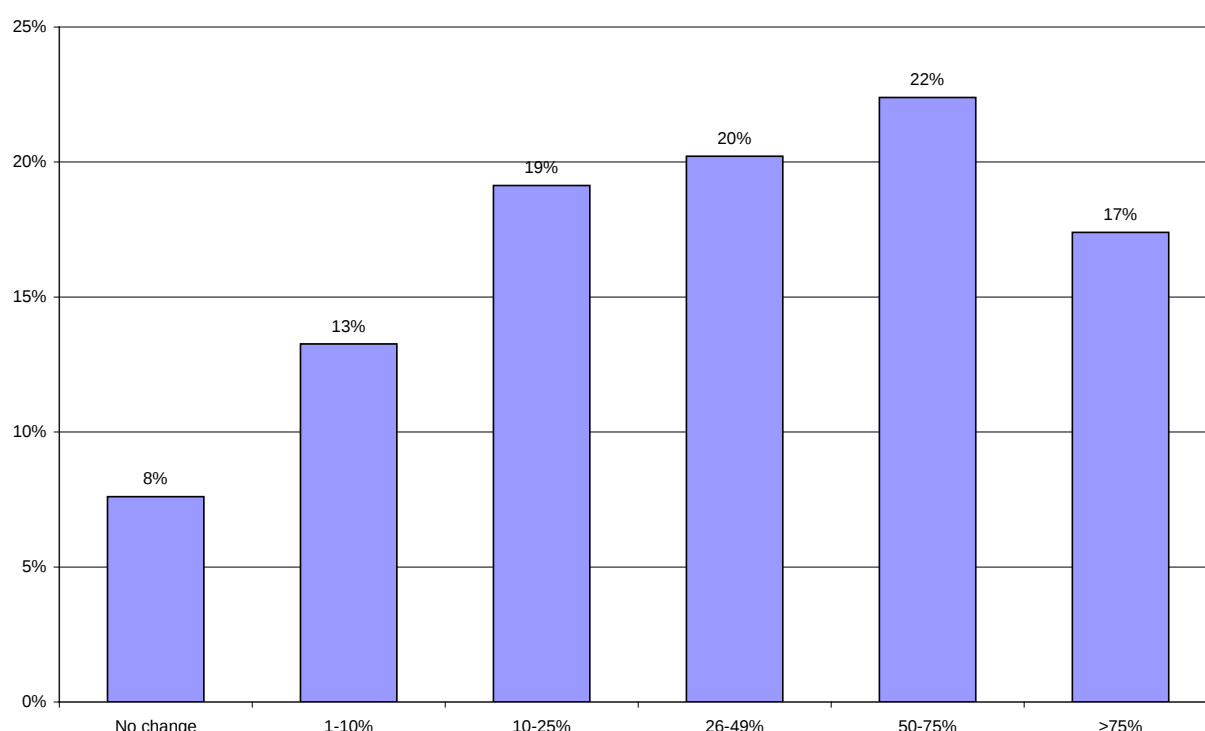


Figure 15: Impact of using ESDS data and services on research efficiency (N=460).

Source: User Survey, Authors' analysis.

⁹ One US-based post-graduate student reported that using ESDS longitudinal survey data had had a negative impact on his/her research efficiency, and no impact on teaching efficiency. The same respondent reported not having used any of the supporting tools and services in question 13, and answered few other questions.

There was a range of comments on efficiency impacts and the difficulties involved in answering such a question (47 comments in all). They included a number of directly relevant responses:

‘ESDS allows us to do analysis more quickly, and allows us to do analysis that would not otherwise be possible / easy.’

‘I have been working with data for over 20 years - I can remember when it used to be delivered on massive reels of magnetic tape and then would be sent to the mainframe computer in Manchester! Obviously there have been a lot of technological changes in that time but the ESDS has also been a very important positive development, resulting in large savings of time and greater efficiency.’

‘I have no idea how I would get my data without ESDS. It would probably involve lots of applications and form-filling and travelling to data repositories; old school data use where you had to travel to London and could only use it on site and under supervision. That would take weeks, whereas ESDS means I only needed a day before I had access to the data I needed.’

Q19: The 353 respondents (46 per cent) who said that their duties included teaching, were asked the same question about the impacts of using ESDS data and services on their teaching efficiency. This resulted in much lower reported efficiency impacts, with a mean 17 per cent increase in teaching efficiency due to the use of ESDS being reported. More than 40 per cent of respondents suggested that the use of ESDS had no impact on their teaching efficiency.

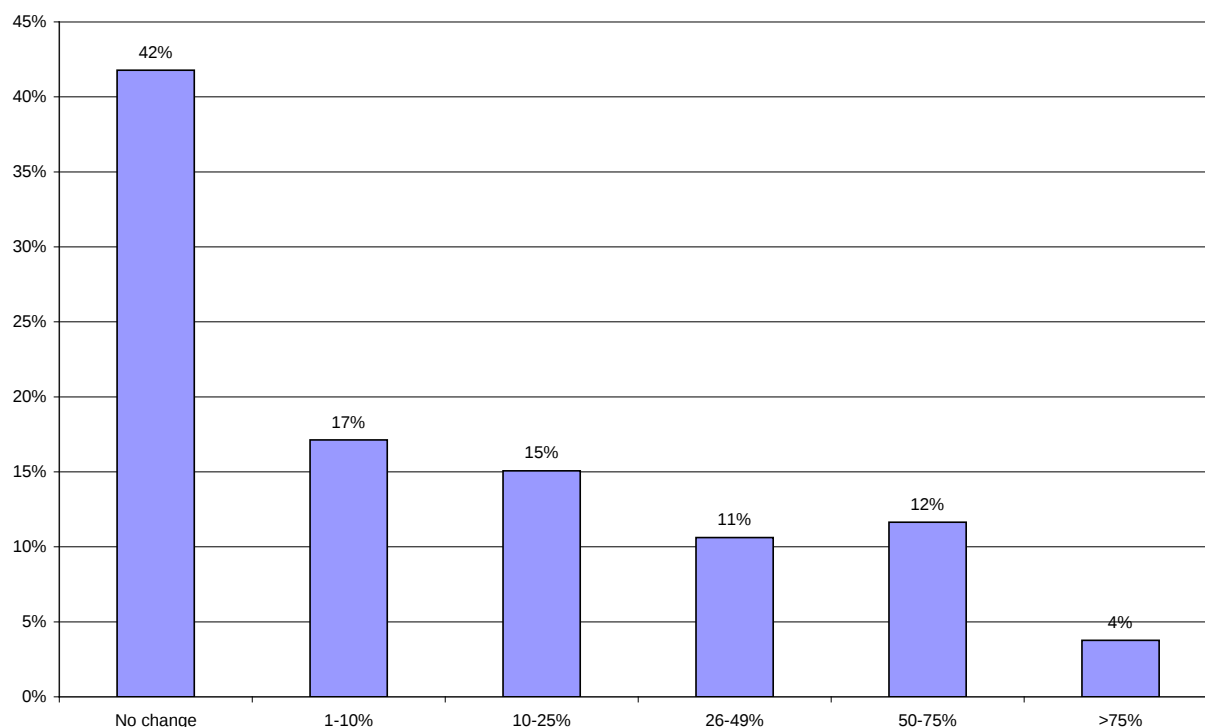


Figure 16: Impact of using ESDS data and services on teaching efficiency (N=292).

