Mind the Gap: Global Data Sharing

by Susan Noble, Celia Russell and Richard Wiseman[1]

ESDS

Abstract

In an increasingly globalised world, the importance of collaborative research into international issues, such as climate change, economic crises, migration and health cannot be underestimated. This type of research requires data sharing at a cross-national level. In this paper we discuss the barriers which prevent data sharing on a global scale and the need to address the lack of data awareness. We will explain the licensing issues which restrict the use of data in the ESDS International portfolio by the UK academic community but will also highlight ways in which the ESDS International service helps the non-UK international data research community by identifying free data sources and producing freely available supporting documentation and other materials.

The paper will highlight the development of e-learning materials based upon a key socio-economic theme: 'the UN's Millennium Development Goals (MDGs)'. As the current international development planning framework, the eight MDGs are time-bound and quantifiable targets for reducing poverty, improving health and protecting the environment. Intended to raise awareness of the potential use of quantitative international data in research and teaching, additionally, these open access e-learning materials have been developed as an interactive online set of materials to be used as a self-teaching resource.

Keywords: international data, social sciences, macro-economic databanks, ESDS International, Millennium Development Goals, e-learning.

INTRODUCTION

The Organisation for Economic Cooperation and Development (OECD) describe climate change as humankind's central long-term challenge[2]. Addressing it successfully requires the combined efforts of researchers and national governments worldwide. Similarly, the international banking crisis cannot be tackled from a single country perspective alone. These types of transnational problems call for multilateral, multilevel responses combining international and national policy-making supported by evidence-based research. So how can data specialists best support research communities through international data sharing? And how do we aid the development of the statistical capacities required to inform policy-making on a global scale?

The international data which feeds into this type of social science research is produced by intergovernmental organisations (IGOs) such as the International Monetary Fund, the International Energy Agency, OECD, the United Nations and the World Bank. These organizations have a presence in every country in the world, the authority to create international standards and the technical and financial capacity to support the development of national statistical infrastructures. They have long produced high quality, regularly updated time series databanks for their own internal use which typically contain a huge range of macro-economic and social indicators aggregated to

Since the creation of ESDS International no longer is access a barrier ...

national or regional level and collectively cover virtually every country in the world. The academic research community needs access to these unique datasets in order to contribute to and comment on policy responses to global issues. Furthermore, access to these data resources give students the opportunity to work with real world data.

ESDS International

In the United Kingdom, ESDS International, a specialist

data service of the wider Economic and Social Data Service (ESDS) was established in January 2003 to address the issue of international data access for the UK academic community. Hosted by Mimas at the University of Manchester, the service helps researchers to locate and acquire international survey datasets such as the Eurobarometer and World Values Survey and provides free access to over 35 socio-economic macro datasets from nine different data providers via a common web-based user interface.

Beyond 20/20 Web Data Server, a web-based data dissemination tool provides the common user interface to the ESDS International macro international data. It requires only a standard web browser, is accessibility compliant and can be used to display, subset, visualise and download time series data. Each of the IGO's provides large and complex data files in various formats and the ESDS International developers convert these data files into a format suitable for delivery via Beyond 20/20.

Each of the datasets is accompanied by a differing quantity and quality of explanatory documentation which ESES International collate and re-format into standard supporting documentation, i.e. dataset user guides which are consistent across each of the datasets available in the portfolio irrespective of the IGO data provider. This set of comprehensive supporting documentation includes details of the topics and time range covered, the countries included, the periodicity of the data and links to relevant documentation.

Free access to the service for the UK academic community has been made possible through a series of ground-breaking national data redistribution licensing agreements between each of the data providers and the University of Manchester (which hosts the service).

For the bulk of the licence agreements ESDS International employed the services of Databeuro[3] – a professional data negotiation agent which already had expertise working with IGOs for access to online data and e-books to the academic, government and corporate sectors. Where possible the data re-distribution agreement negotiated with each IGO was based on a model licence. The model licence adopted was very similar to one developed between the Joint Information System Committee and the Publishers Association for the licensing of commercial datasets[4] . It sets out generic terms and conditions of use (e.g. educational use only) but there was also scope for adding additional clauses to reflect special conditions required by different suppliers. As the licence agreement was between the University of Manchester and the IGO, it was necessary to ensure that the University was not exposed to any financial risk by taking on the role of licensing the data on behalf of the entire UK academic community. For this reason, appropriate clauses relating to limitation of liability and dispute resolution were inserted to protect the University of Manchester.

These UK wide redistribution agreements deliver significant savings to the academic community as institutional or individual subscriptions are no longer required and in some instances additional discount was negotiated by agreeing to a five year deail with a single up-front payment in year one. In addition they remove one of the major barriers to use of these datasets in research and teaching by allowing all members of the UK academic community, irrespective of their institution, to access the data through ESDS International, free at the point of use.

