UNITED NATIONS STATISTICAL COMMISSION and ECONOMIC COMMISSION FOR EUROPE CONFERENCE OF EUROPEAN STATISTICIANS

ORGANISATION FOR ECONOMIC COOPERATION AND DEVELOPMENT (OECD) STATISTICS DIRECTORATE

Joint UNECE/OECD Work Session on Statistical Dissemination and Communication
(14-15 February 2005, Henley-on-Thames, United Kingdom)

Topic (ii): Sketching out the future of statistical publications

DELIVERING THE WORLD:
THE ESTABLISHMENT OF AN INTERNATIONAL DATA SERVICE

Invited Paper

Submitted by ESDS International, MIMAS, University of Manchester[1]

ABSTRACT

Economic and Social Data Service (ESDS) International is a UK national data service providing web-based access to the major socio-economic databanks produced by inter-governmental organisations such as the World Bank and the United Nations. The UK Office for National Statistics Time Series Data is also included in the service. Access to the data is free for the UK academic community and the service has thousands of users from a broad range of social science disciplines.

In this paper we examine how these data is used in teaching and research and some of the problems academics encounter trying to find the data they need. We report on the challenges of converting large and complex datasets from a range of sources into a common user-friendly interface and the role of visualisation. The paper also discusses the data acquisition strategy, the establishment of national licensing agreements with the data providers and how the ESDS International model could be extended to other countries.

Finally, our paper outlines the strategies and value added services employed to promote the use of these previously under-utilised databanks in teaching and research.

Keywords: International data, Social Sciences, macro-economic databanks, Economic and Social Data Service, ESDS International.

---

[1] Prepared by: Celia Russell, Keith Cole, Susan Noble and Nicholas Syrotiuk (ESDS International, MIMAS, University of Manchester, Manchester, M13 9PL, United Kingdom).

I.   INTRODUCTION

1.   June 2003 saw the launch of the Economic and Social Data Service (ESDS), a national UK data service which provides dissemination, user support and training for an extensive range of key economic and social data, both quantative and qualitative, for the UK academic sector. The service, funded by the Economic and Social Research Council (ESRC) and the Joint Information Systems Committee (JISC) comprises four specialist data services: ESDS Government; ESDS International; ESDS Longitudinal and ESDS Qualidata. It has been established as a distributed service, based on a collaboration between four UK centres of expertise – the UK Data Archive (UKDA) and Institute for Social and Economic Research, both based at the University of Essex, and Manchester Information and Associated Services (MIMAS) and the Cathie Marsh Centre for Census and Survey Research, both based at the University of Manchester.

2.   This paper describes the development of ESDS International, the specialist data service which provides the UK academic community with free online access to the statistical databanks produced by international governmental organizations such as the World Bank and United Nations. The service also provides access to international survey datasets.

II.   BACKGROUND

3.   International Governmental Organisations (IGO's) such as the International Monetary Fund (IMF), the United Nations, the World Bank and the Organisation for Economic Cooperation and Development (OECD) have long produced high quality, regularly updated time series databanks for their own internal use. These databanks typically contain a huge range of macro-economic and social indicators aggregated to national or regional level and collectively cover virtually every country in the world.

4.   Recent years have seen a growing requirement from the academic sector for access to these types of data. As the integration of national societies and economies accelerates, the accompanying issues of growth, power and inequity have attracted increasing interest from the research community. Access to international databanks also allows researchers to make cross country comparisons and interpret their findings in a broader perspective. Moreover in social science fields such as crime, employment or economics, a research question that may have been a single country study a few years ago now requires examination in an international context. Issues such as climate change, the spread of infectious disease, international security or other collective global problems can only be addressed at an international level and the academic sector requires an evidence base in order to contribute and comment on these trans-national policy responses. ESDS International was established to address these needs through the provision of free web-based access to a portfolio of authoritative, high quality international databanks.

5.   International bodies such as the World Bank or International Monetary Fund 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. The result is that IGO's produce databanks of tremendous quality, scope and potential value to the academic sector. However, until now these databanks have largely been unavailable or very expensive for people outside these organisations to use with only those academics belonging to an institution with an institutional subscription to each IGO having access to the data.

6.   The establishment of ESDS International gave us an opportunity to remove some of these barriers to use by adopting a new and unique approach to the delivery of international data to the UK academic community.