Source: User Survey, Authors' analysis.

This is probably not surprising given efforts to increase teaching of quantitative methods and use of ESDS Data in teaching are still relatively recent.

There were just 24 comments to this question, fewer than the equivalent question about research efficiency, with a number of respondents saying they had yet to, or had only recently started to use ESDS data in their teaching, so it was too early to say. Other comments included:

'In my seminars, that availability of data has increased and enhanced the quantitative dimension of students' opportunities and work.'

'The learning of data analysis techniques is brought to life using real data.'

'It's not about efficiency, but rather about being able to deliver up to date data-based examples and cases.'

'I only teach first year economics, so I rarely use real data in my teaching.'

The final three questions in the user survey explored contingent valuation in the forms of "willingness to pay" and "willingness to accept". Such questions are difficult, and there is a

considerable literature on both the method and its implementation (DTLR 2002). During pilot testing we found that many people had difficulty with the idea of "willingness to pay" when so many worked in institutional environments wherein access to information was typically handled through central institutional and/or library subscriptions. To that extent, "willingness to accept" may be a better guide to how users value ESDS data and services.

Following standard practice, we offered an open-ended question as an opportunity to explore and identify possible protest answers. There were indeed a number of protest answers, of which many suggested that: they would not pay, their university would; the questions were unanswerable; they had no idea; they would not answer or put zero because they believe data should be freely available; etc. Such protest responses are common to the approach. Weeding out the protest responses (ie 62 for willingness to accept and 44 for willingness to pay), we received the following answers.

Q20: Asked what was the minimum amount they would be willing to accept as an annual payment in return for giving up all their use of ESDS for a year, the mean of responses was £5,039 (median 346) (N=490). As ESDS hosts very different data collections, for which there are very different uses and users, it is not surprising that the responses varied widely.

A number of respondents noted that they would not accept money in return as they believe that data should be freely available, rather than because they did not value ESDS data and services. Unsure whether other zero responses were an expression of this or of the worthlessness of the ESDS to them, we excluded all (23) zero responses and found an adjusted mean of £5,287 (N=467).

Q21: Asked, in a hypothetical market situation, what was the maximum amount they would be willing to pay for access to ESDS data and services either as an annual subscription or on a pay-per-access basis, the mean of responses were £988 per annum (N=531) or £264 per data collection/access (N=522). It is interesting to note that there is reasonable agreement between the expressed annual willingness to pay and the pay-per-access willingness to pay multiplied by the annual frequency of access (ie £988 versus £1,031 per annum), suggesting that the responses were thoughtful and genuine.

Depositor Survey

The depositors' questionnaire began with similar questions about affiliation, local currency and the type of data deposited or supplied to ESDS.

Q1: The majority of depositor respondents were in higher and further education (74 per cent), 22 per cent in a range of government and non-profit organizations, and the remaining 4 per cent in commercial and other organizations. Again, the majority (95 per cent) selected the UK (GBP) as their local currency, and 3 per cent selected the Euro zone.

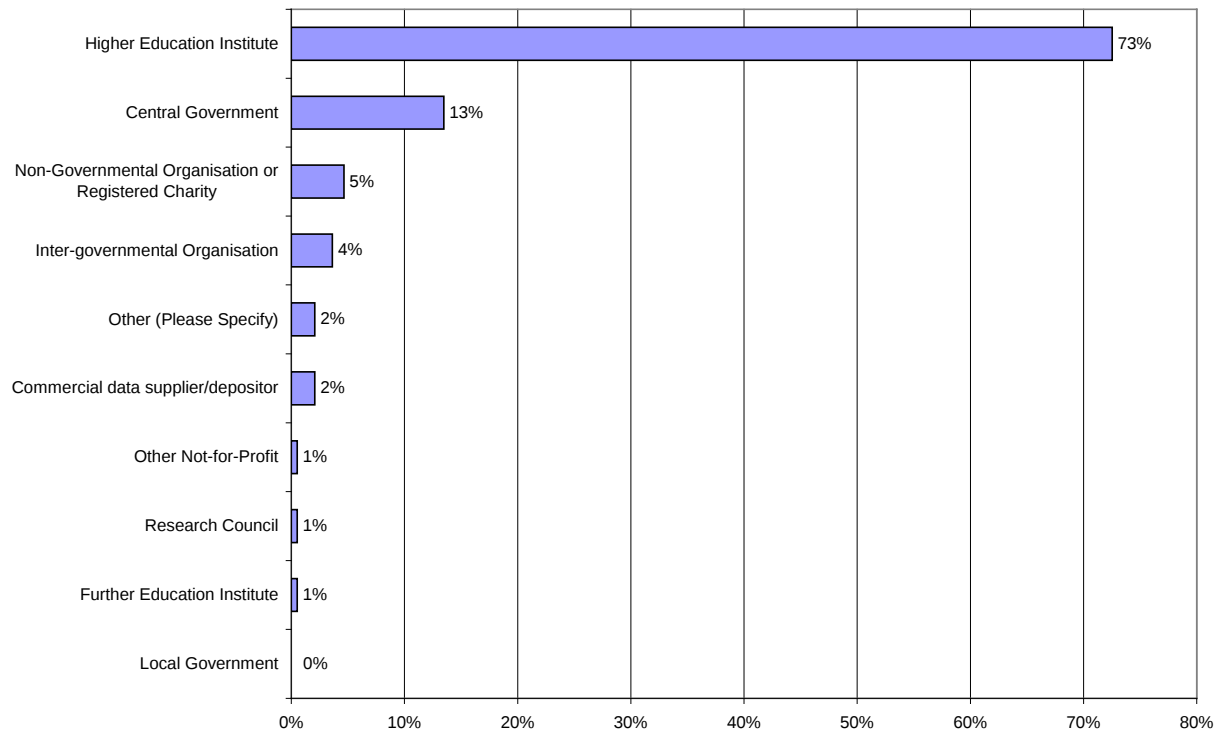


Figure 17: Main affiliation of depositor respondents (N=193).
Source: User Survey, Authors' analysis.

