



Unit 1: Study Guide
Linking International Macro and Micro Data – the basics
MIMAS
The University of Manchester

1.1. Introduction

1.2. Learning objectives

1.3. Data concepts

1.3.1. Survey and aggregate data

1.3.2. Other types of database (longitudinal, qualitative, meso, panel)

1.3.3. Creating surveys (questionnaires, interviews)

1.3.4. Creating aggregate databanks

1.4. International data resources

1.4.1. An overview

1.4.2. ESDS International

1.4.2.1. World Bank databanks

1.4.2.2. IMF databanks

1.4.2.3. OECD databanks

1.4.2.4. IEA databanks

1.4.2.5. UN databanks

1.4.2.6. Eurostat

1.4.2.7. Selected sources of freely available data

1.4.2.8. Data collection activity, Beyond 20/20

1.5. Micro data

1.5.1. An overview of the available micro data sets

1.5.1.1 European Social Survey

1.5.1.2 World Values Survey

1.5.1.3 The global barometers

1.5.1.3 Other international surveys

1.5.2. Data collection activity, Nesstar

1.6. Further exploration

1.6.1 Further exploration, locating data activity

1.6.2 Further exploration looking at sources

1.7. Summary / Reading lists

1.1 Introduction

Recent years have seen a growing requirement from the academic sector for access to high quality international data. As the integration of national societies and economies accelerates, the accompanying issues of globalisation, migration, inequity and growth have attracted increasing interest from the research community. Access to international databanks also allows researchers to make cross country comparisons or interpret their findings in a broader perspective. In social science fields such as crime, employment or health, 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, energy 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.

Through ESDS International, the ESRC provides the UK academic community with access to a stock of world class, authoritative data resources. The service is unique in that it provides online access, free at the point of use, to a comprehensive range of international databanks. These include macro (or aggregate) databanks such as those produced by the World Bank and the United Nations, and micro (or survey) datasets such as the European Social Survey. One of the objectives of this course is: **to promote increased and more effective use of both aggregate and survey data in international comparative research.**

However, there still exists a lack of understanding of how best to establish linkages between international macro and micro data. These linkages are important, not only for reasons of consistency, but also to take account of wider influences on individual level responses and allow researchers to explore how survey data can inform aggregate data research. Hence, the second key objective of this course is to: **increase awareness of the methodological, statistical and technical issues associated with combining aggregate and survey data.**

This first unit covers the basics of international **aggregated macro data** and **survey data**. What are these different data types and how are they produced? The unit introduces the major international aggregate and survey datasets available, and describes the range of themes they cover. We will be looking at how to locate the

data you need and using an example drawn from the World Bank World Development Indicators, the unit will guide you through accessing, subsetting and charting data using Beyond 20/20.

1.2. Learning objectives

By the end of this module you will be able to:

- Understand the basic conceptual differences between aggregate and survey data
- Know how international data is produced and disseminated, and the role of ESDS International
- Know more about the organisations that produce the international databanks
- Outline the content of the major international aggregate databases
- Access, subset and chart data from the World Development Indicators
- Outline the content of the major international survey datasets
- Search for specific data using the ESDS International resources
- Look up original sources and definitions



Example 1 : anonymised data

This is an example of what survey microdata looks like once it's been read into a statistical software package, in this case SPSS.

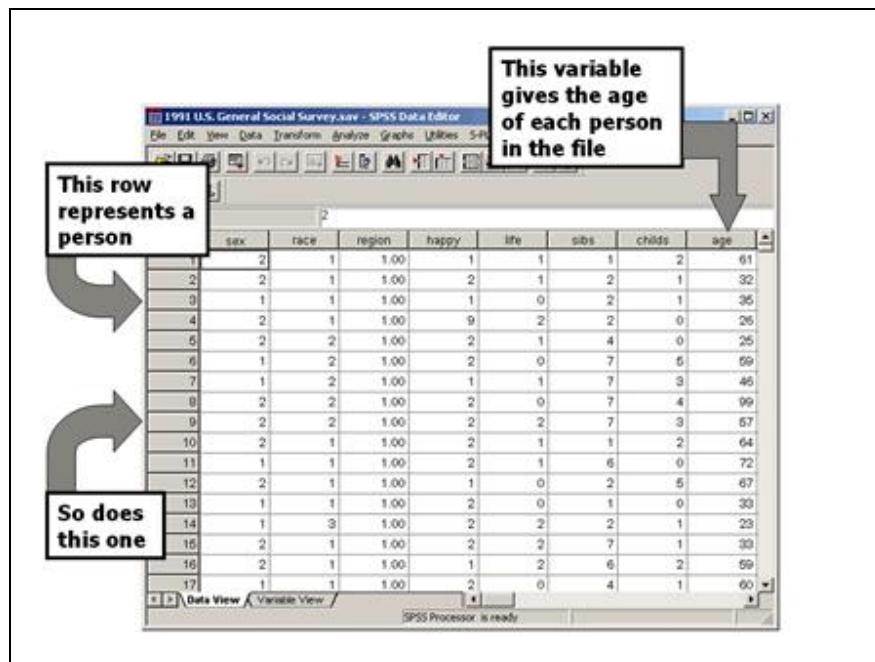


Figure 2: screenshot showing SPSS data view

In Figure 2: we are looking at a screen shot in SPSS data view. In this social survey, each row represents a person. Each column is a variable, a piece of information that was collected about each person.

Figure 3: shows SPSS's variable view, which displays information about the variables that has been entered from a codebook. This is metadata - it is not part of the data itself.

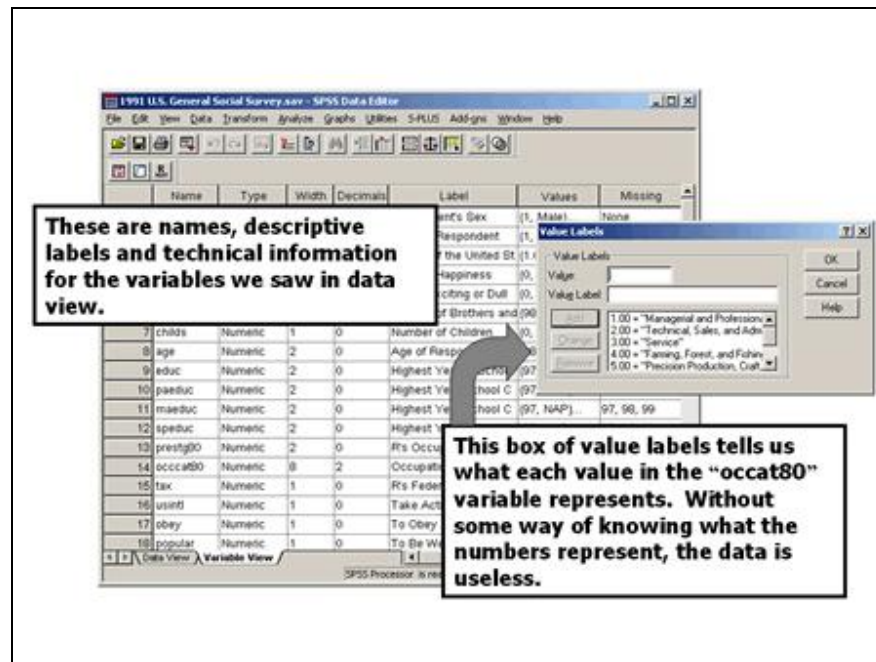


Figure 3 : screenshot showing SPSS variable view

This box tells us how occupation was coded in this dataset. For example, the variable for occupation is **occ80**; a **'1'** means that the person was managerial, a **'2'** that they worked in a technical position, and so on.

A key property of a survey dataset is that it allows you to understand the relationships between different characteristics at the level of the individual such as whether the level of a person's education affects the number of children they have. This makes survey data a very powerful resource for policy makers. **ESDS government** provides access and support for the UK large scale government surveys such as the **General Household Survey** and the **Labour Force Survey**.

Macro or **aggregated** data are data about populations, groups, regions or countries, Examples of aggregate indicators are life expectancy or employment rates. These are data that have been averaged, totalled or otherwise derived from the individual level data found in the survey datasets. Most of the aggregate data we are going to talk about in this section is data about countries



Example 2 : Beyond 20/20

In this screen shot we see some typical aggregate data in Beyond 20/20. The indicator we are looking at is the number of marriages. Time runs across the top and countries down the side.

