

AUTHOR ACCEPTED MANUSCRIPT

FINAL PUBLICATION INFORMATION

Developing Statistical Literacy Using Real World Data : Investigating
Socioeconomic Secondary Data Resources Used in Research and Teaching

The definitive version of the text was subsequently published in

International Journal of Research & Method in Education, 34(3), 2011-10-13

Published by Taylor and Francis

**THE FINAL PUBLISHED VERSION OF THIS ARTICLE
IS AVAILABLE ON THE PUBLISHER'S PLATFORM**

This Author Accepted Manuscript is copyrighted by the World Bank and published by Taylor and Francis. It is posted here by agreement between them. Changes resulting from the publishing process—such as editing, corrections, structural formatting, and other quality control mechanisms—may not be reflected in this version of the text.

You may download, copy, and distribute this Author Accepted Manuscript for noncommercial purposes. Your license is limited by the following restrictions:

- (1) You may use this Author Accepted Manuscript for noncommercial purposes only under a CC BY-NC-ND 3.0 Unported license <http://creativecommons.org/licenses/by-nc-nd/3.0/>.
- (2) The integrity of the work and identification of the author, copyright owner, and publisher must be preserved in any copy.
- (3) You must attribute this Author Accepted Manuscript in the following format: This is an Author Accepted Manuscript of an Article by Carter, Jackie; Noble, Susan; Russell, Andrew; Swanson, Eric *Developing Statistical Literacy Using Real World Data : Investigating Socioeconomic Secondary Data Resources Used in Research and Teaching* © World Bank, published in the International Journal of Research & Method in Education 34(3) 2011-10-13 <http://creativecommons.org/licenses/by-nc-nd/3.0/>

Developing Statistical Literacy Using Real World Data: Investigating Socioeconomic Secondary Data Resources used in Research and Teaching

Jackie Carter (Mimas, University of Manchester)

Susan Noble (Mimas, University of Manchester, UK)

Andrew Russell (Social Sciences, University of Manchester)

and

Eric Swanson (World Bank Group)

jackie.carter@manchester.ac.uk

susan.noble@manchester.ac.uk

andrew.russell@manchester.ac.uk

eswanson@worldbank.org

For Correspondence contact:

Dr Jackie Carter

Mimas, University of Manchester, M13 9PL

e-mail: jackie.carter@manchester.ac.uk

Developing Statistical Literacy Using Real World Data: Investigating Socioeconomic Secondary Data Resources used in Research and Teaching

Abstract

Increasing volumes of statistical data are being made available on the open web, including from the World Bank. This 'data deluge' provides both opportunities and challenges. Good use of these data requires statistical literacy.

This paper presents results from a project that set out to better understand how socioeconomic secondary data are being used in helping to develop quantitative skills in UK university social sciences, and focuses on World Bank data in support of this. Evidence of real world data use to develop statistically literate social science students is presented. This underpins capacity building in UK social sciences, a need recognized by research councils, statistical organizations and other bodies.

The work has helped uncover evidence of good practice in teaching and research to inform others who wish to support their students in becoming statistically literate using real world data. The challenge is to share and replicate this good practice (mainly in economics and econometrics) in other areas of social and educational research which are less well served.

Keywords: *statistical literacy, building statistical capacities, teaching and learning, social sciences, secondary data.*

Introduction:

The amount of statistical data available on the open web is increasing exponentially so that individuals now face a 'data deluge' (The Economist 2010). This provides both opportunities and challenges for researchers and learners in Higher Education. In order for citizens to be able to make use of this data they need to understand both the social construct of data - where data come from, how they are collected, whether they are comparable with other data and consistent over time – and have the skills to interpret and analyze them. The use of real world data in teaching and research can be seen as a real boost to Higher Education since such data are authoritative, abundant and applicable to life outside (or indeed after) academia.

The increased access to global statistics has spilled over into popular culture. In December 2010, BBC4 broadcast a programme *The Joy of Stats* presented by Hans Rosling. It provided an entertaining, thought-provoking, and graphical view of statistics, with Rosling's dynamic

visualisation of lifespan and income in 200 countries over 200 years in 4 minutes (plotting over 120,000 data points) receiving over 3.5 million hits on YouTube (Rosling 2010) provoking considerable debate online and in the twittersphere. Raising the profile of statistics in this way is an indication of the power of statistics and visualisation methods in telling stories with and about data, taking it from the academic to the mainstream, although as Rosling himself noted,

Having the data is not enough: I have to show it in ways that people both enjoy and understand.

Rosling 2010 (0:10 into recording)

Academia is one of many sectors in society where statistical literacy (Schield 2004, Bolton 2010a; 2010b) is increasingly important. The need to improve statistical literacy in both the academic and wider population is recognized by research councils, statistical organizations and other bodies. The prospect is that more data will become available both nationally and globally as development agencies such as the UK Department for International Development (DfID) and the World Bank, one of the world's largest International Financial Institutions, are supporting programmes to increase the quality and quantity of statistics produced by developing countries. More important, both are also supporting open access to these public data.

This paper presents results from a project that set out to better understand how statistical data are being used in helping to develop quantitative skills in the UK university social sciences, focusing on the use of data accessed from ESDS (Economic and Social Data Service) International, a UK online academic data service, and from the World Bank. This contributes to the field by presenting evidence of real world data use in UK Higher Education to improve statistical literacy among social science students and by contributing to capacity building in the sector.

Work on use of secondary data sources for educational research is scarce (see Smith 2008). Although texts exist for some social science disciplines (e.g. for Sociology see Fielding and Gilbert 2006) there is concern about statistical literacy, or quantitative skills, in social science subjects (MacInnes 2009). However the delivery of these skills is thought to be uneven in the social sciences with economics and psychology tending to be better served than other disciplines in the area. The definition of social

sciences is somewhat inexact but any list should include behavioural studies such as those found in Anthropology, Economics, Education, Geography, History, Law, Linguistics, Political Science and Public Administration, Psychology and Sociology. The findings of this work could be of interest to those in the social sciences who might potentially use secondary data resources in their teaching or research, or researchers in other disciplines who are interested in analysis of socioeconomic, secondary quantitative data (Cole *et al.* 2008).