Q3: Asked which types of data they had deposited or supplied to ESDS, 41 per cent reported Qualitative and mixed methods data, 37 per cent Other survey data (eg. British Election Studies, Farm Business Survey or smaller ESRC-funded academic research projects), 21 per cent Large-scale government surveys (eg. the Labour Force Survey and the British Crime Survey), 13 per cent Longitudinal Survey Data (eg. British Household Panel Survey) and 5 per cent Multi-nation databanks and International survey data (eg. European Quality of Life Survey or World Bank's World Development Indicators). The 193 respondents ticked 226 data types.

While the categories are not strictly comparable and the question was open-ended as to the time period (ie asking about the data types they had deposited at any time through their history of depositing), the reported data types deposited differ from the shares of reported data collections acquired during 2009-10, with significantly more respondents reporting deposit of qualitative and mixed method data than is reflected in ESDS acquisitions, and fewer multinational and international data (ESDS 2010a).

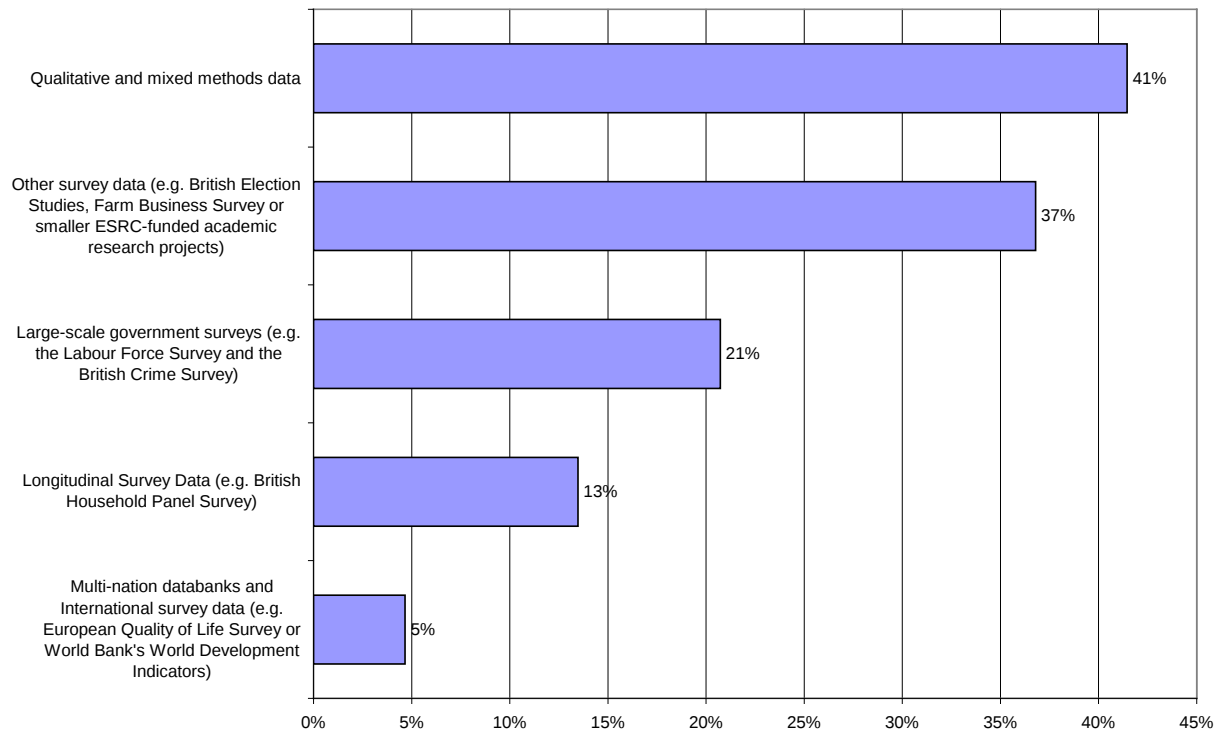


Figure 18: Types of data deposited/supplied to ESDS (N=193).
Source: User Survey, Authors' analysis.

Q4: Asked how many times they had deposited/supplied the various data types during the last three years, respondents reported some 44 deposits of Longitudinal Survey Data (eg. British Household Panel Survey), 17 of Multi-nation databanks and International survey data (eg. European Quality of Life Survey or World Bank's World Development Indicators), 88 of Large-scale government surveys (eg. the Labour Force Survey and the British Crime Survey), 86 of Qualitative and mixed methods data, and 112 of Other survey data (eg. British Election Studies, Farm Business Survey or smaller ESRC-funded academic research projects). Again, >4 deposits was interpreted as 5, so these would be minimum counts.