Table View - Microsoft Internet Explorer

File Edit View Favorites Tools Help

Address http://devs.mcs.ac.uk/WDCI/LINKTableViewerPublications.aspx

Microsoft and Excel are trademarks

Reports Marriages, urban and rural [code 150820] Help

Actions

OTHER:

Year		1948	1949	1950	1951	1952	1953	1954	1955	1956	1957	1958	1959	1960
Class 52	Country or Area	000	000	000	000	000	000	000	000	000	000	000	000	000
	Germany, Federal Republic of	493,806.0	476,306.0	505,708.0	527,944.0	493,388.0	462,101.0	453,168.0	465,818.0	479,373.0	482,590.0	494,110.0	501,983.0	513,445.0
	Gibraltar	207.0	191.0	210.0	215.0	222.0	211.0	297.0	307.0	264.0	338.0	239.0	484.0	543.0
	Greece	42,600.0	42,118.0	58,482.0	63,245.0	49,644.0	60,909.0	-	-	55,233.0	68,818.0	69,178.0	74,213.0	58,168.0
	Greenland	154.0	343.0	249.0	224.0	187.0	203.0	192.0	218.0	195.0	212.0	207.0	286.0	-
	Guadeloupe	1,576.0	1,501.0	1,151.0	1,049.0	1,265.0	1,183.0	2,486.0	1,223.0	1,254.0	1,352.0	1,275.0	1,410.0	1,593.0
	Guam	-	477.0	351.0	371.0	406.0	365.0	343.0	339.0	386.0	333.0	279.0	383.0	443.0
	Guatemala	8,095.0	9,332.0	10,344.0	10,584.0	11,381.0	11,618.0	12,915.0	12,167.0	16,230.0	17,094.0	16,783.0	19,258.0	16,561.0
	Guernsey	389.0	460.0	413.0	380.0	338.0	310.0	314.0	300.0	363.0	373.0	338.0	325.0	361.0
	Guinea-Bissau	435.0	608.0	-	-	-	-	-	-	-	-	-	-	-
	Guyana	1,249.0	2,003.0	1,983.0	1,942.0	1,943.0	1,889.0	1,949.0	2,097.0	2,089.0	2,524.0	1,948.0	2,098.0	2,259.0
	Holy See	-	-	-	-	-	-	-	-	-	-	-	-	-
	Honduras	-	-	-	-	-	-	-	-	-	-	-	-	-
	Hungary	97,710.0	107,820.0	106,245.0	93,362.0	104,836.0	91,622.0	107,368.0	103,020.0	94,133.0	97,992.0	92,439.0	95,313.0	88,566.0
	Iceland	1,142.0	1,877.0	1,217.0	1,199.0	1,151.0	1,225.0	1,417.0	1,339.0	1,316.0	1,315.0	1,331.0	1,345.0	1,309.0
	Indonesia	-	-	-	-	-	-	-	-	-	-	-	-	1,213,622.0
	Iran (Islamic Republic of)	-	-	-	-	-	-	-	-	-	-	-	-	-
	Iraq	-	-	-	-	-	-	-	-	-	-	-	-	24,145.0
	Isle of Man	480.0	451.0	428.0	409.0	380.0	333.0	326.0	352.0	363.0	343.0	297.0	332.0	326.0
	Israel	-	12,076.0	16,019.0	-	17,598.0	15,329.0	14,515.0	14,742.0	15,098.0	15,983.0	16,562.0	16,143.0	16,532.0
	Italy	189,034.0	199,722.0	198,494.0	230,610.0	237,047.0	242,814.0	259,911.0	266,718.0	265,734.0	265,243.0	273,752.0	281,222.0	287,683.0
	Kazakhstan	-	-	-	-	-	-	-	-	-	-	-	-	-
	Kiribati	-	-	311.0	385.0	381.0	377.0	338.0	-	-	-	-	-	-

Start

Microsoft PowerPoint - [0...]

Table View - Microsoft

Internet

Figure 4 : Beyond 20/20 data view

Aggregate country data is usually in the form of time series data, with repeated observations of the same indicator for the same country at regular intervals.

World CO₂ emissions from energy use, by region

Million tonnes

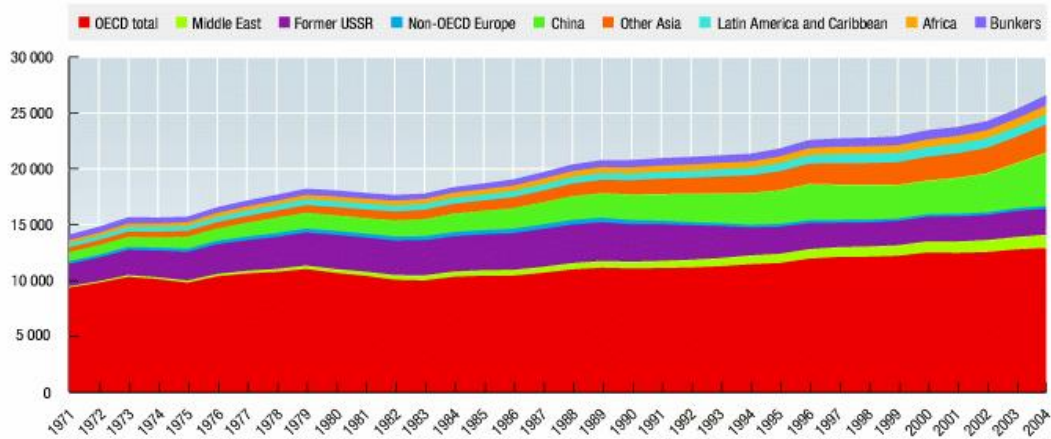


Figure 5 : Time series data showing CO₂ emission by region, source OECD Factbook 2007, see Statlink : <http://dx.doi.org/10.1787/363278342610>

"Global emissions of carbon dioxide have risen by 88% since 1971 and are projected to rise by another 52% by 2030. In 1971, the current OECD countries were responsible for 66% of the total. As a consequence of rapidly increasing emissions in the developing world, the OECD contributed 49% to the total in 2004, but this is expected to fall to 38% by 2030. By far, the largest increases in non-OECD countries occurred in Asia, where emissions in China have risen by 5.5% per annum between 1971 and 2004. The use of coal in China increased levels of CO₂ by 3.2 billion tonnes over the 33-year period.

Two significant downturns can be seen in OECD CO₂ emissions, following the oil shocks of the mid-1970s and early 1980s. Emissions from the economies in transition declined over the last decade, helping to offset the OECD increases between 1990 and the present. However, this decline did not stabilise global emissions as emissions in developing countries grew.

Disaggregating the emissions data shows substantial variations within individual sectors. Between 1971 and 2004, the combined share of electricity and heat generation and transport shifted from one-half to two-thirds of global emissions.

Fossil fuel shares in overall emissions changed slightly during the period. The relative weight of coal in global emissions has remained at approximately 40% since the early 1970s. The share of natural gas has increased from 15% in 1971 to 20% in 2004. Oil's share decreased from 49% to 40%. Fuel switching and the increasing use of non-fossil energy sources reduced the CO₂/total primary energy supply (TPES) ratio by 7% over the past 33 years."

Source OECD Factbook 2007,
<http://lysander.sourceoecd.org/vl=29027040/cl=23/nw=1/rpsv/factbook/>

1.3.2 Other types of database

Other types of database used in the social sciences include *longitudinal*, *panel*, *qualitative* and *meso*.

Longitudinal or **panel** data surveys involve repeated surveys of the **same** individuals, households, firms, or governments at **different** points in time.

Longitudinal data are often used to examine individual life courses, enabling researchers to investigate the associations between different events or circumstances in a person life. For example, you would use a longitudinal dataset to study the effect of birth weight on later educational attainment.

One example of a longitudinal dataset is the British Cohort Study which is following the lives of the 17,000 babies born in the UK between 5th and 11th April 1970. The study talked to mother and midwife about the birth of each of these babies. At ages 5, 10 and 16, class teachers and parents were interviewed. Now the study-subjects are adults and they are interviewed every 3 or 4 years about their work, relationships, health, family, politics and values. The data in this study is all anonymised before being released to researchers. **ESDS Longitudinal** provides access and support to the major UK longitudinal studies.

Qualitative data is data that is non-numeric, for example, interview transcripts, focus groups, oral histories, case notes, records of meetings, diaries and mass observations. Difficult to digitalise, qualitative data often includes some diverse data types such as audio, video, photos and text. **ESDS Qualidata** provides access and support for a range of social science qualitative datasets.

Mesodata, are data that fall into an intermediate conceptual layer between aggregate and survey level databases. A mesodata database is typically a survey dataset that has been partly aggregated, or aggregated to very small geographic units, or a mixture of survey and aggregate data. It often retains the granularity required to show some of the relational or hierarchical structures of the original survey database.