Background and Context

Quantitative Methods Skills Deficit in Social Sciences

The (UK) Government designated quantitative social sciences as a strategically important and vulnerable subject when it identified a skills deficit in quantitative methods

HEFCE 2008

Although the cutting edge of social science research is dependent on use and analysis of real world data there is concern in the UK that there is a lack of highly trained quantitative researchers and teachers. The UK's Economic and Social Research Council (ESRC), as part of its national strategy to build a strong research base in quantitative methods in UK social sciences has been working for several years to address the skills gap through its Quantitative Methods Initiative (ESRC 2011), with an aim:

to combat the skills deficit in quantitative research methods across the social science research base [recognising that] there is a need to enhance quantitative skills across the full breadth of the 'educational lifecourse', from building new capacity at the undergraduate level to refreshing the quantitative skills of mid-career academics that teach undergraduates and supervise PhD students ('training the trainers').

ESRC 2011

Moreover, there is a great desire to improve research-led teaching in this area and build on the ESRC's investment in secondary data resources. In January 2009, the ESRC appointed a Strategic Advisor on the Undergraduate Teaching of Quantitative Methods in the Social Sciences with a brief of developing a coherent programme of activity aimed at enhancing undergraduate teaching of quantitative methods across the social sciences and producing a report of recommendations (MacInnes 2009). Recent efforts have focused on building statistical capacities in the undergraduate as well as the postgraduate population.

The UK provides access to a rich set of social science data resources with ESDS regarded as a jewel in the crown for the UK's social science data community. The challenge for educators in addressing quantitative literacies lies in promoting students' use of data, but the benefits in doing so can improve both academic performance and job prospects for students.

Carter 2010a.

The work reported here is therefore apposite, even more so as budgets in social sciences are under threat (Campaign for Social Science, CFSS 2011) so that social science faculties find themselves in need of delivering critical analysis skills, including the ability to solve quantitative problems, to their students in increasingly efficient ways.

Consequently, this paper has been driven by two interlinked developments; the identification of a quantitative skills deficit in UK social sciences and the remit for UK research council funded data services to support academic users, including undergraduates, in building statistical capacity in the social sciences. A mutual interest to explore data use (in actual and potential terms) resulted in a project between a data service provider to UK academics, the Economic and Social Data Service (ESDS) International service, and the World Bank as one of the data providers to the service, with the aim of working work collaboratively to explore these issues. ESDS International is especially interested in the application of lessons learned through this work to social science disciplines other than those for which they were developed. The World Bank, having opened up access to its data, is interested in learning how the data could be used in educational contexts.

The Data Deluge and Opening up of Statistical Data

The early 21st Century witnessed a huge increase in the amount of high quality data resources available for public and academic consumption. The 'data deluge' is increasingly being used as a term to describe the issues associated with mass release of data (The Economist 2010). National initiatives in the UK - including the freeing up of government data under open access agreements (<http://data.gov.uk>) - and elsewhere -such as the open data initiatives in New Zealand (<http://cat.open.org.nz>) and the USA (<http://data.gov>) - provide an opportunity for modern citizens to investigate the world. International interest in the open data movement (Guardian 2010) means that it is likely that the data deluge is unlikely to abate. On April 20, 2010 the World Bank opened up its international database (World Bank 2010b) through its

Open Data Initiative, having previously made data accessible through subscription (though 54 widely used economic and social indicators had been freely available online for many years). These open access data resources complement the data services funded to support social and economic research and teaching, including the Economic and Social Data Service (ESDS at <http://www.esds.ac.uk>), which provides UK academic, registered users access to a comprehensive portfolio of data.

As if to signal the opening up of real world data, the first ever World Statistics Day was declared by the United Nations in October 2010 (actually 20/10/2010). On the same day the UK Royal Statistical Society (RSS) launched its ten-year statistical literacy campaign, *getstats* which forecast a ‘society in which our lives and choices are enriched by an understanding of statistics’ (<http://www.getstats.org.uk>). There are three main objectives to the *getstats* campaign:

- To close the gap between the knowledge, skills and perceptions of statistics that people currently have and what they need to make statistics work for them
- To create a new culture towards statistics by altering beliefs around the role of both statistics as data and statistics, the discipline
- To reposition statistics so that it is recognised and desired as a valuable life skill.

getstats 2010

The *getstats* campaign is aimed at elected representatives, the media, education, employers and the wider public as the RSS sought to raise awareness of the benefits of statistics, and create paths to the knowledge and skills needed to make statistics fit for purpose. The *getstats* campaign is a response to the increasing recognition of a need for *statistical literacy*. Schield (2004) regards statistical literacy as a branch of information literacy pointing out that:-

a great deal of information involves statistics. ...It seems difficult to be considered information literate in the 21st century without being statistically literate.

Schild 2004, 6

The Guardian’s DataBlog (<http://www.guardian.co.uk/news/datablog>) and use of infographics for depicting complex data in visual format has exploited the release of open data and attracted a significant following. Two examples of innovative graphics are the

Mortality Statistics in England in 2009 (Guardian 2011) and The Income Gap (Guardian,2009), based on data used in the controversial book *The Spirit Level*. The BBC's *More or Less* radio programme (http://news.bbc.co.uk/1/hi/programmes/more_or_less/default.stm) deals with the popular use and abuse of numbers raising important questions about understanding of statistics, and contributing to the statistical literacy debate.

Following the opening of its databases in 2010, the World Bank announced its *Apps for Development* (World Bank 2010a), an open competition aimed at developing tools for visualizing, analyzing, and sharing World Bank data online, requiring skills from collectively statisticians, subject matter specialists, and software engineers. Over 100 applications were submitted and \$55,000 in prizes were awarded (World Bank 2011). At the same time four conferences on 'Innovative Approaches to Turning Statistics into Knowledge' sponsored by the OECD and other partners (OECD 2010) have shown the institutional desire to encourage analysis of real world data and visualizing statistics. At their best, such innovations have the ability to make statistics readily accessible to ordinary citizens. As Michael Migurski commented in *The Joy of Stats* (2010),

People interacting [with a display of crime data] think it is much more like browsing a website or shopping at Amazon. They are looking at data and don't realize they are doing statistics.