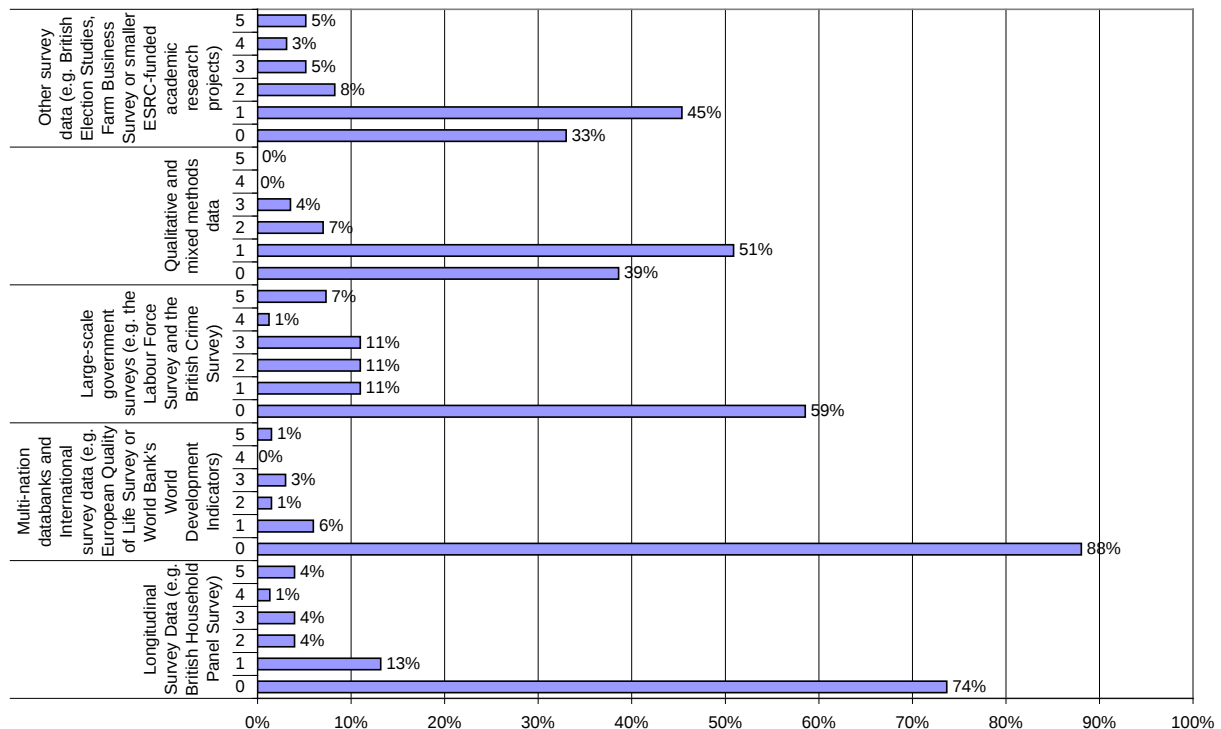


Figure 19: Frequency of deposit over the last 3 years by data type (N=177).

Source: User Survey, Authors' analysis.

Subsequent questions in the depositor survey sought information on data creation costs, and the time and costs involved in preparation for, and deposit with ESDS.

Q5: Using a "critical incident" approach, the next question asked depositors to estimate the cost of creating the last data collection they deposited. Some 64 respondents skipped the question, but the mean of responses was £768,753 (N=129). Putting the reported costs and deposit frequencies together suggested a mean annual data creation cost among respondents of £745,999.

Q7: The next question asked if this last incident was typical and, if not, whether they could estimate how much more or less than average it was on a scale of "- <10 per cent of the usual cost" through to ">200 per cent of the usual cost" - the former was interpreted as -9 per cent and the latter as 201 per cent, although there was only one incident of each. Using this adjustment to the critical incident cost to estimate an average cost suggested a mean annual average cost of £734,950.

Q8: To explore deposit-related costs, respondents were asked to estimate how long it took to prepare the data specifically for deposit with ESDS and to make the submission. Due to the diversity of data types and likely wide range of times, the question proved difficult to formulate in such a way as to make it easy to respond. Following feedback from pilot testing,

drop-down ranges were offered in months, days, hours and minutes, with ranges from 0 to 7+ for months, 0 to 25 days, and 0 to 7 hours. Working, as in the formulation of these ranges, on the assumption that there are 7.5 hours in a working day and 18 working days in a month, respondents reported a mean of 185 hours preparation and deposit time (N=152).

To convert this to a cost we used the same full economic costing approach used in the user survey, based on average academic salaries (£64 per hour). This suggested a mean cost of £11,841 per data collection or £10,796 per annum (based on cost and frequency). Not surprisingly, there was considerable difference between data types. Putting creation and deposit costs together, and taking account of deposit frequency, suggested mean depositor-side costs of £615,664 per data collection or £761,983 per annum (amongst those responding to both creation and preparation questions), or a similar £756,796 per annum (as the sum of the means from the question independently).

Open-ended comments on the time taken elicited a number of responses, with many suggesting that the time was in permission and/or anonymising the data, rather than the depositing processes itself. A number also suggested that they were inexperienced at data deposit and it had taken a lot of time to do it for the first time, but that experience and 'designing in' the expectation of depositing would make the process much easier in the future.

Q9: Depositors were asked to state the level of benefits they receive from depositing their data with ESDS. Most of the depositors reported benefiting highly from the following features: data preserved long-term, dissemination targeted to academic community, wider exposure and data more discoverable, single deposit and licence provides access to many users, and fulfilling grant obligations. They reported benefiting somewhat less from fulfilling organizational mandate. The level of benefit from training others in your data and outsourced user support was reported to be low. All attributes except long-term preservation of data mattered somewhat less for depositors of small data collections than for depositors of big data collections.

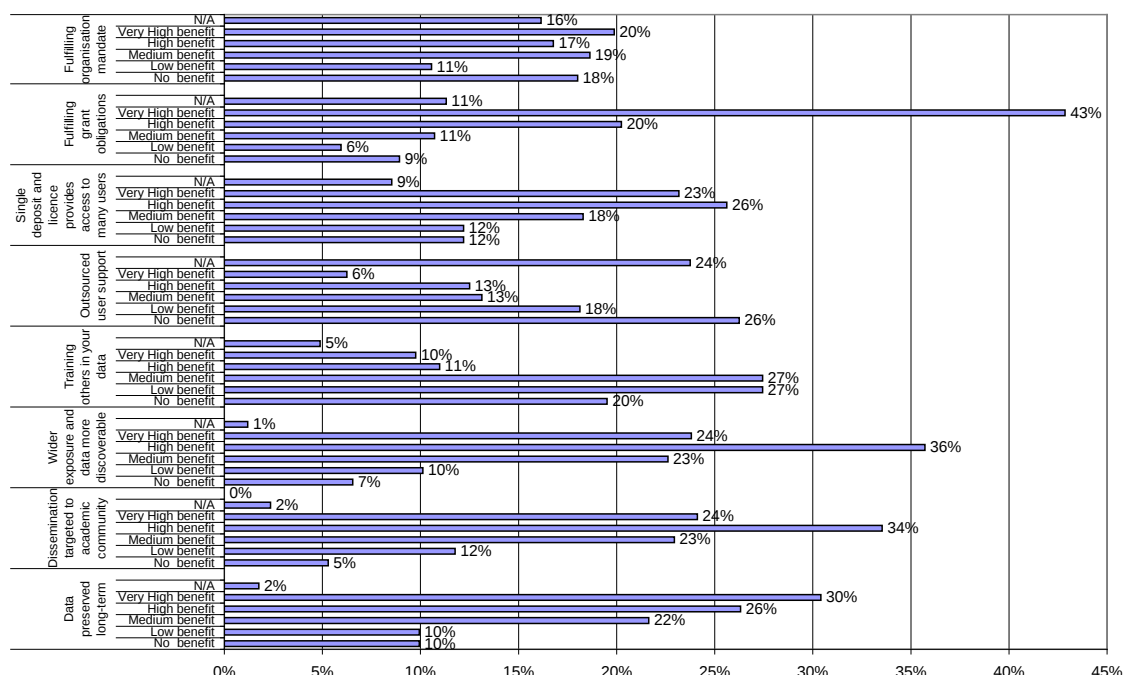


Figure 20: Level of benefit from depositing data with ESDS (N=172).