1.3.3 Creating surveys

A survey typically consists of collecting data in a standardized form from a relatively large number of individuals over a short time frame. The design of a questionnaire can affect how people respond to questions and the response rate. A well-designed survey asks unambiguous questions that provide a valid measure of the research question. It is carried out on a large, representative sample and attains high response rates, which will enable researchers to generalise from the sample to the population with greater accuracy.

For international surveys it is also important that questions are translated carefully to produce functionally equivalent questionnaires, and that consistent methods of fieldwork and coding are used across the participant countries.

There are many methods used to conduct surveys, for example, by post, telephone or face to face interview. The method chosen depends on the resources and time available. The European Social Survey is an attitudinal survey asking quite personal questions about people's values and beliefs. For this reason, the ESS is, where possible, conducted through face to face interviews with each interview lasting about an hour. This allows the interviewer to clarify questions and the interviewee to expand their answers.

1.3.4. Creating aggregate databanks

Aggregate data are derived from micro-level survey datasets by applying statistical methods such as summing or averaging, then applying additional calculations such as weighting and estimation of sampling error. These procedures are designed to provide reliable inferences about an entire population based on data collected from the set of samples surveyed.



Example 3: Labour Force Surveys

Unemployment is one of the most closely watched measures of the health or otherwise of a nation's economy. But how is it calculated?

The standard measure of unemployment is the **unemployment rate** which measures the number people who are unemployed as a proportion of the number of people who are economically active. This unemployment rate is derived from survey data. In the UK, the Office for National Statistics uses the Labour Force Survey as a basis for calculating the unemployment rate. The Labour Force Survey is conducted every 3 months and interviews around 101,000 people over the age of 16. Each respondent is classified as either in employment, unemployed or economically inactive. The unemployment rate is then derived from these responses.

So, the unemployment rate is an example of how an aggregate indicator is produced from a survey. However, basing unemployment rate on a survey raises several methodological issues. Firstly, although the Labour Force Survey covers a large, representative sample, not 100% of the over-16 population is included in the survey. This means, a different sample would produce a slightly different result. The spread of results from different samples is known as the sampling variability. The sample variability places a confidence interval in the final unemployment rate produced by the survey. Secondly, as it is based on a survey the unemployment rate also needs to be adjusted to reflect the population as a whole, perhaps by adjusting the results for a particular local area by age group and gender. For example, the 2001 UK census revealed that the UK population was 1 million lower than had been previously estimated. This meant a new set of adjustment factors based on the revised population figures had to be applied to all the Labour Force Surveys from 1992 onwards.

A second measure of unemployment is claimant count. This is a headcount of everyone claiming Job Seekers Allowance and National Insurance credits. Since it is a 100% count, the claimant count is unaffected by sampling variability and so can be used as an indicator of those without work at very small levels of geography. However, as claimant count only measures

people claiming benefit, the unemployment rate is considered to be a more accurate measure of actual unemployment and is the standard indicator. In fact, it is a legal requirement for every country in the EU to produce a Labour Force Survey, and Eurostat and the OECD then use these surveys to publish a monthly rate of unemployment for each of their member countries.

1.4. International data resources

1.4.1. An overview

Inter-governmental organisations (IGOs) such as the International Monetary Fund and World Bank are the primary actors in the collection and dissemination of international aggregate data. These organisations have the capacity to produce very high quality multi-national databanks as they have a presence in every country in the world, the leverage to charge countries to deliver the data and the power to create international standards. National statistical agencies are the main source of data, and one of the functions of international agencies is to provide technical assistance and financial resources to national statistical agencies that are struggling to build their national statistical infrastructures. Data are also collected from central banks, government agencies and through specialist surveys.

Inter-governmental organisations have been central to the establishment and adoption of the common standards and definitions used by countries worldwide to report their data. For example, balance of payment data (which record of the flows of goods, services and finance between an economy and the rest of the world) is collected using a methodological framework developed by the IMF. The collection of standardised, comparable and timely balance of payment data is seen as a core task of the Fund, as one of its primary functions is to prevent financial crises and assist countries in balance of payment difficulties. As a consequence, the Fund produces the **Balance of Payments Manual**, used by countries worldwide to produce their balance of payments data in a common format. The manual was first published in 1948 and successive editions were published in 1950, 1961, 1977, and 1993 to reflect changing economic circumstances and analytic requirements. The Fund is now consulting on a new version of the manual that will clarify reporting on more recent issues such as monetary union (the euro-zone for example) or migratory workforces (with increasing globalization, more people have connections to two or more economies).

Some IGO's have open dissemination policies. The World Health Organization, for example, makes all the data it produces freely available online through its **Statistical Information System**. However, for most IGOs, the dissemination of data is limited and often carried out through subscription services. The UK is unique in that, through ESDS International, the UK academic community has free web-

based access to the major databanks produced by the IMF, World Bank and other major international governmental organisations. There is more about ESDS International in the next section.

1.4.2 ESDS International

ESDS International

ESDS International is a specialist data division of the Economic and Social Data Service (ESDS), the principle UK data service for the social sciences. It is funded by the ESRC and JISC. The service provides free, online access to the major databanks produced by the IMF, OECD, World Bank, UN, IEA and Eurostat.

How does it work? ESDS International has negotiated national license agreements with several international governmental organisations. Through these agreements, the UK academic community has free access to the full versions of the databanks covered by the licenses. The databanks are updated regularly with the latest releases. In many cases, this means monthly.

How long for? The current licenses expire in 2008. We are currently in talks with the ESRC and the data providers to extend the agreements until 2012.

Who can use the data? Under the terms of the licensing agreements we have with the data providers, you must be a member of staff or student to access the databanks through ESDS International. You can only use the data you access through the service for teaching or research purposes. When you first register to use the data you will need to agree to this and the other terms and conditions before you can access the databanks.

How is the data delivered? The data arrives at ESDS International in all sorts of different formats. We convert it all into the same format, Beyond 20/20, for web delivery. Beyond 20/20 is the name of a propriety software package developed in Canada for the web delivery of aggregate databanks. The OECD and IEA use it for their own in-house web delivery systems. The other databanks in the portfolio are converted into Beyond 20/20 by ESDS International.

We will now look in more details at the content of the databases in the ESDS International portfolio.

1.4.2.1 World Bank databanks

Created along with the IMF in 1944 to rebuild Europe, the main focus of the World Bank is to alleviate poverty and improve the living standards of people in the developing world. It is a development bank which provides loans, policy advice and technical assistance to countries and it collects data on all aspects of human development worldwide.

The World Bank's primary database, the **World Development Indicators** (WDI) is one of the most widely used and frequently cited of the international databanks. The database contains statistical data for over **630 development indicators**. The extensive collection of development data includes social, economic, financial, natural resources and environmental indicators covering the period 1960 - onwards for over **200 countries** and 18 country groups.

The main topics in the WDI are as follows:

People

- Population and demographics
- Labor and employment
- Poverty and income distribution
- Education
- Health

Environment

- Land use and agricultural production
- Energy production and use
- Urbanization
- Emissions

Economy

National accounts (local currency)
 National accounts (US\$)
 Derived national accounts
 Purchasing power parity
 Trade
 Government finance
 Monetary
 Balance of payments
 External debt

States & Markets

Investment and risk
 Financial depth
 Tax and trade policies
 Prices and exchange rates
 Military expenditures and arms trade
 Transport, power, and communications
 Information and technology

Global Links

Investment and trade
 Financial flows
 Development assistance and aid
 Foreign labour and population in OECD countries
 Travel and tourism

The database is designed to be used for making cross-national comparisons in that the same indicators tend to be reported in the same units for every country. Series that are reported in local currencies almost invariably have a equivalent series in a more comparable unit, such as \$US or % of GDP. All the data in the WDI are annual data. The database is updated once a year.

**Example 4 : WDI - Global Urbanisation**

This graph, based on the World Bank World Development Indicators shows the how the percentage of population that live in a cities has changed over time for selected Asian countries. From the graph, you can see that in the city states of Hong Kong and Macao, almost everyone lives in a city. The industrialising countries are also urbanising at a rapid rate.