Statistical Literacy

Neither data – nor the need to interpret them wisely - are modern inventions. Nevertheless their ready accessibility via modern technology might give the appearance that statistical data are suddenly everywhere. As Google's Chief Economist, Hal Varian, said in an interview with *The McKinsey Quarterly* in 2009;

The ability to take data - to be able to understand it, to process it, to extract value from it, to visualize it, to communicate it's going to be a hugely important skill in the next decades, not only at the professional level but even at the educational level for elementary school kids, for high school kids, for college kids. Because now we really do have essentially free and ubiquitous data. So the complementary scarce factor is the ability to understand that data and extract value from it.

... statisticians are part of it, but it's just a part. You also want to be able to visualize the data, communicate the data, and utilize it effectively. But I do think those skills - of being able to access,

understand, and communicate the insights you get from data analysis - are going to be extremely important.

Varian 2009

In order for students to be able to acquire skills to achieve insight from data, and think critically about statistics as evidence for inference, they need to develop statistical literacy. Schield discusses the need to be able to understand the social construction of statistics, how statistics are prepared, who counted what and why, as well as the importance of context (for example use of counts or rates), and spurious relationships caused through confounding variables (so correlation does not imply causation). He argues the importance of what he calls 'data literacy' identifying the need to be able to understand and manipulate the data using software packages and tools. Another important dimension of data literacy is familiarity with the standards and systems of classification used to construct datasets. In practice data literacy should be seen as integral to statistical literacy, with students learning about the sources of data and how to manage data sets whilst analysing and querying that data, often through exploration of a substantive issue. Some work has been undertaken investigating the use of secondary data in the social sciences; Stephenson and Caravello (2007) report how they used real world data in support of data literacy in the sociology curriculum at UCLA; Rice and Fairgrieve (2003) investigated use of numeric data in social science teaching in the UK; and Carter *et al.* (2005) summarised a group of projects funded to enhance the use of socioeconomic data in teaching. However, there is much still to be done in the development of this area.

Examples of the use of data in Economics

The data deluge then has meant that there may have never been a better time to engage students with statistical data and a more necessary time to upskill students, future researchers and professionals to handle and analyse such data. Although economics, psychology and educational research have claims to be better served than other social sciences (Gorard *et al.* 2004) quantitative social science remains a 'strategically important and vulnerable subject' area. The following section gives examples of how data use in economics teaching can help students develop statistical literacy through handling real world data. The two examples – both taken from economics – are provided in order to demonstrate where good practice is already established to develop skills in one particular social science discipline; the aim being to learn from this and apply to other discipline areas. The first is a case study undertaken by

ESDS International, the second from a text book on using econometrics and data for developing countries.

Teaching Econometrics. Dr Paul Turner who lectures at Loughborough University uses international time series data in his econometrics teaching to undergraduates. The aim of his Introduction to Econometrics module is to introduce basic and more advanced techniques (regression analysis, hypothesis testing, autocorrelation, heteroskedasticity, dependence between regressors and errors; dummy variables; parameter stability; specification errors; dynamic econometric modelling; and the use of simultaneous equations in econometrics) to second year undergraduates allowing economic theories to be tested using econometric methods and through the use of real world data. He tells his students:

if there's one course they may actually make use of when they leave - it's probably going to be this one.

ESDS 2011

Having been introduced to ESDS International data each student is then assigned a project topic from a combination of three economic variables (annual times series data on consumption, investment, imports extracted from IMF's International Financial Statistics) and seven nations (Australia, Belgium, Denmark, France, Germany, Italy, Japan, UK, USA) such as 'investment in Belgium', 'consumption in Australia' or 'imports in Denmark'. Students submit a short but precise diagnostic report that interprets the economic data worth 20 percent of their final module mark. This approach which has taken three years to refine now works effectively with the large teaching classes.

Student feedback to this approach has been positive. Feedback has been particularly glowing about the modules' applicability to life in the world of world after graduation teaching skills such as decision-making, numeracy, communication and the critical appraisal of important econometric methods. Turner has shared teaching resources through professional teaching networks and in spaces such as Youtube and ESDS International's Teaching Tools (for more details see <http://www.esds.ac.uk/international/elearning/teaching-tools/tools/econometrics/print.asp>).

Modelling socioeconomic outcomes. Mukherjee, White, and Wuyts (1998) provide an extended example of how simple graphical methods can be used to interrogate a dataset

before applying more complex, econometric methods. Their example involves cross-country data on Gross National Income (GNI) per capita, the crude birth rate, and infant mortality. Beginning with simple histograms, they encourage students to notice the dominant feature of each data set; next they consider bivariate scatter plots, and then scatter plots of algebraic transformations of the variables, eventually leading to a model of the fertility rate that is linear in the log of GNI per capita and the square root of infant mortality. They caution students,

Results which look good at first sight may be riddled with problems if we care to look at our data more carefully

Mukherjee *et al.* 1998, 16.

These two examples illustrate how working with real world data can assist students in grappling with real world problems, and demonstrate that data are messy and need to be investigated in order to be understood. The findings section of this paper will highlight an example of statistical literacy that was presented through the project undertaken, using the measure of GDP (Global Domestic Product) as an example.

Which data?

The availability of ever increasing amounts of accessible data comes with a price. Understanding and interpreting data however requires a set of critical skills that are in need of development. The term ‘data’ is used throughout here to mean ‘international macrodata’ or more specifically socioeconomic, secondary quantitative data (see Cole *et al.* 2008, 366) - in other words quantitative data that can be used for secondary analysis.

Macrodata are aggregated to a country or regional level, as opposed to microdata which are usually survey or questionnaire datasets collected from individuals, households, or firms. Macrodata can be presented in the form of time series at monthly, or longer, frequency. They are typically collected by a national statistical offices, administrative agencies or sectoral ministries. Microdata typically record characteristics of the reporting unit, permitting analysis at the individual (sometimes household) level.

An example of international macrodata is the International Energy Agency’s World Energy Statistics which has been used to analyse, among other things, the impact of cross-national co-operation on the price of electricity provision (Boonyasana 2010). An example of international microdata can be found in attitudinal surveys. The New Russia-Barometer, New

Baltic-Barometer and New Euro-Barometer survey series have facilitated cross national analysis of attitudes in Eastern Europe and Russia over a period of time of considerable political and economic change (ESDS 2010).

For data to be useful for cross-country comparisons they must be compiled using standard methodologies or standardized by intergovernmental organisations such as the Organisation for Economic Co-operation and Development (OECD), International Monetary Fund (IMF) or United Nations and its specialized agencies such as the World Health Organisation, International Labour Organization, and UNESCO.