III.   DATA ACQUISITION STRATEGY

7. The strategy developed by ESDS International for the acquisition of data aimed in the first instance to ensure that the UK academic community would have continued access to regularly updated versions of the international macro datasets previously available through various service such as the UK Data Archive. This initial data portfolio had a strong macro-economic theme covering topics such as economic performance, trade and international flows of capital and included the IMF's Direction of Trade Statistics, International Financial Statistics, Balance of Payments, Government Finance Statistics, the OECD's Main Economic Indicators and UNIDO Industrial Statistics and Demand Supply databases.

8. Additional funding from the ESRC enabled the service to enhance the data portfolio to include datasets on the social effects of greater global interdependency, economic growth, industrialization and changes in public policy priorities. These datasets cover topics such as human development, globalisation, migration, labour markets, social expenditure, demography, environment, education and science and technology.

9. A variety of selection criteria were used to evaluate whether a dataset should be included in the portfolio. A literature survey using the John Rylands University Library of Manchester e-journals service was carried out on leading journals covering a range of social science disciplines and where relevant, the empirical data section of each paper was reviewed to identify any international datasets used in the research.

10. A data mapping exercise provided content mapping between the topics covered in each potential dataset, the ESRC research priorities and the 19 subject categories used by the Council of European Social Science Data Archives to classify socio-economic datasets. An ESDS user consultation survey identified international datasets potential users of the service would like to gain access to.

11. Priority was given to 'research quality' datasets, that is those that are regarded internationally as being key sources of high quality, authoritative, reliable and up-to-date statistics with good temporal and spatial coverage. The databanks were to provide long term potential for research and teaching, with long and consistent time series, relatively stable data domains and strong opportunities for comparable research. Final consideration was given to those datasets where prohibitive data license costs were previously an obstacle to use.

12. The data acquisition strategy identified a data portfolio that collectively would chart over 50 years of global economic, industrial and social activity.   The ESDS International macro data portfolio is shown in Appendix 1

13. Through a series of reciprocal agreements with Data Archives worldwide, ESDS International also provides access to international survey datasets such as Eurobarometer, International Social Survey Programme and the World and European Values Surveys. These micro level datasets cover a range of social science topics including household and demographic information, income, employment, education and housing.

IV.   DATA RE-DISTRIBUTION LICENSING

14. One of the key achievements of ESDS International has been the successful negotiation of UK wide academic data redistribution agreements with a number of IGOs including the International Monetary Fund, the World Bank, the Organisation for Economic Co-operation and Development (OECD), the United Nations, and the International Labour Organisation (ILO). In many instances, this was the first

time that an IGO had agreed to country wide data redistribution agreements for academic access to their databases.

15. Previously, in order for a UK academic to obtain access to SourceOECD, UN Common Database and the World Bank's Development Indicators and Global Development Finance statistics, their institution would have to have taken out an institutional subscription. The requirement for an institutional subscriptions, which have to be funded by the library or a department, has significantly reduced the number of UK academics who could have access to key international datasets and it also made it difficult for academics from different institutions to undertake collaborative research. For example, although the UN Common Database is a relatively new product, there were in May 2003 a total of 185 academic subscribers worldwide of which 81 were in the US with only 2 academic subscribers in the UK.

16. The aim was to negotiate a UK wide redistribution agreement that would deliver significant savings to the community as a whole and remove one of the major barriers to use of 'commercial' datasets in research and teaching. In some instances, additional discount was provided by negotiating a five-year deal with a single up-front payment in year 1. Indeed, the consortium purchase approach for the portfolio has produced significant savings and opened up access to international macro data to the entire UK academic community.

17. For the bulk of the licence agreements, ESDS International employed the services of Databeuro - a professional data negotation agent. Databeuro (http://www.datbeuro.com) were selected as they already had expertise in working with intergovernmental organisations (IGO's) such as OECD, IMF, World Bank, and UN for access to on-line data and e-books to the academic, government and corporate sectors.

18. Where possible, the data redistribution agreement negotiated with each IGO was based on a model licence. The model licence adopted was based on one developed between the JISC and the Publishers Association for the licensing of commercial datasets (http://www.jisc.ac.uk/index.cfm?name=wg_standardlicensing_report). The model licence set out generic terms and conditions of use (e.g. use of data for educational purposes only) but there was also scope for adding additional clauses to reflect special conditions required by different suppliers. For example, the World Bank requires a user to obtain written permission if they wish to use more than 10 indicators in teaching materials. In circumstances where an IGO had its own model licence agreement (e.g. OECD) it was necessary to get them modified to ensure that they were in line with all the other agreements in terms of key definitions and conditions of use.