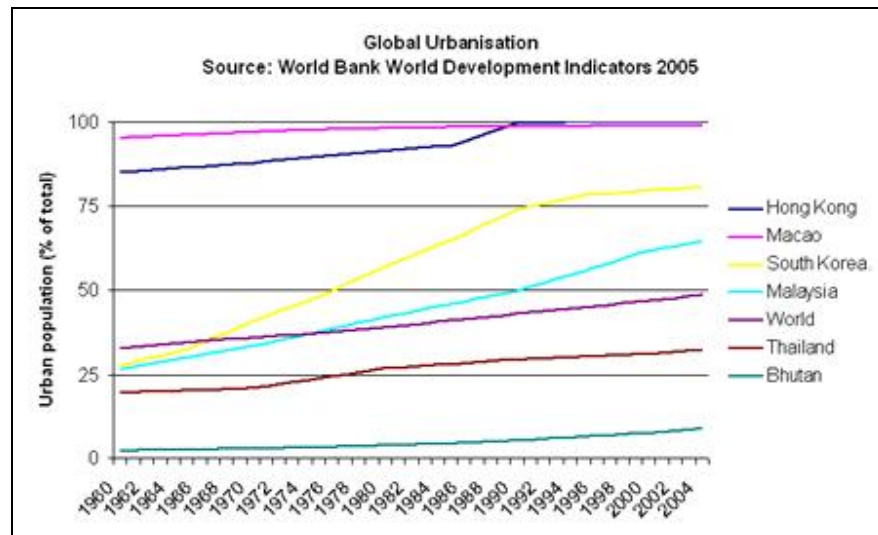


Figure 6: WDI graph showing global urbanisation in selected Asian countries

The process of urbanisation tends to be most advanced in the higher income countries, where almost 80% of the population already live in metropolitan areas. This figure drops to 50% for middle income countries. In low income countries the majority of people still live in the countryside. However, these regions are showing the fastest rate of urban growth and it is projected almost all the world's total population growth between 2000 and 2030 will take place in the cities of these countries.

The graph also includes the data for the world as a whole. According to the UN, by 2007, for the first time in human history, more than half the people in the world will be living in cities. As the chart shows, this is the result of the result of a continuing movement of people that has lead to a tremendous growth in urban areas around the world.

The Global Development Finance database is also produced annually by the World Bank. This database contains more than **200 debt and financial flows indicators** for over **130 countries** that report public and publicly-guaranteed debt to the World Bank Debtor Reporting System. Topics covered include external debt stocks and flows, major economic aggregates, and key debt ratios as well as average terms of new commitments, currency composition of long-term debt and debt restructuring. The database runs from 1970 onwards and also contains future projections of countries' debt repayments for the next ten years.



Example 5 : Workers remittances

This example from the World Bank Global Development Finance shows workers remittances, or the money people send home when they work abroad. Remittances tend to be counter-cyclical, buffering other shocks, since natural disasters or economic downturns encourage additional workers to migrate abroad and those already abroad increase the amount of money they send to families left behind. Remittances now exceed official development assistance. They are set to rise further as banking restrictions are lifted and it becomes easier to send money home. You can see India has high level of remittances due to its highly educated and historically very mobile workforce. Remittances are also very high for Mexico because of its proximity to the USA.

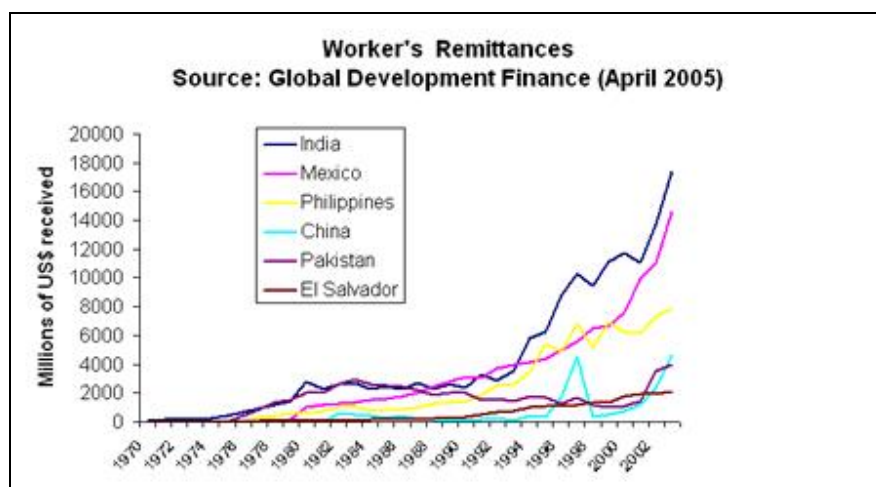


Figure 7 : example from the World Bank Global Development Finance showing workers' remittances

1.4.2.2 IMF databanks

The International Monetary Fund or IMF is the **central institution** of the international monetary system (the international system of exchange rates and payments between countries). The primary purpose of the Fund is to maintain international **financial** stability and prevent crisis in the system. It does this through the provision of advice and financial assistance to countries that experience serious balance of payment difficulties. The Fund is often criticized for the terms and conditions it attaches to these loans, which frequently include internal changes (such

as cutting social expenditures, privatisation and anti-corruption measures) as well as external ones, in particular, the reduction of trade barriers.

The Fund **monitors** economic and financial developments and policies and collects detailed macro-economic data on all its member countries.

It produces four major databanks:

- International Financial Statistics
- Government Finance Statistics
- Balance of Payments Statistics
- Direction of Trade

These four databanks contain data on national accounts, trade, the balance of payments and government spending for around 190 countries. Collectively, they provide a global picture of economic development and international trade over the last 50 years.

The **International Financial Statistics(IFS)** is the IMF's principal statistical publication and is the standard source for all aspects of international and domestic finance. It was first published in 1948 and is updated every month. It covers over 200 countries and reports time series data on exchange rates, balance of payments, international liquidity, money and banking, interest rates, prices, production, international transactions, government accounts, national accounts and population. It also contains prices of internationally traded commodities such as coffee, oil and gold.

The IFS is split in three parts: country tables, world tables are commodity prices. Data are reported in a mixture of indexes, national currencies, US Dollars, SDRs (the basket currency the Fund uses as a unit of account) and other units. As the database is primarily used for countries to benchmark their own progress, you need to take extra care if using the IFS to make cross national comparisons. For example, countries may use different base years to reporting indexes and many of the series in the country tables are in local currencies. The IMF pulls the more comparable series together to form the world tables.

The Balance of Payments Statistics contains time series data covering the standard balance of payment components and international investment positions of countries worldwide. Balance of payment data record of the flows of goods, services and finance between an economy and the rest of the world. Since one of the primary functions of the IMF is to prevent financial crises and assist countries in balance of payment difficulties, the collection of standardised, comparable balance of payment data is seen as a core task.

A country's Balance of Payments (BOP) is a statistical statement that summarizes, for a specific period (typically a year or quarter), the economic transactions of an economy with the rest of the world. It covers:

- all the goods, services, factor income and current transfers an economy receives from or provides to the rest of the world
- capital transfers and changes in an economy's external financial claims and liabilities

The Balance of Payments methodology uses a double-entry accounting system. This means that every recorded item should have a debit and a credit, and there should be a net balance of zero. In practice, the figures rarely balance to the point where they cancel each other out. This is the result of errors or omissions in the compilation of statements.

Balance of Payments credits denote a reduction in assets or an increase in liabilities; debits denote an increase in assets or a reduction of liabilities, summarized in the table below:

Balance of Payments credit and debit table	
Credit	Debit
Exports of goods and services	Imports of goods and services
Income receivable from abroad	Income payable abroad
Transfers from abroad	Transfers to abroad
Increases in external liabilities	Decreases in external liabilities
Decreases in external assets	Increases in external assets

The **Government Finance Statistics** database was introduced by the IMF to provide current and internationally comparable data on the finances and fiscal

policies of Fund member governments. The database contains information on government income and expenditure i.e. how governments raise revenue (tax, lending, bonds, etc), and how they spend it (on defence, education, health, financing debt repayments etc) for all levels of government (national, state, local).

In April 2004, the IMF introduced a new classification system for the Government Finance Statistics that only supports data from 1990 onwards. The IMF now split the GFS into 2 databases: - the **Government Finance Statistics Current (1990 -)** contains data using the new classification system from 1990 onwards, and the **Government Finance Statistics Historical (1972 - 2003)** contains data using the old classification system.

The **Direction of Trade Statistics** database contains data on the value of merchandise exports and imports between each country and all its trading partners. Total bilateral and multilateral exports and imports are aggregated at national or regional group level. For each reporting country or group, all the trading partners are listed. The corresponding monetary values of total imports and total exports are then provided as time series for each country / trading partner pair. Along with data from reporting countries, estimates are provided by partner countries for non-reporting countries or for those that are slow to report.