Further information about use of macrodata and microdata made available through ESDS can be found online (www.esds.ac.uk). Case studies based on use of the data are reported in the ESDS Annual Report (for example, see ESDS 2010) each year and regarded as an opportunity to capture impact of the downstream use of the data and at the ESDS International website (<http://www.esds.ac.uk/international/casestudies/>).

ESDS International

As a specialist data service, ESDS International provides all UK academics and students free online access to a large, high quality socio-economic data portfolio. Details about ESDS International and how it has developed can be found in Russell *et al.* (2005).

The service hosts key international time series statistical databanks produced by international governmental and statistical organizations including the World Bank, the International Monetary Fund (IMF), the International Energy Agency (IEA), the United Nations (UN) and the Organisation for Economic Co-operation and Development (OECD). Data cover topics from finance, trade and industry to education, employment, health, migration, social expenditure, human development, energy and the environment. By far the most popular (accessed and downloaded) are The World Bank's datasets (which include the World Development Indicators (WDI), the Global Development Finance (GDF) and the Africa Development Indicators (ADI)), followed by the OECD datasets and the IMF's International Financial Statistics (IFS) (Carter and Shaw 2011). Since its launch in 2003, the service has recorded over 582,600 data access sessions by 30,840 unique users, from 217 different institutions (to Nov 2010). At registration users are asked to select the discipline area to which they belong; the resulting usage figures reveal that for ESDS International use is

dominated by economics and econometrics (of the 30,840 users 17,703 were from Economics and Econometrics, 3,752 Business and Management Studies, 1,909 Politics and International Studies and 1,783 Accounting and Finance). Whilst this is unsurprising given firstly the type of data, and secondly that quantitative courses tend to be taught more heavily in the economics and related disciplines, the service is actively raising awareness of the potential for using international data in other disciplines.

ESDS International also helps users to locate and acquire international micro level datasets, such as Euro-Barometer. The service supports the use of these macro and micro databanks through the provision of a comprehensive website, a helpdesk, a suite of supporting documentation and a number of teaching and learning resources including Case Studies of International Data Use (<http://www.esds.ac.uk/international/casestudies/>), Countries and Citizens: Linking international macro and micro data (<http://www.esds.ac.uk/international/elearning/limmd/>) and the Teaching Tools area of the website launched in October 2010 aimed at lecturers interested in using real world international data in their teaching (<http://www.esds.ac.uk/international/elearning/teaching-tools>).

The World Bank

The World Bank is a vital source of financial and technical assistance to developing countries around the world. Its mission is to fight via local empowerment and the provision of resources, sharing knowledge, building capacity and forging partnerships in public and private sectors. The World Bank's Development Data Group provides a collection of national and international statistics to clients within and outside the Bank, and works towards improving the capacity of member countries to produce and use statistical information. The Data Group works with other organizations on new statistical methods, data collection activities, and statistical capacity-building programmes. This group also coordinates the World Bank's analytical and statistical work related to monitoring progress towards the Millennium Development Goals (MDGs) (<http://www.un.org/millenniumgoals>). The World Bank and ESDS International have collaborated since 2003 to provide World Bank data to all UK academic institutions through ESDS International's online delivery platform.

Improving the quality, timeliness, and coverage of international statistics has received greater attention since the announcement of the MDGs in September 2000. Around the same time a

consortium of international and national statistical agencies was formed under the title of Partnership in Statistics for Development in the 21st Century, or Paris21. The purpose of Paris21 is

to promote, influence and facilitate statistical capacity development and the better use of statistics.

www.paris21.org

Paris21 maintains an annual Partner Report on Support to Statistics (www.paris21.org/Press). Over the past decade the World Bank has provided over \$450 million in loans and grants to developing member countries to strengthen national statistical systems. Bilateral aid donors are also important supporters of statistical capacity building programs in developing countries.

One of the best used databases made available by the World Bank and through ESDS International is the World Development Indicators (WDI) database, a compilation of 1239 aggregate indicators for 213 countries and territories since 1960 (see <http://data.worldbank.org/data-catalog/world-development-indicators>). The unique feature of the WDI database is that it brings together a wide variety of statistical indicators relevant for monitoring social and economic characteristics of developing and developed countries.

A ten-year retrospective by Paris21 (PARIS21 2010) found evidence of increased availability of social indicators used to monitor the MDGs. An analysis of economic statistics from the WDI database also found significant improvements in data coverage.

<<Table 1: About Here >>

The World Bank maintains a Bulletin Board on Statistical Capacity (BBSC at bbsc.worldbank.org) which scores developing countries' statistical systems based on methodology, data sources, and periodicity and timeliness. The BBSC scores document

overall improvements in the quality of statistical systems, as do the IMF's Reports on the Observance of Standards and Codes (IMF 2010). Nevertheless changes in methods and practices can lead to large revisions to data series and breaks in comparability over time. Ghana, for example, has recently revised its national accounts, leading to an upward revision in GDP of more than 50 percent, which will cause it to be re-classified by the World Bank from a low-income to a lower-middle-income economy.

The Project

The project was made possible through funding under the University of Manchester's Simon Industrial Fellowship Fund for the promotion of research and teaching in the Social Sciences. Dr Swanson of The World Bank Group took up the invitation to visit the UK to work collaboratively with the ESDS International team on this project. This work builds on a series of presentations and papers given throughout 2010 on the use of real world data in teaching (Carter 2010a; 2010b; 2010c; 2010d).

The project ran in the last quarter of 2010. The primary aim was to work with the World Bank to strengthen links between the service, the data providers and the data users. The move by the World Bank to provide free and unrestricted access to its databases changed ESDS International from the principal gateway in the United Kingdom for the World Development Indicators to being one among many potential sources, including the World Bank's own website. It was therefore timely to focus on academic users of World Bank data in order to better understand the benefits afforded through ESDS-International, and any barriers to use.

The results reported here focus on engagement with data users based on a series of visits and talks to the Universities of Manchester and Oxford and the London School of Economics and Political Science, and interviews with data users, with a focus on the World Bank's data. It was also possible to meet with social science researchers at the ESRC's Social Science Festival, and with the ESRC's strategic advisor for quantitative methods at undergraduate level. The final presentation as part of the fellowship was at the ESDS International's annual conference through participation in a panel session on Helping to Build Statistical Capacities.