Source: User Survey, Authors' analysis.

Q10: When asked what impact it would have on their work if they could not deposit/provide data to ESDS, half of the depositor respondents reported that their work would be severely to moderately affected, 22 per cent reported a slight impact and a quarter of the respondents stated that the absence of ESDS would not harm their work at all. Depositors of big data collections suggested that they would be more affected than depositors of small data collections.

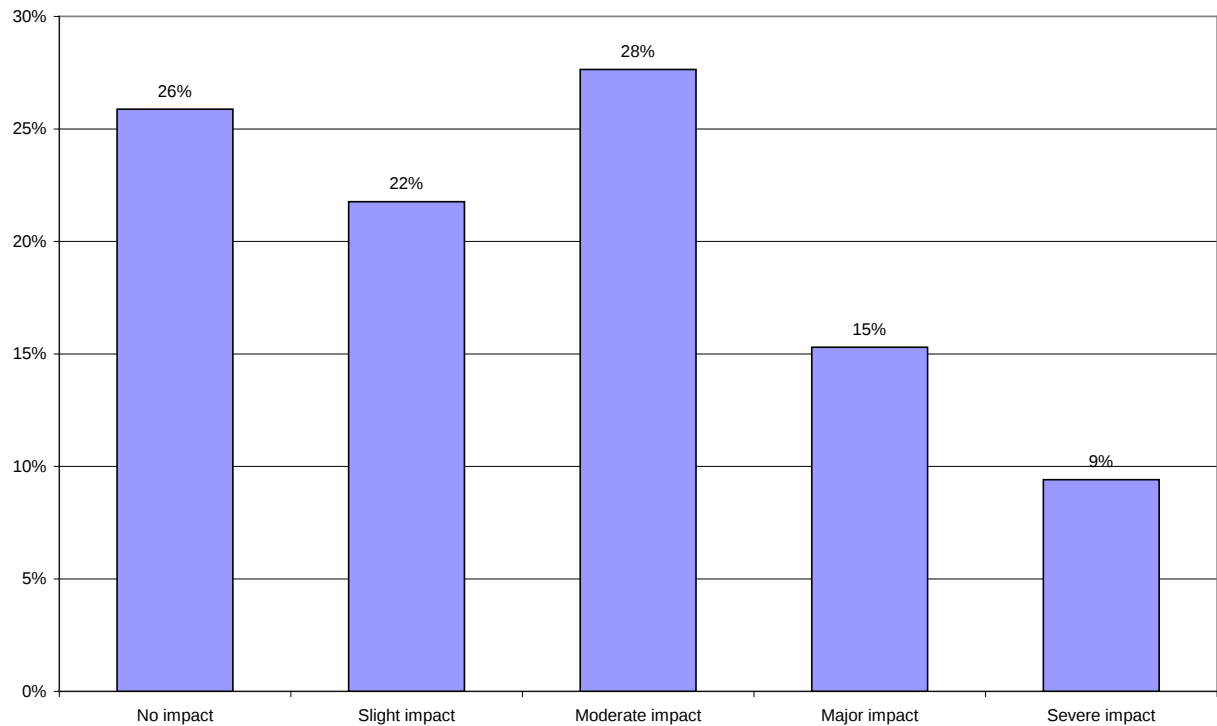


Figure 21: Impact on depositors' work if they could not deposit data with ESDS (N=170). Source: User Survey, Authors' analysis.

Appendix B: Summary of Interviews

As noted in section 2.2, twenty-five structured individual or group interviews involving thirty-three individuals were carried out with key stakeholders. The principal purpose of the interviews was to provide in-depth responses from a relatively small set of individuals that could help frame the study and the larger body of evidence being gathered via the online surveys. Findings from the interviews also helped to inform development of the survey questionnaires and the analysis. In addition they provide in-depth insights for the study or “mini-case studies” that can illustrate some of the findings in the survey and economic analysis.

Insights have been selected from interview notes and grouped under a set of relevant headings. The selected extracts are largely allowed to speak for themselves with a small amount of commentary if needed. The quotes are extracted from the study notes of the interviews (interviews were not transcribed) and the nature and role of the interviewee identified after each extract.

Defining Research Data Service Infrastructure

During interviews, ESDS staff and users emphasised that ESDS (or research data service infrastructure generally) is not solely about data: there is a broad spectrum of value-added activities.

It is worth considering in the study therefore how research data service infrastructure might be defined and how its components contribute to economic impact. On the basis of the interviews, we believe the broad four-part division of intellectual capital//human capital//organisational capital//relationship capital proposed (Hunter 2006) for intangible assets covered by digital preservation or digital curation can provide with some adaptation a suitable definition for ESDS or ESRC funded research data service infrastructure. The components of that research data service infrastructure would be:

- Research data packages [Intellectual Capital]
- Staff knowledge and skills [Human Capital]
- Technical and organisational environment [Organisational Capital].
- Disciplinary and professional networks [Relationship Capital].

The Value of ESDS to Users and Depositors

Value for research

‘She emphasised again that it would not be possible to do her research without the datasets accessible via ESDS. If these were available, but each from a separate source, it would require the efforts of an administrator to access each one individually. This would cost possibly £15,000 per year – that’s just to access data. The Swedes have a register system – you pay separately for each longitudinal dataset. In Sweden administrative data is linked so people have not necessarily consented for sharing their details – this means an ethics

committee has to give permission for each dataset request – that costs around £5000. Then a register has to be constructed for each project – this costs £9,000. No re-use of this data is allowed. Therefore, if you had a new idea – such as the factors which create badly behaved youth – the required dataset would cost £14,000 (ethics and dataset construction) plus the associated administrative outlay - and would take a year. In most countries this would be impossible. It is cumbersome and expensive [compared to assessing data via the ESDS].’

Academic research user (Medical Sociology)

‘He would be hugely affected [if he did not have access to ESDS data and other services], mainly because of the portal nature of the service. Although he feels it would be possible to obtain what he needs from elsewhere, it would be ‘more problematic’ to get it at source – in terms of finding it, possibly negotiating access etc. He has very limited time, and the two days he mentioned as working with ESDS data would be multiplied considerably if he had to go elsewhere (as he hadn’t needed to explore other sources he couldn’t really put an exact figure on the extra time required).’