The unique licensing arrangements have certainly addressed the issue of data access for the UK academic community, as the service has witnessed an astonishing growth in the UK international research community from a few hundred users at a handful of institutions in 2003 to over 30,000 users representing over 200 different institutions across the UK as of November 2010. However, there is still an issue concerning data access elsewhere in the world, and there the barriers which prevent global data sharing still exist. ESDS International is frequently contacted by non-academic organisations and non-UK academics to request data access or provide support and further information (62 queries logged to our helpdesk between January 2009 and November 2010). In general response to this apparent gap in data provision ESDS International now:

•	Provides helpdesk support where possible – e.g. will point user to an alternative source of free data if available – for example, will advise a user from a developing world who to contact to obtain the data.
•	Produces comprehensive user guides for each dataset which include details of dataset topics, as well as its geographical and temporal coverage and makes them available to anyone from the ESDS International website.
•	Regularly updates a guide to freely available international data and resources.

Since the creation of ESDS International, no longer is access a barrier for instructional data use in teaching, but rather there is a lack of awareness of data resources and their potential for use in a teaching environment. User consultation on this issue identified the following key points regarding the relatively low use of international data in teaching:

A lack of awareness of the potential use of ESDS data for teaching was highlighted. It was felt that the use of international data at an institution was very dependent on pro-active data librarians/ tutors seeking out information about what data is available and promoting it within their institution or department. Librarians appear to play a key role in assisting academic users to identify and access data for research and teaching purposes and should be targeted with information about international datasets.

 A lack of teaching materials based on international data – so even when the service is well promoted it is primarily used for post-graduate research and tutors are unsure of how to use the data in a teaching environment. Tutors felt the existence of teaching materials would encourage more extensive instructional use of this type of data. Suggested teaching materials included thematic case studies and teaching datasets (to allow demonstrations of particular estimation and inference methods), and some supporting guidelines and notes to explain how the data might be used. Suggested themes included 'Science, technology and innovation', the 'United Nations (UN) Millennium Development Goals (MDGs)', and 'Environment and economy'.

Access problems were highlighted by a number of tutors - some commented that the registration process was too complicated to get all their students registered prior to using the datasets in a lesson. Feedback about the provision of support services indicates that courses should be made available via the web, since face to face courses are not always convenient to attend due to timing and location.

In 2007, ESDS International started to address these issues with the launch its first major learning resource entitled, "Countries and Citizens".

This self-guided training resource, including online tutorials, activities, study guides and videos, is designed to show how to combine socio-economic data from country-level aggregate databanks (macro data) with individual-level survey datasets (micro data). This resource has proved extremely popular with an average 20,000 page views per year since its release.

The extraordinary level of usage combined with a growing demand for entry level statistical learning resources fuelled the decision to develop a new online teaching resource based on the MDGs. In addition, the development of a new learning package based on the MDGs aligns well with ESDS International's strategy to promote and encourage the use of international databanks in teaching and research and helps to fulfil its responsibility to raise statistical capacities.

The UN Millennium Development Goal e-learning materials

The UN's MDGs offer a wealth of freely available, cross-national, policy relevant data. The goals themselves have formed the policy framework for international development through public investment, capacity building, domestic resource mobilization and the targeting of official development assistance. Progress towards each goal is measured by a series of quantifiable indicators. For example, *Millennium Development Goal 5: Improve maternal health* is measured by the following quantifiable indicators for each country:
- Maternal mortality ratio
- Proportion of births attended by skilled health personnel
- Contraceptive prevalence rate
- Adolescent birth rate
- Antenatal care coverage (at least one visit and at least four visits)
- Unmet need for family planning

While the MDG data is freely available, there are still many potential data users who are simply unaware of how this global data resource could better inform their research and teaching. In order to rectify this issue the ESDS International service has developed a suite of e-learning materials based upon the socio-economic theme of the UN's MDGs. The suite has been designed as an interactive online set of materials to be used as a self-teaching resource intended to raise awareness of international data resources and help students develop their own data handling skills using real world data. The online materials are targeted at postgraduates and undergraduates mirroring the current user profile of the ESDS International service which counts almost 75% of its registered users as postgraduate or undergraduate students.

The learning package is available as an open access resource via the ESDS International data service. This fact meant the course had to cater to users with differing levels of expertise in statistical analysis techniques and/or awareness of international data resources.

Implementation

A number of possible implementations to satisfy the requirements for the MDGs learning package were considered. For example, we looked into developing a blended approach (i.e., face–to-face and virtual elements) but decided we did not have the resources to dedicate to this. In addition, it was thought that there would be some key benefits in developing an entirely online e-learning package, for example, it would be available to any institution at any time and users can pick and choose elements from the package for inclusion in their specific learning environment.

The e-learning package was developed as web-based materials rather than within a Virtual Learning Environment (VLE) such as Blackboard, as

ESDS International is a national data service and, as such, must ensure its resources are available to users from any UK institution no matter what VLE their institution supports.

Consultation with a number of key data users who use international data in the classroom contributed to the identification and development of the e-learning package contents. The content was produced by Susan Noble and Celia Russell, Richard Wiseman developed the e-tutorials based on Camtasia Studio 5. Susan Noble also developed the framework and programming for the resource.