The objectives of the joint activity were as follows:

- To better understand what the academic community do with the Bank's data, why they do/or don't use it and what would help them to use it more.
- To find out if users of one dataset tend to use other datasets too and if so what obstacles (if any) they face when accessing data from multiple domains.
- To contribute to improving statistical literacies and capacity building; trying to learn from good practice where this exists
- To take good practice from one discipline and extend into other social science disciplines if possible; to avoid 're-inventing the wheel'.
- Identify users who could help to build narratives around dataset use through either case studies or contribution to the Teaching Tools area of the website through future activity.
- Communicate the benefits of using international data in teaching and research across the social science community.

Methodology

A lack of information gathered from data users means the ESDS International service has to employ other methods to investigate how and why its data is utilised. Some of this is gathered as feedback from the many talks, presentations and training sessions given by the service. For example from August 2010 to January 2011 over 250 people attended ESDS International training courses, and provided feedback, across four universities (full list of events at <http://esds.ac.uk/international/news/intpastevents.asp>).

However, data are notoriously hard to track and until data citation is standard practice the need to be innovative in tracking users and use will remain. So for instance, researchers are encouraged to submit case studies of their use of international data for publication on the ESDS International website (with a small incentive offered) , searches are carried out on Google Scholar for any publications referencing data obtained from ESDS International (knowing that many do not reference data and so this is a poor substitute), and delegates of ESDS International's annual conference publish their papers and posters on the conference proceedings section of the ESDS International website.

In addition to the series of talks and visits to Universities discussed earlier, this project sought input from experienced users of World Bank data, and therefore a small non-random sample was selected from among the 40 most active ESDS International World Bank data users; additional requests for feedback were posted to relevant mailing lists, for example Quantitative Methods teaching list; a focus group at the London School of Economics was interviewed; and finally a number of targeted emails were sent to key contacts. This resulted in follow-up interviews with 12 key users (researchers and staff from six universities - Edinburgh, Hull, Leicester, LSE, Manchester and Middlesex) who were asked a number of questions about their data use, to investigate:

- principal research interests
- data sources used
- experience of using these databases (such as feedback on ease of use of interfaces, clarity of data content and definitions).
- tools used for data management and analysis
- use of international data in teaching

Results and Findings

The results reported focus on an example of statistical literacy through choice of data as presented in one of the talks, and findings from the interviews conducted with respondents. Supplementary information was gathered from discussions with staff at other universities, for example University of Oxford through courses given on MA in Development Economics, and panel members at the ESDS International conference.

An Example of Statistical Literacy using World Bank data - GDP

The example presented here of the need to alert students to the choice of indicators for Gross Domestic Product (GDP) formed part of the presentation given in the talk at University of Manchester (Swanson 2010).

The assertion of Schield (2004) that statistical literacy must include knowledge of data sources and their construction is borne out by this example of the delivery of knowledge and

skills in econometrics. By considering two measures of Gross Domestic Product (GDP) we can see how the perception of the world – and thus policy preferences - is altered by the choice of data. It is vital for social science to understand the context of data and for educationalists to equip students with the necessary knowledge and skills to undertake meaningful study and research.

Data from the World Bank's *World Development Indicators 2010*, showing the share of global output measured firstly by gross domestic product (GDP) converted to U.S. dollars using market exchange rates and secondly GDP converted at purchasing power parity (PPP) are presented below (Figure 1)

<<Figure 1: About Here>>

The two resulting graphs in Figure 1 present very different pictures of global output. Using 'GDP by exchange rate' data high income economies accounted for 76 percent of global output, but by contrast the 'GDP by PPP' data reduces their share to 57 percent, whereas that of the lower middle income group doubles from 12 to 24 percent. The reason for the shift is that PPP is the exchange rate at which the money you exchange would buy exactly the same basket of goods in both countries and as such, PPPs are necessary for making meaningful international comparisons of welfare or standards of living. However, they are not appropriate for evaluating financial resources, which must be obtained through market exchange rates. Hence it is important for students to be aware of the existence of both indicators, the conceptual differences between them and when it is appropriate to use each statistic according to the analysis being undertaken (see Swanson 2010 for further details of this example).

The Respondents

From the formal interviews respondents were chosen from a variety of backgrounds and research and learning needs. Seven of the 12 were PhD or Masters level students, one was a

Research Librarian, one was a senior lecturer, one a post-doc, and two professors in their HEI. It was important to interview students as well as researchers and teachers since we were interested in the ‘bottom-up’ perception of the use of real world data in teaching and research as well as the ‘top-down’ model that was provided by the teachers themselves. Discipline areas covered were Economics and Development Economics (7), Health and Population studies (2), Statistics (1), Finance (1) and Library and Management (1). Those researching and/or teaching were interested in the following fields of research: trade, poverty, energy poverty, financial development, concept of household in developing country surveys, the impact of family planning and reproductive health on social outcomes, evasion and corruption, financial regulation and credit availability and labour markets in developing countries. This reflects a highly diverse range of topics covered by a relatively small group using international data as part of their research or teaching, representing a huge spread of substantive research topics.

Data used

The group tended to use a variety of different international data sources as well as World Bank data, all of which are available through ESDS International, for example, UNIDO’s industrial statistics, the IMF’s International Financial Statistics and Government Financial Statistic datasets, International Energy Agency (IEA) and UN Common Database. Some of the respondents reported that they used microdata as well as macrodata, for example the Demographic and Health Surveys (DHS) from Macro International, and Living Standards Measurement Study surveys available through the World Bank; these data are not available through ESDS International. Most of the students encountered World Development Indicators (WDI) data in masters-level programs where they had neither the time nor the training to undertake primary data collection. The WDI was seen as particularly useful because it contains long time series for a wide variety of internationally comparable indicators. This allows students to apply newly acquired statistical methods to a standard data set, with a particular emphasis on cross-country and panel data analysis. One institution (the LSE) plans to roll out the use of WDI data in the immediate future with a new quantitative methods first year course for undergraduates that will present case studies of current economic, political, and social issues. Students will be asked to use data (from a variety of sources) to explore these issues (naturally with less emphasis on rigorous statistical analysis

but in order to explore substantive issues using data, with a view to increasing statistical literacy).