19. As the licence agreement was between the University of Manchester (licensee) and the IGO (licensor), 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. As a result, it was necessary to ensure that appropriate clauses relating to limitation of liability and dispute resolution were inserted to protect the University of Manchester.

V.   DELIVERING THE DATA

20. Complex, non web-based interfaces have long been associated with accessing international databanks.   MIMAS for example had previously offered a limited international data service providing access to datasets such as the UNIDO Industrial Statistics via a complicated X-Windows interface which meant knowledge of the UNIX operating system and SAS statistical software package were essential in order to access the data. Each IGO provides custom interfaces to their datasets so a user

wanting to access data from a number of sources would have to navigate a number of different interfaces. A key element therefore, to our data dissemination strategy was the provision of a web-based, user-friendly interface which would be common to all databases within the portfolio. It was also decided to adopt a proprietary solution to minimize interface development and maintenance therefore giving us more time to concentrate on value added activities.

21. Beyond 20/20 Web Data Server (WDS), a web-based data dissemination tool was chosen as the software solution to provide our common user interface. It requires only a standard web browser, is accessibility compliant and can be used to display, subset, visualise and download time series data. In addition it is recognized within the UK Social Science community as the Office for National Statistics use the same interface to deliver their Neighbourhood Statistics.

22. In order to deliver the data via Beyond 20/20 WDS it was first necessary to convert the large and complex files from a range of sources into Beyond 20/20's own table format. This process was problematical as data was provided in various formats with differing quantities and quality of explanatory documentation which in some cases proved difficult to match with a generic product like WDS.

23. The first stage was to understand the contents and structure of the datasets by analysing the data provider's original data files, documentation and user interface. Once the structure of the database was understood it was possible to interpret how this could be presented. Data processing programs were written to re-format the raw data, to load the data and to publish the tables on the web server. Each dataset consisted of a unique set of data files, requiring a unique set of processing programs which varied considerably from dataset to dataset.
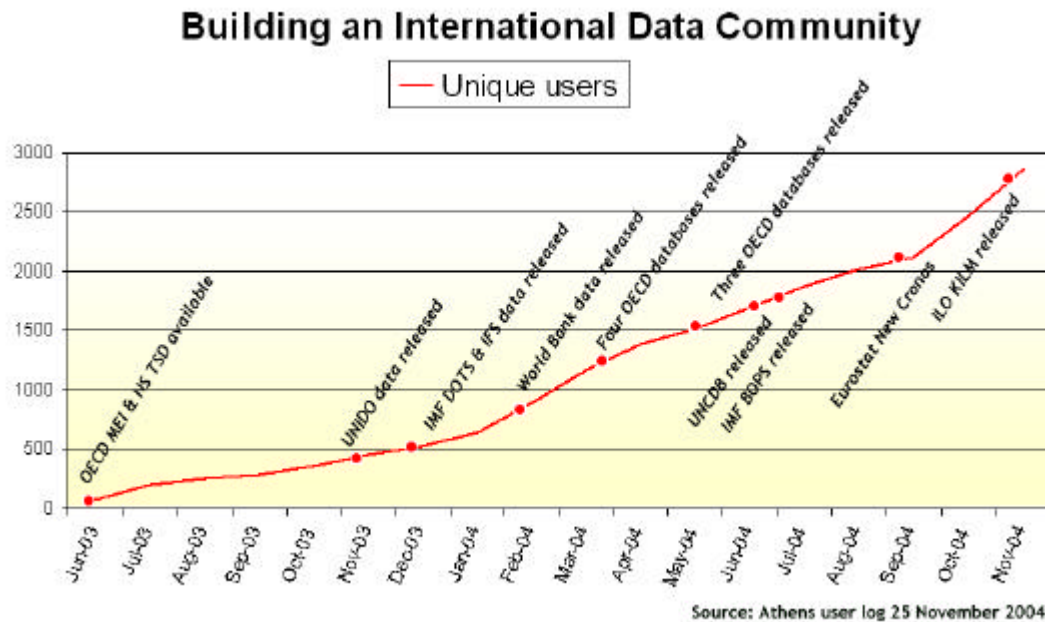
24. A further complication to the delivery of the international data was the requirement to restrict access to UK academic users only. Users are required to register and to authenticate before they can access the data. Registration is completed by accepting the ESDS end user license conditions and authentication is provided by Eduserv Athens using their access management system.