All the data in the Direction of Trade are expressed in \$US. Data originally reported in national currencies are converted by the IMF using the exchange rates published in the IMF International Financial Statistics.



Example 6 : IMF Direction of Trade

Iraq's trade with the rest of the world reflects the country's recent volatile history. The costly and inconclusive war with Iran depressed trade throughout the 1980's. After hostilities ended in 1988, imports and exports recovered briefly, but then fell sharply after the invasion of Kuwait and the imposition of sanctions following the Gulf war. The data shows trade started to increase after the implementation of the UN's oil-for-food program beginning in December 1996, under which Iraq was allowed to export limited amounts of oil in exchange for food, medicine and other basics.

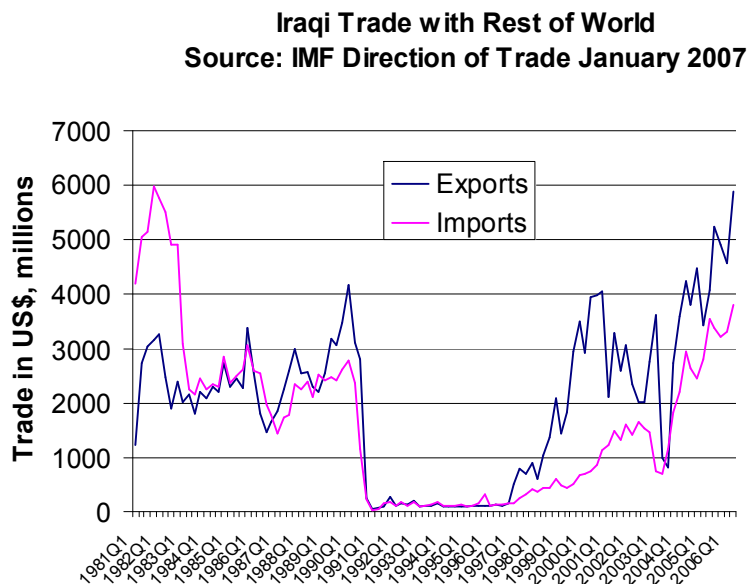


Figure 8 : IMF direction of trade for Iraq

Oil is by far Iraq's major export. We can see the drop in exports between the first and second quarters of 2003 following the US-led invasion in March when the pipelines were blown up and then the rise again as the pipelines were fixed. These data are from the Jan 2007 release of the International Monetary Fund's Direction of Trade, which contains data running up to the middle of 2005. The database is updated every month.

1.4.2.3 OECD databanks

The OECD (Organisation for Economic Cooperation and Development) is a group of the world's thirty biggest market economies all market economies. Collectively its member countries account for about two thirds of the world's GDP. The OECD acts as a forum for member countries to develop economic and social policies and it collects and disseminates economic data on a wide range of industrial and economic indicators.

The OECD data are considered to be accurate and reliable and provide an authoritative means to compare economic, social and industrial indicators across national boundaries. However, the databases tend to only cover a limited number of countries.

Access to the OECD databases is usually by institutional subscription. ESDS International provides free access to UK academics to most of the OECD databases.

Databases produced by the OECD:

- Agriculture
- Banking Statistics (formerly Bank Profitability)
- Economic Outlook: Statistics and Projections
- Education Statistics (formerly Education at a Glance)
- Employment and Labour Market Statistics
- Globalisation
- ITCS International Trade by Commodity
- Indicators of Industry and Services
- Institutional Investors
- Insurance
- International Development
- International Direct Investment
- International Migration
- International Trade and Competitiveness
- Main Economic Indicators
- Monthly International Trade
- National Accounts
- OECD Health Data
- Revenue
- STAN Structural Analysis
- Science, Technology and R & D
- Services
- Social Expenditure
- Statistical Compendium
- Structural and Demographic Business Statistics (formerly Structural Statistics for Industry and Services)
- Taxing Wages
- Telecommunications and Internet

1.4.2.4 IEA databanks

Founded during the oil crisis of 1973-74, the International Energy Agency (IEA) has a very similar membership and structure as the OECD. However, its primary focus is energy and environmental policy. It produces data on energy security, production, trade, stocks, transformation, consumption, prices, taxes and environmental sustainability. Designed for cross country comparability, the IEA databases are global databanks covering around 130 countries. The majority contain annual time series data from 1960 onwards.

Access to the IEA databases is usually by (very expensive) individual subscription.

Databases produce by the IEA:

- Energy Statistics of OECD Countries
- Energy Balances of OECD Countries
- Energy Statistics of Non-OECD Countries
- Energy Balances of Non-OECD Countries
- CO2 Emissions from Fuel Combustion
- Energy Prices and Taxes
- Energy Technology Research and Development Database
- Electricity information
- Coal Information
- Natural Gas Information
- Oil Information
- Renewables Information

ESDS International provides free access to UK academics to the IEA databases.



Example 7 : From the IEA

This example from the IEA's CO₂ Emissions from Fuel Combustion shows how emissions from electricity generation have changed over time in Latin America. The most striking feature is that CO₂ emissions have fallen dramatically since the 1970's and are comparatively low, even for big economies like Brazil. This is because, thanks to the terrain, a very large percentage of electricity in South America is generated from hydroelectric power.

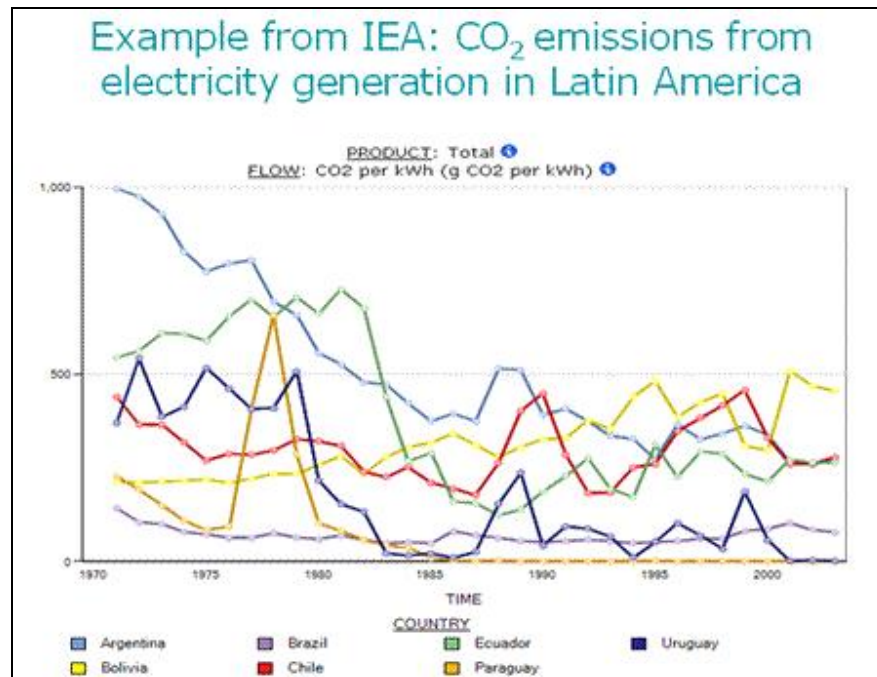


Figure 9 : IEA graph based on CO2 emissions from electricity in Latin America

1.4.2.5 UN databanks

The UN produces a large number of databases. Many are freely available. ESDS International hosts the important subscription databases produced by the UN and its agencies. The **UN Common Database** has been used for many years internally by the UN as a basis for policy formation and covers a huge range of human development indicators such as employment, childhood, health, access to clean water access to the internet and land use. It has a wider range of indicators than the World Bank development indicators but it is a little bit patchy with shorter time series.

The United Nations Industrial Development Organisation (UNIDO) produces the **Industrial Statistics Databases** and **Demand Supply Balance Database**.

These annual databanks cover employment and trade world-wide, broken down by country and manufacturing sector. They provide data that can be used to analyse patterns of growth, structural change, industrial performance and employment by gender and sector worldwide.

The International Labour Organisation, the oldest of the UN agencies, collects data on the labour market and working conditions worldwide. Its principal publication is the

Key Indicators of the Labour Market, a frequently cited database on labour market statistics worldwide.



Example 8 : From ILO Key Indicators

Some typical data from the ILO: the average annual hours worked around the world. You can see workers in Europe do the least!