A number of interviewees from teaching and research backgrounds said that they would find it useful to have access to the World Bank's Worldwide Governance Indicators (WGI) data through ESDS International, and some requested other data or tools for accessing data (such as the World Integrated Trade Solution for the UN COMTRADE data). Several of the interviewees commented on how useful it was having all the data available in a single place and said that they recommend it to their colleagues and students. There was also evidence of research students using real world data on the recommendation of their lecturers and its previous use in taught courses while a number of respondents claimed that although they had found data available on the open web they would continue to access it through ESDS International as the principle access point to international data.

However there were some issues of usability. Although some respondents commented favourably on the ease of use of the interface, and despite the fact that lost data can be easily restored and downloaded again, some complained that the data delivery software used (Beyond 2020) is now looking 'a little outdated' and it is clear from the comments that some users find the process of downloading multi-dimensional data in a format suitable for panel analysis troublesome. This has helpfully highlighted a user support area which needs to be addressed as although it is possible to manipulate the format of the data tables prior to download some users clearly do not find the software very intuitive.

Tools and methods used

Most of those interviewed download data and export it into a separate data management and analysis tool. Tools mentioned include Excel, e-Views, SPSS, Stata, and Matlab with many first processing the data in Excel before going on to analyze it in other software. The data analysis tools used varied according to the level of the respondent, with software such as e-Views and SPSS being used for teaching and STATA for research. WDI data was highlighted for being particularly useful for descriptive statistics.

Of the 12 formal interviewees, five used international data in their teaching as well as their research, six used it for their own research only and one ran training courses on library resources and how to find data. Real world international data was used to develop statistical techniques and methodologies such as ordinary least squares and multiple regression, simple plotting of data, issues of non-stationarity in time series and other more complex analyses with time series and cross sectional data. It was noted that although a number of the interviewees teach courses at 1st year undergraduate level (such as economics, introduction to macroeconomics and maths for economics) these courses did not typically include the use of real world data. Of course this might change given the pressure to close the statistical literacy gap and it is important that data providers meet the challenge of having data that can be use at all levels from first year undergraduate to post-doctoral research work.

Academics and postgraduates interviewed reported using the data in the following ways:

Undergraduate teaching

- Macroeconomic indicators are used to teach multiple regression and more complex analysis with time series and cross sectional data to level two undergraduates.
- Data from WDI and IMF's IFS are reported to have 'significant use' in undergraduate dissertations.
- WDI data is used as teaching aid to identify substantive issues in undergraduate teaching. There is a widespread willingness to use data for policy formation and advocacy and one institution is committed to the planned introduction of a compulsory 1st year undergraduate unit designed to refine research skills so that students 'think like a social scientist'. The course teaching here will be concentrated on quantitative and qualitative data using pre-prepared datasets and no analysis tools.

Postgraduate teaching

- Taught postgraduate programmes use statistical packages (e.g. STATA) to investigate real world data such as the WDI.
- Dissertations and independent research projects at Masters level often use WDI as a ready source of descriptive statistics, and using cross-sectional or panel data, especially in Development Economics.
- PhD research training (for example on financial development and growth and in Africa) utilises WDI and ADI data. Students on these courses download data to Excel and subsequently use e-Views for analysis.

- Masters' level and Doctoral candidates make extensive use of WDI and IMF data as these students typically require large amounts of data to run regressions.
- There was extensive reported use of OECD datasets at PhD level.

Other:

- A post-doctoral fellow maintains a useful "blog" on development data sets (<http://sites.google.com/site/medevecon/development-economics/devecondata>)

Conclusions

This work represents only the initial phase of a much deeper analysis of the use of socioeconomic secondary data sources in research and teaching. As such it has merely scratched the surface although some worthwhile findings have been identified nevertheless.

The demand for downloading data for analysis using others tools is interesting, as this helps data providers to determine the suitability of the data access platform. A lack of demand for data analysis functionality within the data delivery software would influence future development; so if users simply want to access data and process/analyse it elsewhere this prevents data providers providing more sophisticated tools. On the other hand it could be that the current platform does not support data analysis sufficiently for users, and requires development in order to facilitate wider usage. Further work would need to be carried out with users to ascertain this.

There is evidence of the data being used in teaching *programmes* to upskill students in statistical literacy. Two encouraging examples are the introduction of a compulsory first year course for social scientists, and the embedded provision of data access through a University Library. Other examples of good practice tend to be focused in econometrics, supporting the perception that this subject is well served in this area, although we note that our sample was taken from the top users of ESDS International and is therefore biased towards this group. It is proposed that econometrics could serve as a beacon of good practice for the further roll-out of the use of real world data in research and teaching in order to address the statistical literacy gap in the social sciences and beyond.

Teachers and students use a variety of techniques and tools for statistical data analysis. Sharing this information across the sector would be beneficial. The Teaching Tools area of the ESDS International website stands to benefit from collation of this information. Indeed it is tempting to call for further consolidation of teaching materials in teaching and learning repositories in order to share good practice via communities of practice throughout the sector.

There is support for the notion that a single-site aggregating data sources adds value to its development in research and teaching in social science. Clearly ESDS International has a reputation as a good place for discovery of data despite the fact that more open data is appearing on the web, including from data providers such as the World Bank. Some interviewees commented that even though data were appearing elsewhere the ‘one-stop-shop’ approach provided through ESDS International would continue to provide value. On a similar note it an important aspect of promoting the use of data is advocacy and support; the interviews showed that where students have been supported in their use locally through lecturers and library staff this helped give them confidence to use data in their own work. This highlights a further important aspect for statistical literacy. Additionally the development of pre-prepared datasets might assist in teaching requirements. This could be an activity that services such as ESDS International could undertake. Several interviewees talked about having enough data to undertake statistical analyses such as multiple regression – and this knowledge could be shared with others alongside the datasets (or variable/country selections) to help teachers get started in using data themselves.

Future research

This project has given us an opportunity to start to gather evidence of use. Further work that is planned is to extend this work by following up these initial contacts to gather teaching resources and narratives (including video recordings) about data use. We would also plan to focus on research methods, so work with data users to gather more information on which methods are used with international macrodata, and whether there are limitations with these methods given the available data. It is important though to not focus exclusively on the methods, but to understand the substantive issues that researchers are investigating – this approach has proved successful in some of the activity that has been undertaken through the

ESRC Quantitative Methods Initiative (ESRC 11), especially in disciplines that are less quantitatively driven such as sociology.