25. A key element of ESDS International's data delivery strategy was to enable users to search across the full portfolio of international datasets. However, as the data comes from various data providers with no common way of describing their data, it was necessary to create collection level descriptions of the datasets. This was accomplished by creating metadata records containing subject headings and classification numbers. These were all assigned using international metadata standards, including Dewey Decimal Classification, Library of Congress Subject Headings, Humanities and Social Science Electronic Thesaurus and the UNESCO Thesaurus. Users can currently search in the ESDS catalogue across all ESDS International's datasets and within the MIMAS metadata database across all of MIMAS's services.

VI.   BUILDING A USER COMMUNITY

26. Another key activity in the establishment of the international data service has been the promotion of international data use to raise awareness of the datasets and their research potential. Awareness days and workshops have been organised by ESDS and the service ensures it has a presence at relevant social science events. Introductory and advanced training courses on data analysis are provided and an online JISC mailing list is used to disseminate information to users of international data, including details of new data releases, updates, courses, workshops and events.

27. The establishment of the service has created a growing community of international data users. The service now has over 3,600 users with increases of several hundred users each month. The release of new datasets (i.e. UNIDO data in November 2003, IMF data in December 2003, World Bank data in February 2004) and Eurostat (November 2004) has also generated a substantial increase in users.

Building an International Data Community

Source: Athens user log 25 November 2004

28. One of ESDS International's aims was to attract users from a broad range of disciplines both in FE and HE and analysis of the ESDS International user-base allows us to see who is actually using the service. Appendix 2 lists the number of users from each discipline as recorded in the ESDS user registration database. As expected economists are the largest single user group but users from other subject areas such as Accounting and Finance, Environment, Geography, Sociology, Politics, International Studies, History, Pure Mathematics, Health, Asian Studies and Engineering are also represented. Data Librarians and data support staff are the second largest group highlighting the key role they have in promoting and supporting use of the service locally. The majority of our users come from Higher Education institutions, but a very small percentage of users (1.5%) are from Further Education institutions. The below table presents a breakdown of the types of users accessing the data showing that post-graduates are the largest single group with student users in total forming over 60% of the registered users.
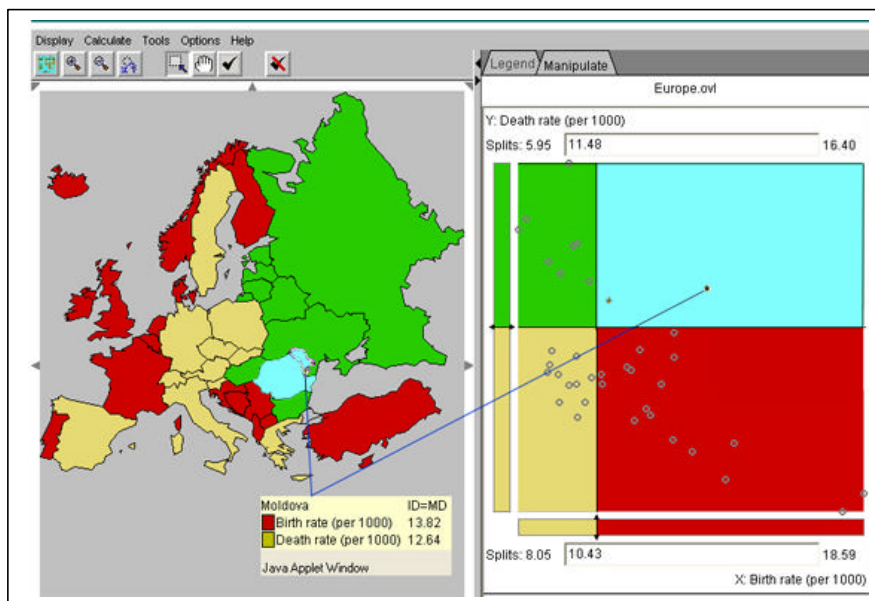
ESDS Registered Users (Jun 2003 – Feb 2004): User Type	No. of users (836 in total)	% of total
Staff – Class Tutor	110	13.2
Staff – Other	210	25.1
Postgraduate Student	334	39.9
Undergraduate Student	156	18.7
Student – Other	23	2.8
Academic Visitor	3	0.3

VII.   VALUE ADDED ACTIVITIES

29. In order to further promote the use of international databanks and increase the service's effectiveness across a broad range of social science disciplines and user types, a number of 'value added' services are offered by ESDS International.