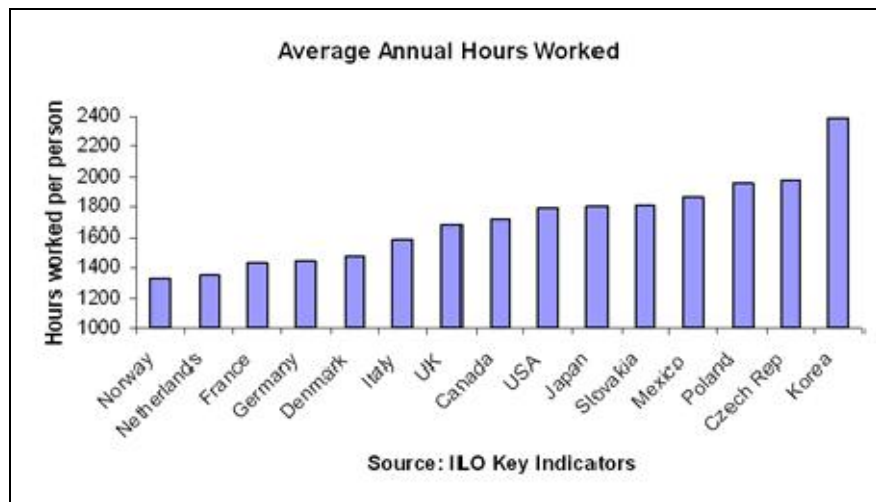


Figure 10 : ILO graph showing European workers' annual hours

1.4.2.6 Eurostat

Eurostat is the statistical arm of the European Union. It produces detailed data on the member states of the EU. Eurostat does not collect data itself (this is done by the statistical authorities of member states), instead its role is to consolidate the data, ensure methodologies are harmonized and provide the European Union with the data required to formulate policies and make comparisons between countries and regions. In particular, Eurostat produces the regional data which guide the EU's structural policies, and the macro-economic data used by the European Central Bank in the development of its monetary policy for the euro.

The main database produced by Eurostat is New Cronos, which contains detailed data on a wide range of social and economic themes at a national and regional level for the EU. Many of its tables also cover candidate member countries, central European countries, and the main economic partners of the European Union. The data may be daily, monthly, quarterly, half-year or yearly depending on the statistical field covered. Eurostat New Cronos is the only major pan-European

database containing data at a regional level for the EU. Eurostat New Cronos is now freely available from the **Eurostat website** and ESDS International also host Eurostat data with an alternative, more user-friendly interface.

1.4.2.7 Freely available databanks

The primary focus of ESDS International is to provide free access to databases that have a subscription charge. In this section, we highlight selected sources of freely available data.

The IGO's are all publicly funded organizations and many of them make their data freely available. In particular the **World Health Organisation** makes all the data it produced freely available. Data on food and agriculture are available through the **FOA website** and the **Bank for International Settlements** (the world's central bank) disseminates global finance statistics. The World Bank, IMF and UN all make small subsets of recent data freely available from their websites. ESDS International has created a **Guide to freely available international data resources**. This guide also contains links to national statistical office worldwide, and resources on methodology, standards and good practice in the compilation of international data.



Activity 1 : Data collection - Beyond 20/20

You will need to go online to complete this activity – printed step by step instructions can found at the back of this workbook. This task is to use the World Bank's annual World Development Indicators (WDI) to visualise and download male and female life expectancy figures in 5 year intervals for the 1980's and 1990's for various regions of the world. You will need your Athens username and password to complete this task.

1.5. Micro data

1.5.1. An overview of the available micro data sets

Almost every country produces its own household surveys. The International Household Survey Network maintains a web-based **Central Survey and Census Catalogue** that holds the metadata created by these surveys. Although ownership and control of access to datasets remains with each participating depositor, the catalogue enables you to see a short abstract and the questionnaire used in each survey (and sometimes the data).

The World Bank has also created a Microdata Management Toolkit to help national statistical offices organise their census and survey data . Survey producers can feed their data files from standard formats (SAS, SPSS, STATA, or others) into the Toolkit alongside the documentation and related metadata (literal questions, sampling methodology, period and method of data collection, information on producers, description of computed, imputed and recoded variables and so on) . The Toolkit checks that all variables and values are properly labelled and the metadata is in order, and then creates a Nesstar datafile. All original datafiles and the related metadata are stored in one single file coded up to international standards, which makes the archiving, cataloguing and dissemination of datasets considerably easier.

Aside from the international databanks produced by intergovernmental organisations, many smaller organizations produce cross-national survey datasets. These typically are academically-driven surveys that compare opinions, life and individual experiences in several countries; examples include the European Social Survey and the World Values survey. ESDS International provides access to these survey datasets alongside the macro-databanks through a series of reciprocal arrangements with the host institution or data archive. In the rest of this section we are going to have a brief look at the most important of these datasets.

1.5.1.1 European Social Survey

The **European Social Survey** (the ESS) examines the interaction between Europe's changing institutions and the attitudes, beliefs and behaviour patterns of its diverse populations. Now in its third round, the survey covers over 20 nations.

The ESS questionnaire includes two main sections, each consisting of approximately 120 questions or topics; a 'core' module which will remain relatively constant from round to round, plus two or more 'rotating' modules, repeated at intervals. The core module aims to monitor change and continuity in a wide range of social variables, including media use, social and public trust; political interest and participation; socio-political orientations, governance and efficacy; moral, political and social values; social exclusion, national, ethnic and religious allegiances; well-being, health and security; demographics and socio-economics.

The ESS describes itself as a "*methodologically bullet-proof study of changing social attitudes and values*", and through its sampling procedures and translation protocols goes to great lengths to ensure that the data is comparable between countries.

1.5.1.2 World Values Survey

The **World Values Survey** is academically-driven social survey originally developed to answer the question, "What makes people happy?" It investigates socio-cultural, moral, religious and political values of different cultures around the world and looks for the relationships between people's values and beliefs and their overall levels of well-being. First run in 1981, the survey is currently takes place around every 5 years, the most recent wave being 2005-6. The questionnaires from the most recent waves have consisted of about 250 questions. Now covering around 60 countries, the questionnaires are administered a worldwide total of about 92000 interviewees.



Example 9 : World Values Survey

In this example from the World Values Survey we find a classic combining micro and macro data finding. The plot shows how level of happiness vary with a measure of a country's income per person. The way this graph is usually interpreted is that at low levels of income, a small increase in income results in a large increase in happiness. At higher levels of income, changes in average wealth have a smaller effect on how satisfied people feel about their lives.

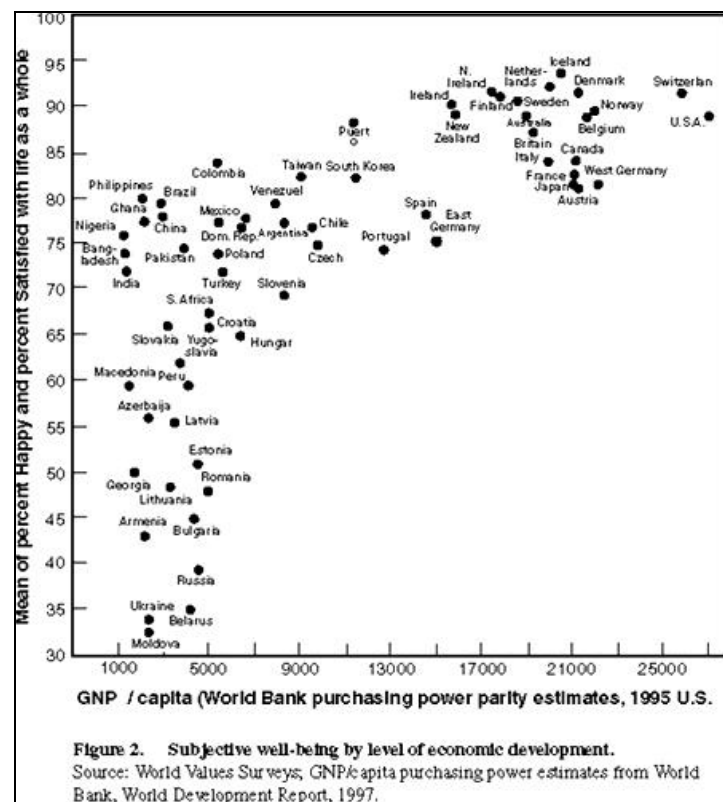


Figure 11 : This plot shows how level of happiness vary with a measure of a country's income per person

1.5.1.3 The global barometers

Eurobarometer is a series of surveys cross-national and cross-temporal comparative social research conducted on behalf of the European Commission and designed to monitor social and political attitudes. Since the early seventies nationally representative samples in all European Union member countries are interviewed at least twice a year. Separate supplementary surveys on special issues, such as attitudes towards European Union enlargement, have been conducted in nearly every wave.