One of the authors has been granted an Open University Fellowship under the SCORE (Support Centre for Open Resources in Education) programme (www8.open.ac.uk/score/fellows). This will enable the continuation of the work reported here, through engagement with lecturers and researchers in the sector in support of sharing teaching resources to address the ESRC's recommendation for 'more accessible and better teaching resources available on the web' (MacInnes 2009) in support of quantitative methods teaching.

Summary

The work reported in this paper through the project undertaken with the World Bank and ESDS International has provided useful information about secondary data use in practice in social sciences in UK universities. The interviews focused on data use in econometrics and economics, reflecting the majority of researchers and teachers are, but this is to be welcomed as it helps us better understand where good practice exists. One of the primary aims of the project was to identify this in order that firstly it could be better understood and secondly it could help inform future work in extending into other social science discipline areas. The Teaching Tools part of the ESDS International website has already started to build narratives of data use from real teachers. We can now use this evidence of use from other teachers identified through this project with the World Bank to populate the Teaching Tools resource with course plans and examples from economics and importantly other subject areas in support of building statistical capacity across the sector.

However this activity will not be without its challenges; to engage researchers from disciplines that do not traditionally deal with quantitative methods will be hard. Having a body of evidence – no matter how small – that shows this is achievable in other area will however be advantageous. Showing that using statistical data can help align universities with national strategies (such as those of the ESRC and RSS for instance) should help. Having

demonstrable evidence that shows using data also helps with the higher skills agenda (see Carter 2010a) should also help encourage students to get on board. Having a bank of initial teaching resources, including teaching datasets, and methods used in different disciplines and at all levels to show how others are grappling with the data issue is a start; building on this so that it is evident across the social sciences, in related subjects such as environmental sciences, and in educational research is a longer term goal.

ESDS International will continue to align its work with the statistical literacy community, and welcome continued interaction with the World Bank and other organisations that are developing statistical capacities globally. The need for data services to demonstrate impact on the communities they serve is paramount; ESDS International is regarded as a jewel in the crown for UK social sciences and needs to be part of the endeavour to raise statistical literacies too.

References

Bolton, P. 2010a. Statistical Literacy, *Social Indicators*, Research Paper 10/65, pp. 9-13, House of Commons Library, <http://www.parliament.uk/briefingpapers/commons/lib/research/rp2010/RP10-065.pdf>.

Bolton, P. 2010b. How to spot spin and inappropriate use of statistics, *Statistical Literacy Guide*, House of Commons Library, <http://www.parliament.uk/documents/commons/lib/research/briefings/snsg-04446.pdf>.

Boonyasana, K, 2010. World Electricity Co-operation. ESDS International Case Study at <http://www.esds.ac.uk/international/casestudies/electricity-coop.asp>

Carter, J., Rice, R., and Cole, K. 2005. JISC's Learning and Teaching (5/99) Programme: the Data Cluster, *VINE*, Vol. 35, Issue: 1/2, pp.24 - 27.

Carter, J., and Shaw, L. 2011. Pushing the boundaries: building on a UK success story of taking sharing beyond national boundaries, paper presented at INTED 2011 Spain, http://www.iated.org/concrete2/list_accepted_abstracts.php?event_id=9.

Carter, J. 2010a. EDULEARN “Real Data: Real World; Real Stories: A Case Study Approach to Demonstrating Impact on the Student Experience” EDULEARN10, Barcelona, 5-6 July 2010. Paper in EDULEARN10 Conference Proceedings ISBN: 978-84-614-2439-9 (with abbreviated online Case Study at <http://www.esds.ac.uk/international/casestudies/real-data/real-data.asp?print=1>)

Carter, J. 2001b “Real Data: Real World: Real Stories” at Coalition for Networked Information Spring Meeting, Baltimore, 2010 <http://www.cni.org/tfms/2010a.spring/Abstracts/PB-real-carter.html>

Carter, J. 2010c “Telling stories at ELI2010” <http://newsletter.alt.ac.uk/1jlqks6612e>

Carter, J. 2010d “Telling stories about and with data” IASSIST 2010, Cornell University,

CFSS, 2011 Campaign for Social Science <http://www.campaignforsocialscience.org.uk/>

Cole, K., Wathan, J. and Corti, L., 2008 The Provision of Access to Quantitative Data for Secondary Analysis in Fielding, N., Lee, R.M. and Blank G. (eds) *The SAGE Handbook of Online Research Methods* Sage Publications, London.

The Economist, 2010. The Data Deluge <http://www.economist.com/node/15579717>

ESDS, 2010. Understanding Global Change: new data on attitudes across Russia and Eastern Europe. <http://www.esds.ac.uk/news/publications/ESDSAnnualReport2009-2010.pdf>

ESDS, 2011. ESDS International Teaching Tools: Economics at Loughborough University <http://www.esds.ac.uk/international/elearning/teaching-tools/tools/econometrics/print.asp>

ESRC, 2011. Economic and Social Research Council: Quantitative Methods Initiative, <http://www.esrc.ac.uk/funding-and-guidance/tools-and-resources/research-resources/initiatives/qmi.aspx>

Fielding, J., and Gilbert, N. 2006. *Understanding Social Statistics*, 2nd Edn, SAGE Publications, London.

getstats 2010, Royal Statistical Society *getstats* Campaign <http://www.getstats.org.uk>

Gorard, S., Rushforth, K., and Taylor, C. 2004. Is there a shortage of quantitative work in education research? *Oxford Review of Education*, Vol. 30, No. 3.

Guardian, 2009 Inequality: mother of all evils?: <http://image.guardian.co.uk/sys-files/Guardian/documents/2009/03/13/inequality.pdf>

Guardian, 2010, DataBlog: World Government Data <http://www.guardian.co.uk/world-government-data>

Guardian, 2011, DataBlog: Mortality Statistics England & Wales http://www.guardian.co.uk/news/datablog/2011/jan/14/mortality-statistics-causes-death-england-wales-2009?CMP=twf_fd#zoomed-picture

HEFCE, 2008 <http://www.hefce.ac.uk/aboutus/sis/socialsci/>

IMF ,2010. Reports on the Observance of Standards and Codes (ROFCs) <http://www.imf.org/external/np/rosc/rosc.asp>

MacInnes, J. 2009. Final Report: Strategic Adviser for Quantitative Methods: Proposals to support and improve the teaching of quantitative research methods at undergraduate level in the UK. Economic and Social Research Council, http://www.esrc.ac.uk/images/Undergraduate_quantitative_research_methods_tcm8-2722.pdf.