30. This includes the provision of specialist advice via a dedicated helpdesk where the team's familiarity with the contents and structure of the various databases within the portfolio enables timely responses and has proved integral in building service-user relationships. Online teaching and learning materials, a structured guide to freely available international data and regularly updated frequently asked questions are provided from the service's website. A programme of introductory and advanced training courses is provided and can be carried out at individual institutions on request.

31. In addition to the data delivery, the web-based geographical information system software CommonGIS has been used to create an interactive exploration and visualisation interface to a set of freely available international data to encourage the use of the data in learning and teaching. This has proved very popular at ESDS International workshops and our aim is to produce a number of themed GIS interfaces to the datasets we host using world boundary data.

CommonGIS cross classification map for Europe. Y axis = Death rate, X axis = Birth rate

VIII. CONCLUSION AND THE FUTURE FOR ESDS INTERNATIONAL

32. This paper has presented the approach taken by ESDS International to provide the academic community in the UK with access to regularly updated, high quality international data.

33. The service monitors demand for international datasets and works closely with ESRC to identify any additional international datasets that might be required to support future research programmes.

34. The service will look at published research based on the databanks to discover how the data provided by ESDS International is being used and the types of research question being addressed through the use of international data.

35. Promotional activities and materials will continue and be enhanced, for example with the provision of themed guides on concepts implicit in the databanks such as economic stability and growth and the development of support materials on the possibilities of linking macro and micro international datasets.

36. A review of requirements to support teaching use of the service will be carried out and used as the basis for developing a guide to creating teaching datasets and using international macro data in teaching. There will be a particular emphasis on raising awareness of the use of international data within teaching in order to encourage a new generation of quantitative data users with the knowledge, skills and confidence to utilize international data resources in their future work.

37. Through ESDS International, the UK has demonstrated how a successful a model licence consortium based approach can be for the international databanks. It has also proved beneficial for the IGO's in that it has considerably increased the use and awareness of their databanks in the academic community. A number of other countries have expressed an interest in developing similar consortium agreements based on our model, and we believe the success of the service, combined with the internationalisation of social science research itself, will prompt the creation of similar services elsewhere.

Appendix 1: ESDS International macro data portfolio

OECD Main Economic Indicators
OECD Main Science and Technology Indicators
OECD Quarterly Labour Force Statistics
OECD Social Expenditure Database
OECD Measuring Globalisation
OECD International Development
OECD International Direct Investment Statistics
OECD International Migration Statistics
OECD Services Statistics
UK Office for National Statistics Time Series Data
UNIDO Industrial Statistics
UNIDO Industrial Demand Supply
IMF Direction of Trade Statistics
IMF International Financial Statistics
IMF Balance of Payment Statistics
IMF Government Finance Statistics
World Bank World Development Indicators
World Bank Global Development Finance
United Nations Common Database
Eurostat New Cronos
ILO Key Indicators of the Labour Market

Appendix 2: ESDS International Registration Statistics – Discipline

836 Users registered with ESDS in period from June 2003 – February 2004.

Economics and Econometrics (264)
Library or Data/Information (118)
Economics, Labour and Employment (104)
Business and Management (70)
Business Studies and Accountancy (44)
Geography (26)
Accounting and Finance (22)
Politics and International Studies (22)
Economic and Social History (16)
Sociology (14)
Environment, Housing and planning (13)
Social Administration and Social Policy (9)
Computing Service (8)
Education (8)
Statistics, Computing and Methodology (7)
European Studies (6)
Political Science and International relations (6)
History (5)
Administration (4)
Communication, Cultural (4)
Computer Science (4)
Engineering (4)
Manufacturing Engineering (4)
Psychology (4)
Asian Studies (3)
Built Environment (3)
Health and medicine (3)

Hotel, Catering, Travel (3)
Mathematics (3)
Statistics and Operational Research (3)
Agriculture (2)
Chemistry (2)
Electrical and Electronic Engineering (2)
Environmental Sciences (2)
Information Science (2)
Law (2)
Librarianship, Information (2)
Medieval and Modern History (2)
Philosophy (2)
Anthropology (1)
Archaeology (1)
Art and Design (1)
Biological Sciences (1)
Biology and Biochemistry (1)
Classics, Ancient History (1)
Community based clinical (1)
Mathematics (1)
Mechanical, Aeronautical and
Media Studies, Journalism (1)
Other studies and Professions Allied to medicine (1)
Research Council (1)
Social Anthropology (1)
Socio-legal Studies (1)
Town and Country Planning (1)