Afrobarometer measures social and political attitudes in Africa. In common with the other barometers, it examines issues of trust in institutions and society, evaluation of the country's political and economic performance, participation and spiritual beliefs. It has a particular focus on how people feel about their living conditions. In the most recent round, Round 3, surveys were conducted in 18 countries from March 2005 through February 2006.

Latinobarómetro is an annual public opinion survey of approximately 19,000 interviews in 18 countries in Latin America. The survey started in 1988 in Argentina, Brazil, Uruguay and Chile and focused on the social underpinnings of these new democracies. As with the other barometers, Latinobarómetro surveys opinions, attitudes, behaviors, and values of the people of the region on such issues as the economy, politics, culture, social policies and the distribution of wealth. Latinobarómetro has now been acquired by ESDS International for the UK academic community.

1.5.1.4 Other international surveys

ESDS International also provides access to a number of other cross-national surveys including:

- **E-Living : Life in a Digital Europe**
- **European Election Study**
- **International Social Survey Programme**

1.5.2 Nesstar

Nesstar is a web-based data delivery software system that allows users to search for, locate, browse, analyse and download data. It differs from Beyond 20/20 in that it is specifically designed to deliver survey data. Data delivered using Nesstar include many large-scale survey datasets, such as the UK Labour Force Survey and the General Household Survey, and international survey datasets such as the European Social Survey.

ESDS produces an introductory guide to using Nesstar at

www.esds.ac.uk/support/guides/A2.pdf

1.6. Further exploration



Activity 2 : Locating Data Using the ESDS catalogue

You will need to go online to complete this activity – printed step by step instructions can be found in the **Activities** section at the back of this workbook.



Activity 3 : Locating Data Using the ESDS International dataset user guides

You will need to go online to complete this activity – printed step by step instructions can found in the **Activities** section at the back of this workbook.



Activity 4 : looking at sources

You will need to go online to complete this activity – printed step by step instructions can found in the **Activities** section at the back of this workbook. You will need your Athens username and password to complete this task.

1.7. Suggested Reading

The ESDS International website brings together information on all the major international databases discussed in this unit into a single resource:

<http://www.esds.ac.uk/international>

For each database the website provides an online dataset guide containing information on the topics and countries covered, periodicity and time range of the database. In addition, dataset specific step-by-step guides to accessing international data via Beyond 20/20 WDS are available. The website also provides themed data guides, user guides relating to methodological issues, links to data provider documentation and a structured guide to freely-available international data.

The OECD Factbook 2007 is an online resource that provides data and commentary on world economic, social and environmental trends.

http://www.oecd.org/document/62/0,2340,en_21571361_34374092_34420734_1_1_1_1,00.html

More general introductory books on the statistical concepts in this unit include:

Wright, D. *Understanding Statistics: An Introduction for the Social Sciences*. Sage, 1997

Argyrous, G. *Statistics for Research*, 2nd Edition, Sage, 2005.

Czaja, R. and Blair, J. *Designing Surveys: A Guide to Decisions and Procedures*, 2nd Edition, Sage, 2004.

Robson, C. *Real World Research: A Resource for Social Scientists and Practitioner-researchers*, Blackwell, 2002

Koop, G. *Analysis of Economic Data* John Wiley and Sons Ltd, 2004



Activity 1 : Data collection - Beyond 20/20

This task is to use the World Bank's annual World Development Indicators (WDI) to visualise and download male and female life expectancy figures in 5 year intervals for the 1980's and 1990's for various regions of the world. You will need your Athens username and password to complete this task.

Life expectancy is the average number of years newborn babies can be expected to live based on current health conditions. It reflects environmental conditions in a country, the health of its people, the quality of care they receive when they are sick, and their living conditions.

Visualising and downloading male and female Life expectancy figures in five year intervals for six regional groups

Stage i) Finding and selecting the data

Open Internet Explorer and go to www.esds.ac.uk/international. Click on **Access Data** then select the **World Bank World Development Indicators** from the list of databanks to see the Summary Table. Open the database using the [**** Access World Bank data via Beyond 20/20 WDS ****](#) link in the Table.

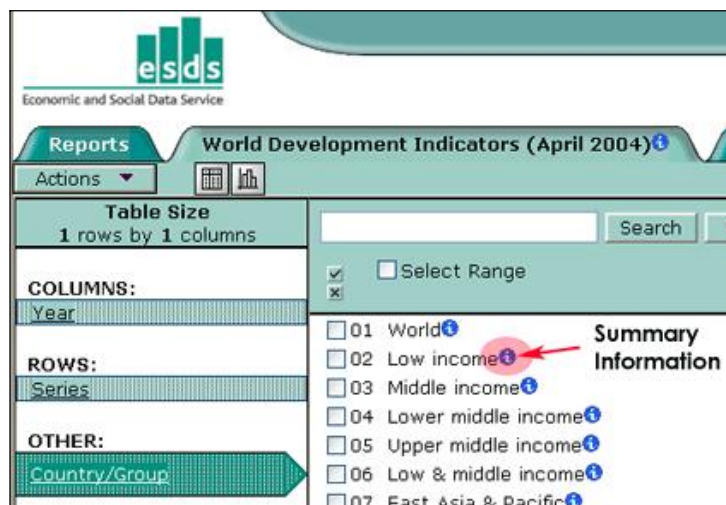
If you have opened a new browser session, the Athens Authentication Point window should appear prompting you to input your Athens user details. If you have not already registered with the ESDS you will then be automatically redirected to complete an online registration the first time you access the service, and to the terms and conditions page for each family of databases.

On entering the Beyond 20/20 Web Data Server site for the World Bank data you should see a folder labeled World Bank. It contains both of the World Bank databases hosted by ESDS International: the World Development Indicators and Global Development Finance.



Figure 12 : Beyond 20/20 screenshot 1

Select World Development Indicators to open the database. At the top of the screen you will see three tabs:



- Reports: returns to list of World Bank datasets.
- World Development Indicators, <month of release>: identifies the current report
- Help: Beyond 20/20 help

In the World Bank World Development Indicators, the data is broken down by series, country and year. These are called dimensions in Beyond 20/20. On the left hand side of the screen you will see the available dimensions under Columns, Rows and Other headings. When you open the database you are first taken to the Country/Group dimension view where you can select the geographical area of interest.

The current dimension is highlighted in darker green, for example in the screen shot above the Country/Group dimension is highlighted and a list of the countries and groups available for selection are shown on the right hand side of the page.

A small blue 'i' icon indicates summary information about a table or dimension item is available. For example if you click on the icon next to **08 Europe & Central Asia**, the following regional notes will be displayed:

Definition: Europe and Central Asia regional aggregate (does not include high-income economies). The economies included are: Albania. Armenia. Azerbaijan. Belarus. Bosnia and Herzegovina. Bulgaria. Croatia. Czech Republic. Estonia. Georgia. Hungary. Kazakhstan. Kyrgyz Republic. Latvia. Lithuania. Macedonia, FYR. Moldova. Poland. Romania. Russian Federation. Serbia and Montenegro. Slovak Republic. Tajikistan. Turkey. Turkmenistan. Ukraine. Uzbekistan.

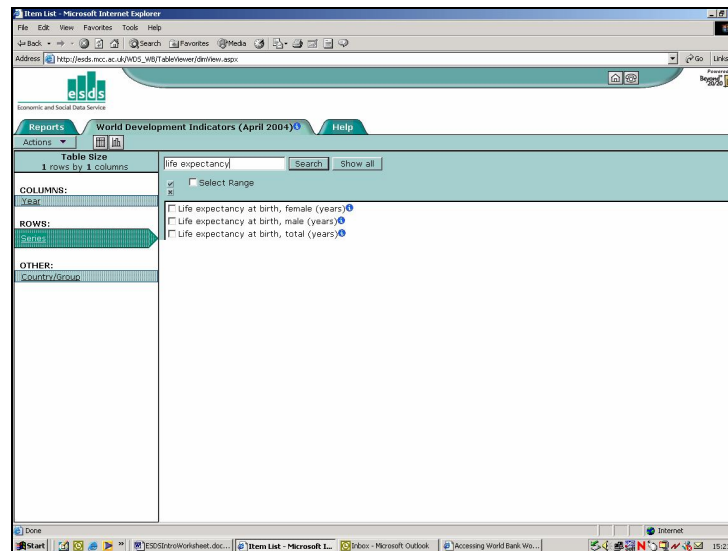
We will now select the desired country dimension items by placing a tick in the check box next to the region name. Scroll down and select the following Country/Groups by ticking checkboxes next to the following regions:

- **East Asia and Pacific**
- **Europe and Central Asia**
- **Latin America and Caribbean**
- **Middle East and North Africa**
- **South Asia**
- **Sub-Saharan Africa**

Once all six regions have been ticked, move onto next dimension, Series, using the Actions option from the toolbar. Click on **Actions > Select items to view > Series**.