Academic research and teaching user (Social Sciences)

‘The online portal works well – there are no requests, persuasion etc. Without ESDS, those costs would be incurred in negotiating access from eg. a European foundation. ESDS saves transaction costs – correspondence and speed of access. In response to a request on how much time (and therefore money) he would spend without ESDS he calculated this using the Chinese Stockmarket data example. For this he needed to register, acquire a user name, learn the system. There might be correspondence (paid) involved in other circumstances. In all, it may take three days as opposed to 5 minutes to obtain and download [ESDS] data. He later qualified this, saying that the estimate for ESDS of only 5 minutes was probably an exaggeration. He explained that if one were to assume no prior registration with ESDS and no knowledge of their systems (ie akin to the CSMAR situation) then one might estimate half a day to register, find a dataset and download it. Of course, in his case he did have knowledge and had registered, so his answer was still legitimate. Costing that difference at £650 per day, would add up to around £2000 [additional cost compared to ESDS].’

Academic research user and depositor (Economics)

‘It would be impossible to do his job without ESDS – he made the point (quite forcefully) that ESDS was a monopoly provider of the datasets he uses, and thus his whole research output was reliant on ESDS (minus the data which he himself deposits).’

Academic research user and depositor (Economics/Social Policy)

‘Thinking about the last time you accessed ESDS data how long do you estimate it took you to find and access it? Only one or two minutes. It was the Annual Population Survey. He knew about the survey and where it would be on the website. However, the short access time would be typical. Even when he doesn’t know what dataset he needs it is extremely quick to find one, using the search facilities. If there was no ESDS and the government did not deposit datasets, his work could not be done at all. The cost for each ‘one-off access’

would be substantial – several thousand pounds. It would also take a long period of time. Calculating any access costs and the time taken to negotiate licences etc. he estimated ‘maybe’ £5,000 - £10,000 (assuming access could only be given by the data providers themselves). On the other hand some data providers could make their data freely available, so it depends on the data required and the potential provider. However, the government has a commitment to provide data: if there was no ESDS it would have to function itself as an ESDS substitute, such as via the ONS...

...What would you do and how long would it take if ESDS did not exist? He calculated three months – if you knew there was a dataset and you were trying to get access. He offered an interesting parallel. This was the way the European Commission Eurostat ran the European Community Household Panel Survey (ECHP). It issued licences to get access. You had to apply, pay a lot of money, and give assurances about data use. Even after that was undertaken, the data would take weeks to arrive and not be in good condition (ie poor meta-data, labelling etc.) He said to be fair this was now a bit easier. However, this is a good illustration of what would happen if ESDS did not exist. ‘

Academic research and teaching user (Sociology)

‘For his own research too, ESDS is very important. He can do things a lot more directly now, by using a centralised source. Having said that, he used to have a subscription (£150, paid out of general research funds) to access International Financial Statistics (IFS) data, which provided him with over 90 per cent of the resources he needs, although with less documentation and harder interface.’

Academic research and teaching user (Economics)

‘Datasets are cleaner and better with ESDS. There is good quality control. It is also a one-stop-shop. All this saves time and enables more concentration on actual research.’

Academic research and teaching user (Economics)

‘It is very difficult to know what the comparisons might be – ie what additional preparation if any might be needed for deposit with ESDS, as they prepared the data for archive and re-use anyway (she conceded that people don’t always document data properly, but in her case it was a very important part of her work) - for complex datasets you need to put all the work in that you would for deposit with ESDS anyway.’

Academic research depositor (Education - Longitudinal Data)

‘Only ever been a depositor, and then only once. She has made one deposit, in the Qualidata archive. The main issue she found was what could and could not be put in (in terms of ethical permissions etc.) because of the nature of the project. The only part of the project deposited was the focus group transcripts – but that represented only a small part of the project. In addition, the project included big public events, individual interviews etc. It would have been very difficult to put all of the data together for anyone else – it was very specific to the project and the time required to index and document it would have been prohibitive.’

Value for teaching and training

'He only needs to actually download data twice a year. However, once downloaded he both manipulates the data and then puts it on the student VLE for use by students, who use it extensively on their statistical courses. Regarding the former, he tends to cut down the data so that it is more manageable for the students. He gave the example of the BSA dataset. He said this had over 850 variables, which he cut down to 75 for teaching purposes. This involves assessing which variables would be the most appropriate and usable for his students. To do this he said he needed basic SPSS skills. The work took around two days...

...He has a "class licence" for his teaching so he can provide all ESDS material for his students without any negotiation. In addition, if his students want to study a topic for their dissertation on a subject about which there are datasets in ESDS they can also register themselves. In the past the students used to have to use an Athens password – and some of them could not remember these. Signing on to ESDS is far simpler...

...In addition to the teaching sets and other documentation, ESDS also provides NESTAR which allows the students to visualise and manipulate data without having any specialist knowledge of SPSS. He uses this a lot with undergraduates and feels it to be a very real benefit. Finally, the search engine is a very good resource. For the first class sessions the students discuss various topics with him, from which they develop appropriate search terms. They enter this into the ESDS search engine to find which datasets are available which contain information on the topics chosen. He says that there is always an appropriate dataset.'

'She encourages all of the students whose dissertations she is supervising to explore the archive. She said this gave them a way of "x-raying" a piece of research, how it was designed, what the outputs look like etc. The ESDS provides all the contextual information surrounding the research, which is very useful...

...The assessment on one of her courses would collapse without ESDS data. This is the Advanced Qualitative Research Unit, where data analysis is needed. The students are required to produce a 4,000 word report, and have to select a particular dataset to analyse. She provides small subsets, and they have to do original analysis of data on those. Using ESDS data not only gives them opportunity, but the whole unit would fall apart without it. She would have to ask colleagues for data or for databases. This would make too many demands on her time and would prove inconvenient for her colleagues...

...Although the [download access] figures are small, once they are downloaded they can be used extensively. She also made the point that she is a conduit for 40 people getting access. She registers for the students and they all use and manipulate the data. In terms of UG student numbers: I would say that 60 first year students are encouraged to use it. 60 second year students are encouraged to use it and 55 final year students are encouraged to use it. I

only act as a conduit for PG students in my advanced qualitative unit- where 25 or so students need access to materials...

...if she didn't have access, it would take far more time. The data is all in one place, the quality is guaranteed, the ethical permissions have been sorted out, ESDS have cleaned it. She said all that was very important. She couldn't put a number on the saving, however...