Mukherjee, C., White, H. and Wuyts, M., 1998. *Econometrics and Data Analysis for Developing Countries* Routledge, London.

OECD, 2010. Summary Report on the Seminar on Innovative Approaches to Turning Statistics into Knowledge: <http://www.oecd.org/dataoecd/22/21/46754884.pdf>

PARIS21 2010. Paris21 at Ten: Improvements in Statistical Capacity since 1999, *PARIS21 Consortium Report*, October 2009, <http://www.paris21.org/sites/default/files/P21-at-10.pdf>

Rice, R., and Fairgrieve, J. 2003. An enquiry into the use of numeric data in learning and teaching in UKHE. *Radical Statistics*, Issue 81, pp 18-38.

Rosling, H. 2010. *The Joy of Stats* - Hans Rosling's 200 Countries, 200 Years, 4 Minutes - BBC Four: <http://www.youtube.com/watch?v=jbkSRLYSojo>

Russell, C., Cole, K., Noble, S., Syrotiuk, N. 2005. Delivering the World: The Establishment of an International Data Service, Joint UNECE/OECD Work Session on Statistical Dissemination and Communication, United Kingdom, United Nations Statistical Commission and Economic Commission for Europe, <http://www.unece.org/stats/documents/2005/02/dissemination/wp.8.e.pdf>.

Schild, M. (2004) *Information Literacy, Statistical Literacy and Data Literacy*. IASSIST Quarterly, Summer/Fall (online at http://www.iassistdata.org/downloads/iqvol282_3shields.pdf)

Smith, E. 2008. *Using Secondary Data in Educational and Social Research*. Open University Press, McGraw-Hill Education ((with accompanying website <http://www.secondarydataanalysis.com/>)) .

Stephenson, E. and Caravello, P. S, 2007. *Incorporating data literacy into undergraduate information literacy programs in the social sciences: A pilot project*, Reference Services Review, Vol. 35 Iss: 4, pp.525 – 540

Swanson, E. 2010. *Working with International Development Data*, ESDS International, <http://www.esds.ac.uk/international/news/news.asp#16dec10a>.

The Joy of Stats, 2010. BBC Four. Broadcast 7th December 2010.

Varian, H. 2009. Hal Varian on How the Web Challenges Managers , *McKinsey Quarterly* 2009, http://www.mckinseyquarterly.com/Hal_Varian_on_how_the_Web_challenges_managers_2286.

United Nations, 2010. World Statistics Day 20.10.2010 <http://unstats.un.org/unsd/wsd/Default.aspx>

World Bank, 2010a World Bank Apps for Development: <http://appsfordevelopment.challengepost.com/challenges/78>

World Bank, 2010b. World Bank Frees up Development Data <http://go.worldbank.org/MZ29WQLWK0>

World Bank, 2011. World Bank Apps for Development Submissions: <http://appsfordevelopment.challengepost.com/submissions>

World Bank. World Bank. Annual. *World Development Indicators*. Washington, DC.

Websites (accessed between Dec 2010 and May 2011)

A Gateway to the UN System's work on the MDGs: <http://www.un.org/millenniumgoals/>

BBC More or Less: http://news.bbc.co.uk/1/hi/programmes/more_or_less/default.stm

Economic and Social Data Service (ESDS): <http://www.esds.ac.uk>

ESDS International Case studies: <http://www.esds.ac.uk/international/casestudies/>

ESDS International Countries and Citizens: Linking International Macro and Micro Data: <http://www.esds.ac.uk/international/elearning/limmd/>

ESDS International Teaching Tools: <http://www.esds.ac.uk/International/elearning/teaching-tools/>

Mimas: Building Statistical Capacities: <http://mimas.ac.uk/news/2010/11/building-statistical-capacities/>

PARIS21 Partnership Report on Support to Statistics: <http://www.paris21.org/Press>

World Bank and United Nations Virtual Statistical System: <https://www.virtualstatisticalsystem.org/>

World Bank Bulletin Board on Statistical Capacity: <http://bbsc.worldbank.org>

Number of MDG Indicators with at least two data points	2003		2006		2009	
	Number of Countries	Percentage of Countries	Number of Countries	Percentage of Countries	Number of Countries	Percentage of Countries
0-5	31	19	9	6	9	6
6-10	49	30	19	12	8	5
11-15	79	48	31	19	28	17
16-22	4	2	104	64	118	72
Total	163	100	163	100	163	100
Source: Paris21 2010						

Table 1: Improved coverage of MDG monitoring indicators, 2003 to 2009

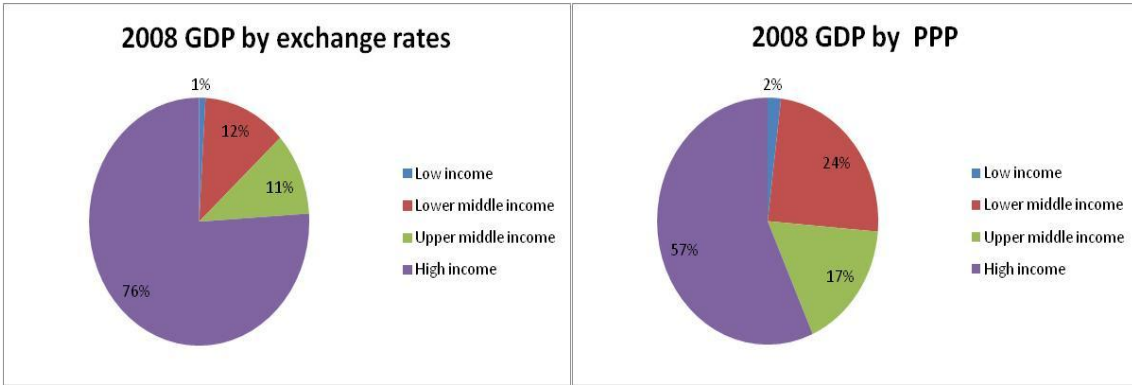


Figure 1: 2008 GDP by (a) exchange rates and (b) Purchasing Power Parity (PPP)