The MDG e-Learning Resource

The e-learning materials have been designed primarily to help learners explore the United Nation Millennium Development Goals. In addition, they guide participants through using the ESDS International data service and provide learners with an increased understanding of data availability and potential data use for research and teaching. The materials comprise four main sections:

1. Guide to ESDS International - provides a comprehensive overview of the international data service including the topics, time range, frequency and geographical coverage of the data provided.
2. Guide to the MDGs - describes the background to the MDGs, the relationship between the goals, targets and indicators, identifies the data sources and how progress towards the goals is measured.
3. Activities section - provides step-by-step activities enabling learners to, for example, search the ESDS International data portfolio or carry out a simple data visualisation task. All activities are available as e-Tutorials.
4. Links and resources - contains links to relevant UN MDG related interfaces such as UN MDG Gapminder and a CommonGIS interface to a subset of MDG data.

The resource is based on a blend of instructional and constructivist elements, allowing students to discover the content by a variety of means. For example, in the guides, the target material is structured into a formal hierarchy, enabling the learner to work through sequentially. Alternatively, students can tackle real tasks using the data by attempting related activities at the end of the section.

One particularly successful element of the e-learning materials is the e-tutorials. These are video screen capture demonstrations built using Camtasia Studio 5. Five e-tutorials were developed collectively demonstrating how to access, search, source, visualize and cite cross-national data.

The resource uses real-life examples drawn from the MDG indicators as a basis for the interactive activities. For example, in the data access e-tutorial, students look for data that supports two of the indicators for MDG Goal 5: improve maternal health. By basing the activity on authentic data gathered for a genuine purpose (maternal deaths represent the greatest indicator of inequity between rich and poor women) students discover for themselves the worrying fact that in many countries there has been little improvement in maternal death rates over the past two decades. These hands-on activities, which also explore poverty and hunger (Goal 1) and women in parliament (Goal 3), are key constructivist elements of the resource. As they move through the activities, students learn how to access, visualise and cite real MDG data by taking part in these authentic tasks. They can also discover how to take action in support of the MDGs by exploring the 'What you can do?' area of the Links and Resources section.

Future work for these materials could include the development of teaching datasets, including simple data analysis tasks. Further, the

materials could be made available as a re-usable Content Package which can be downloaded into an institutional VLE.

Conclusions

Within this paper, we have discussed the barriers that prevent international data sharing, relevant data access and licensing issues and mentioned ways in which these barriers can be reduced.

The UK model for providing access to international data through national licensing agreements has shown to be a very effective way of increasing the use of cross-national datasets and is one that can be replicated elsewhere.

Traditionally most intergovernmental organisations make some of their data freely available, but large portions are only available at a cost. This trend is reversing. In recent years the United Nations and the International Labour Organisation have made data freely available. In April 2010, the World Bank publicly released all of their data. The removal of the key barrier of cost is, of course, beneficial, as a larger amount of people will be able to access the data and indeed the World Bank has highlighted this increased opportunity to turn data into knowledge[5].

We have also introduced e-learning materials based upon the UN's Millennium Development Goals, and highlighted how this approach reduces the gap between the richness of available data resources and their uptake. A central aspect of the new resource is that it is completely based on genuine data collected to monitor progress towards international development. Real world data is often messy, in particular from the developing world where "exceptions are the rule"[6] . For this reason in many teaching environments data is pre-fabricated, outliers removed, missing data omitted, resulting in sufficient data for a problem-free analysis. However, outliers provide interesting insights into the issue at hand and managing missing data is a skill often required in the outside world. Handling real world data gives students important skills they can use in their later careers. Moreover, the pre-fabrication of data disengages students from the real-life aspects of the task.

The MDG indicators monitor ongoing processes and, as such, are constantly being modified, updated and extended. As the Millennium Development Goal e-learning materials described here sit on top of the UN's own MDG indicator data, they are well-suited to accommodate this changing data environment. In other words, by delivering learning resources through live web interfaces, underlying data can evolve naturally, thus engaging students with the changing world they see around them.

ESDS International's teaching and learning resources, including the UN Millennium Development Goal e-learning materials can be found at: http://www.esds.ac.uk/international/resources/learning.asp

Notes

1. Susan Noble, Dr Celia Russell and Richard Wiseman. All of ESDS International, Mimas, University of Manchester, UK. (contact at: susan.noble@manchester.ac.uk, celia.russell@manchester.ac.uk, richard.wiseman@manchester.ac.uk).
This work was presented in session 'C2: Sharing Data: High Rewards, Formidable Barriers' at the 2009 IASSIST conference.
2. Angel Gurría, Secretary-General, OECD. Opening speech at the OECD Forum 2008: http://www.oecd.org/dataoecd/25/5/40762048. pdf.
3. Databeuro - http://www.databeuro.com/.

4. JISC and Publishers Association Model Licence - http://www.jisc. ac.uk/aboutus/committees/workinggroups/disbanded/standardli- censing/report.aspx
5. Dr Eric Swanson, World Bank (2010) 'Working with International Development Data', University of Manchester: http://www.esds. ac.uk/international/news/news.asp#16dec10a
6. Levine, Joel H, 1993 Exceptions Are the Rule: An Inquiry into Methods in the Social Sciences Boulder, Westview Press,