...The external examiner for her masters programme has identified her course as noteworthy and has really liked the assessment – and the use of ESDS Qualidata in the teaching. Without this it would be impossible to run the course as it is...

...She felt that the ESDS datasets constituted 'an archive in itself', representing a history of social science research and data gathering. Thus it had an intrinsic value. She actually gave some examples. Townsend's work on poverty has become a part of social history. Similarly Savage's work on class has given us an understanding of the issues that has now been built into popular discourse. She tries to impart all of this to her students – in other words, the value of the archive goes beyond individual datasets.'

Academic teaching user (Sociology)

'Most ESDS surveys, however, are too complex for UG and even PG students – so he takes a subset with a narrower range of variables, which he re-labels to aid comprehension. Also useful are the teaching datasets ESDS have invested in...

...He also compared ESDS-accessible data to a hypothetical data archive without the front-end provided by ESDS. Without the signposts provided it was difficult to locate data. ESDS makes people aware of what there is and how to get hold of it, which makes things infinitely easier. ESDS has growing range of teaching resources, so yes, beneficial in helping students understand the structure of data, what is available and how to interrogate it.'

Academic research and teaching user (Sociology)

'He runs a 2nd year course, the objectives of which include subject knowledge and understanding; accessing and using databases, and transferable skills. ESDS International data is used in this. Students are required to undertake an activity in which they examine consumption, investment and imports for a number of countries (up to nine). A large part of their work involves rearranging the data – getting the columns into national currencies, calculating log values to standardise it, and calculating growth rates. ESDS is very important for this, as it provides all the data in one location – the students don't have to look at different national statistics databases, or log into different systems or services. The main benefit is in offering the students the transferable skills they will need when they leave university, principally in how to interrogate databases, learning what material is available and the accompanying documentation...

...The effect of ESDS is far bigger on teaching, where non-access to ESDS material would result in at least twice the time spent by students in accessing the material elsewhere and therefore less teaching which would be more rushed – hence both efficiency and quality

being affected [for a numerically large number of students]. In his research, the quality would not suffer too much, because of the availability of the data elsewhere.'

Academic research and teaching user (Economics)

'His job is very teaching focused – he teaches a lot of courses. For example he has 300 students on an economics introductory course, for which he uses databases giving world development indicators and a 2nd year course on economic development, which uses these also. The big one however, is being in charge of UG and some Masters dissertations, all of whom do applied work which requires data available via ESDS to the extent that they have a talk from an ESDS staff member. The data they tend to access is from IMF, OECD, some mix and match; world development indicators, ONS statistics etc...

...He made the point that he viewed much more than he downloaded – approximately 50 per cent. Partly this was checking data and its relevance, and partly because he might only need to check its format, for use in teaching or even to obtain a small amount of information for which data manipulation was not required...

...He felt that the time taken to access data from other sources would be detrimental to his teaching in all these areas. Also, the documentation in ESDS is very good, which he uses and thus adds to the quality and, he feels, novelty of his teaching (no other database gives such good documentation and he would not have time to explain it all etc...).

...Before, every individual student had to register with each provider, which entailed completing a form, sending emails etc. Even having done that, usage was far lower than for ESDS because it took so long to log in. He estimated that the time taken to access what they do on ESDS was less than 10 per cent of what it would otherwise have been...

...Another point was that in classes he makes assessments on how well students can do things – this is a lot easier with ESDS as they have a range of data they can access without all the 'hoops' to climb through. Thus, he can look at their accessing, interpretation and use of various data without waiting for them to constantly access and log on to different services...

...Because it is benefitting master's students yes it is very useful and will have positive impacts in the future. This is because their understanding of the issues inherent in economics is now more connected with data, and so the approaches are more data driven. This will have important impacts when these students begin their professional work – being able to approach and solve problems on evidence...

...The teaching and general documentation is better on ESDS than on provider sites (eg. such as the IMF - which isn't at all as clear) which helps in teaching and in the students' understanding of the services. Another benefit was that students are exposed to what it is possible to access – even if they don't need everything on their current course. It all helps in their future careers.'

Academic teaching user (Economics)

Value for service provision

'His funding and work require him to deposit data. Indeed, he collects them specifically to be used by wide range of other researchers – not for his own primary research (for which he uses Government datasets eg. labour force survey)...

...ESDS are better at publicising it than his group would be, and they are also better at data security – they can invest more. He guessed that usage would fall if ESDS wasn't there by 'maybe' 20-25 per cent...

...By depositing with ESDS there is heavier usage of the data, which helps publicise his work...

...More people will find the data if it is released through ESDS, because it is a single point for such survey data so people who do not know about our data sets will find out about them because they go to ESDS for other reasons or because they have a research interest and the ESDS data discovery tools will allow them to identify it as the most appropriate data set. In other words we would have to do more promotional work as well as direct work on data dissemination and it is less easy for us to target the whole of the audience.'

Academic research depositor (Economics and Sociology - Longitudinal Data)

'He is currently undertaking secondary analysis for the government, for which the ESDS is a convenience rather than a necessity – he can access the data online and doesn't have to bother the government for the data – this would involve an administrative process including requesting it, negotiating a licence etc. In addition the government would have to package it up. Thus, the value of ESDS data is that it solves any licensing issues etc. The Government benefits too – it is able to put data in to what is effectively a library (ESDS), with the latter bearing all the costs...

...Do ESDS save any work? Yes – as mentioned, one doesn't have to deal with users (registration, correspondence, licensing). The cost of this would be enormous. It would take the work of a librarian. ESDS makes it easily accessible. The cost of preparing is all is on him, but after that ESDS pays for the curation, access and other aspects of the dataset. His guess is that he would have to employ somebody two days a week, for two years, and calculated around £10-15k per year.'

Academic research user and depositor (Economics)

'ESDS is basically an aggregator of data which it provides on to its clients. This arrangement is good for us as it means that we can negotiate and work with only one customer but reach several clients...

...The main advantage is that data is not lost. Much research funding requires researchers to deposit data with ESDS. The UK is doing a good job in respect of data gathering and archiving. There is no equivalent to ESDS in other countries, so lots of data gets lost or obscured in universities [elsewhere]...

...There is, in his view, a question about whether ESDS should continue to gather and host data (if they don't add value) when it is already freely available – such as with that offered by the World Bank. It would be easier just to make a link to the original dataset.'

Non-academic depositor (International Data)

'Differences between ESDS and the local archive we have created: In ESDS you download data at the project level. The data is beautifully stored but not well organised, because you can't do refined searches. This implies a lot of work for the individual researcher, who has to dip into various projects to see what is there – it needs a process of serendipity. ESDS provides infrastructure to preserve data for reuse, but patterns of data sharing don't necessarily map onto big national database. The local archive is a specialist resource, very successful, providing refined research tools. They have set up a particular template whereby data sharing is more embedded in the research process. Data management planning is offered by the team, who receive many enquires on the subject. She felt that there would still need to be a national infrastructure but the question is what that infrastructure does. It may be more of a Gateway in future.'

Academic research depositor (Sociology and Social Policy - Qualitative Data)

Value to Innovation and the Profession

Consumer champion

'As an organisation it is making it easier to use data and acting as a customer champion. Previously there was no organisation designed to make it easy to use data...

...Why use ESDS if the data is available elsewhere? Sometimes the speed of supply, you know what they have at ESDS. Sometimes it is useful to have an alternative supplier. I have no sense that ESDS is anything other than helpful and they know well what they've got. In a government department people are concentrated on the demands of the department, not helping external parties to understand the data. At ESDS their function is to help people use data which encompasses customer service skills.'

Research user (UK Government)

'He also noted that the relationship between ESDS and data providers such as the ONS; government etc. helped make things clearer as to needs of academic community – another advantage of having the ESDS.'

Academic research user and depositor (Sociology)

Centre of excellence

'ESDS staff were also concerned we did not overlook the impact from their status in the profession in terms of professional visitors to the archive and ESDS seeking to learn how to set-up a data archive, leap-frogging early mistakes/set-backs etc. They also noted the professional network that ESDS maintains internationally also has a major impact helping to develop international infrastructure and access for UK users. Knowledge transfer to users

from this work is also done to users via their newsletters etc. In a similar vein ESDS receives significant benefits from University of Essex and UK Data Archive R&D project income outwith the core ESRC grant. This research also has an impact.'

ESDS Staff

Innovation in practice

'Can't put a figure on it but the change in use has been quite rapid. There's a strong sense of innovation. The change is accelerated by cutbacks in the rest of the public sector. ESDS is a unique access point, a one-stop-shop for data...

...The speed of deposit from major data suppliers to ESDS is getting much faster- change to the good. Now depositors feel incentive to deposit at earlier point to stop people ringing them up.'

Research user (UK Government)

'She feels it a pity that ESDS is only available in the UK, other universities around Europe could benefit – Eastern European countries especially.'

International PhD student user (Economics)

'He added at the end of the interview that one of the significant economic benefits is to have created a social norm (although ESRC are behind data creation) as a result of that process data is made available, and the overall economic impact of this will be massive as more and more becomes available. If ESDS had never existed and supported, it seems clear to him that most data available would not be, because the norm would not have been created. This is the case in other countries. The default now is to make data available, which can only be a good thing.'

Academic research user and depositor (Economics)

'She gave an opinion on how much she felt research has changed. Data is readily accessible now, and with strategic direction from ESRC people more willing to use and learn about secondary qualitative data – there has been a huge change...

...There are secondary analysis initiatives from funding bodies. However, she confirmed that data re-use is still limited, and that there is still not a huge culture of sharing and re-use.'

Academic research depositor (Sociology and Social Policy - Qualitative Data)

Quantitative methods

'Another aspect is having a place academics can go and conquer their fear of quantitative analysis. So few people believe it is possible to use the data, ESDS acts as a kind of translation tool. It makes it available in a psychological sense for students to learn quantitative methods from the data and documentation.'

Research user (UK Government)

'The problem is that so few people have the skills to use the datasets (although they are easy to access and retrieve data from). Some datasets are better set up – BHP study is very good, well labelled etc. But it has 20 sweeps so you have to use a person-years approach (eg. during the span of a dataset people reach particular ages – eg. they were all aged 16 at some point. This examines what happens to people of that age, be it in 1960 or 1970. Many people are not able to do that – they need training. There is no point in continuing the dataset unless people can use it. A particular example is the question: “Is divorce less harmful to mental health in younger people?” There are lots of quick conclusions you could draw but, but it would be easy to draw the wrong conclusions. This is because other factors may be in play. This shows the requirement for people to understand other factors that you need to include in these studies.'

Academic research user (Medical Sociology)

'Teaching is a major part of his role. He teaches a number of courses on social statistics, including “Introduction to Quantitative Methods”. He finds that his students come to him with a dearth of knowledge about statistics and poor statistical skills. The best way he finds to teach them is to use real datasets – the great majority of which he accesses via ESDS.'

Academic research and teaching user (Social Sciences)

'Capacity building – the quantitative methods teachable via ESDS are important. ESDS makes it easy to get hold of empirical evidence. Before you were entirely dependent on what individual bodies chose to publish.'

Academic research and teaching user (Sociology)

Value to Commercial and Government Users

'The crucial thing is time: the typical deadline is a day or two to come up with answers – for that reason they tend to have to work from existing analyses.'

Research user (UK Government)

'Private companies are not using ONS confidential data –by and large using their own – are they unaware of the opportunity, or is there no benefit in it for them? If it has wider economic impact why isn't there a queue of commercial users at UKDA? May be it is something in ONS data itself – it is not transactional and current – there is a delay until something can be accessed. Time lags are an issue for business. He gave the example of the research for banking sector review in 2009 which had to use data from 2006.'

Depositor (Government Data)

'He made it clear that his non-use of ESDS [as a commercial user] did not degrade the value of the service - universities find great value in their services - just that it was not appropriate for his purposes. He felt that the nature of the resources available via ESDS is not what

many commercial organisations require. He gave the example of an academic who makes time and effort to get the detailed raw data in order to reanalyse or repurpose it. By contrast, a commercial organisation is far more likely to want 'quick information', say on crime etc. and that as such would be unlikely to spend time going through the process of obtaining the source data and analysing it themselves. They require a digested version which summarises findings. He is guided by what commercial clients want, they want 'quick wins' so he shows them data that is easily obtainable. However, if such summary data proves to be of value, there may then be interest in digging more deeply. If the resources available could be made simpler – such as by offering Excel spreadsheets of summary findings - markets for ESDS would expand. An example is the Expenditure & Food Survey – if it could be produced in summary Excel form this would promote greater usage by a wider range of users. This need not be undertaken by ESDS but by the originating researchers – the government department or academic institution. He felt that the ESDS was vital as a repository for government data because if they don't do the job who would?

Non-user (Commercial Sector)