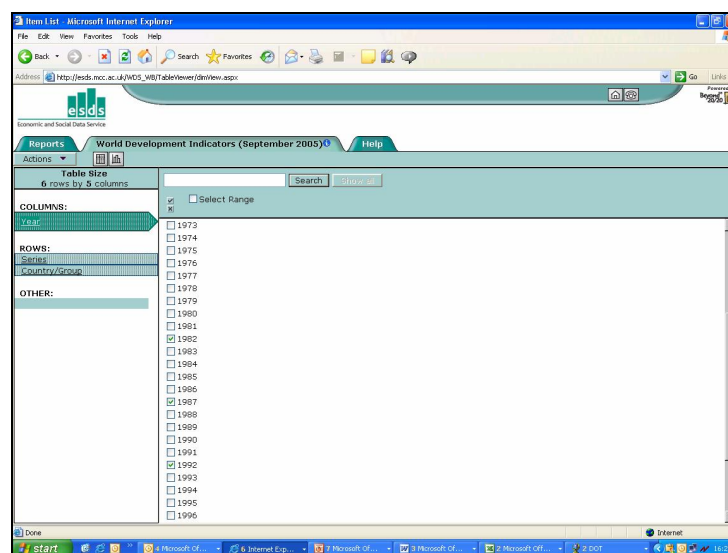
Note: You can also switch dimensions by clicking on the dimension name - both options are illustrated in the screen shot below.

A list of the series in the World Development Indicators appears. Type **life expectancy** into the search input box and click **Search**. The search finds three series- life expectancy for females, males, and a total.



Use the tick boxes to **select** all three series, i.e., Life expectancy at birth, female, male and total.

Use **Actions > Select items** to view to move onto next dimension, **Year**. Scroll down the year items and tick 1982, 1987, 1992, 1997, and 2002. Data on life expectancy is reported every 5 years.



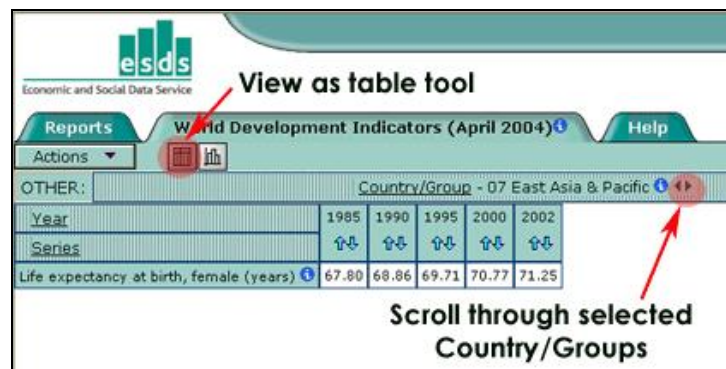
You have now completed the Finding and Selecting data stage and can move on to visualising your data.

Stage ii) Visualising the data

There are two ways of visualising data in Beyond 20/20 - in chart or table format.

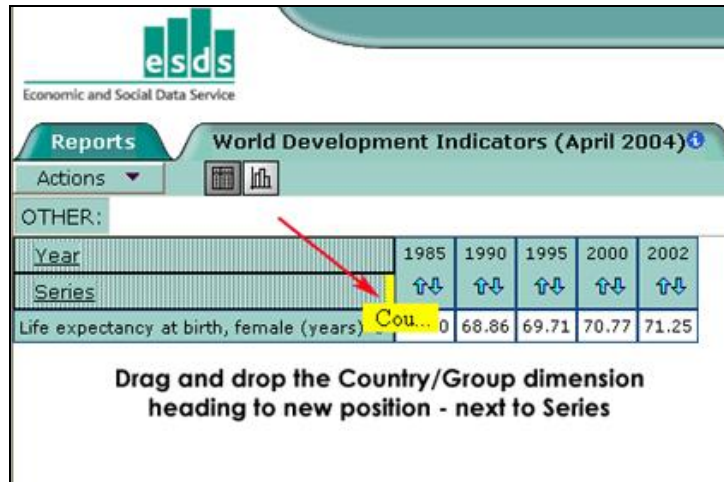
Table format

View the report in table format by clicking on the '**View as Table**' tool at the top of the screen.



In the default view of the table which appears, only the data from one Country/Group is shown - you can use the arrows highlighted in the screen shot above to scroll through the selected regions.

Change the format of the table by dragging and dropping the Country/Group dimension heading so that it is displayed next to the Series heading. To do this, hover over the hatched area of the Country/Group tile until the cursor turns into a four pointed arrow (↔), then left click, turning the heading yellow and use your left mouse button to drag and drop it into the right hand side of the tile marked Series.



esds
Economic and Social Data Service

Reports World Development Indicators (April 2004)

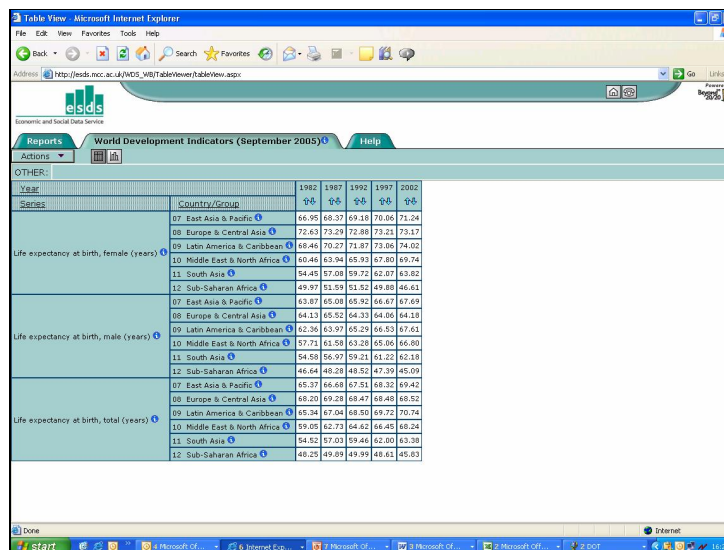
Actions

OTHER:

Year	1985	1990	1995	2000	2002
Series	↑↓	↑↓	↑↓	↑↓	↑↓
Life expectancy at birth, female (years)	Cou... 0	68.86	69.71	70.77	71.25

Drag and drop the Country/Group dimension heading to new position - next to Series

The table reorders as below:



esds
Economic and Social Data Service

Reports World Development Indicators (September 2005)

Actions

OTHER:

Year	Country/Group	1982	1987	1992	1997	2002
Series		↑↓	↑↓	↑↓	↑↓	↑↓
Life expectancy at birth, female (years)	07 East Asia & Pacific	66.95	68.37	69.18	70.06	71.24
	08 Europe & Central Asia	72.43	73.29	72.86	73.24	73.17
	09 Latin America & Caribbean	68.46	70.27	71.87	73.06	74.02
	10 Middle East & North Africa	60.46	63.94	65.93	67.80	69.74
	11 South Asia	54.45	57.08	59.72	62.07	63.82
	12 Sub-Saharan Africa	49.97	51.59	51.52	49.88	46.61
Life expectancy at birth, male (years)	07 East Asia & Pacific	63.07	65.08	65.92	66.67	67.69
	08 Europe & Central Asia	64.13	65.52	64.33	64.06	64.18
	09 Latin America & Caribbean	62.36	63.97	65.29	66.53	67.61
	10 Middle East & North Africa	57.71	61.58	63.28	65.06	66.80
	11 South Asia	54.58	56.97	59.21	61.22	62.18
	12 Sub-Saharan Africa	46.64	48.28	48.52	47.39	45.09
Life expectancy at birth, total (years)	07 East Asia & Pacific	65.37	66.68	67.51	68.32	69.42
	08 Europe & Central Asia	68.20	69.28	68.47	68.48	68.52
	09 Latin America & Caribbean	65.34	67.04	68.50	69.72	70.74
	10 Middle East & North Africa	59.05	62.73	64.62	66.45	68.24
	11 South Asia	54.52	57.03	59.46	62.00	63.98
	12 Sub-Saharan Africa	48.25	49.89	49.99	48.01	45.63

Use the table to answer the following questions:

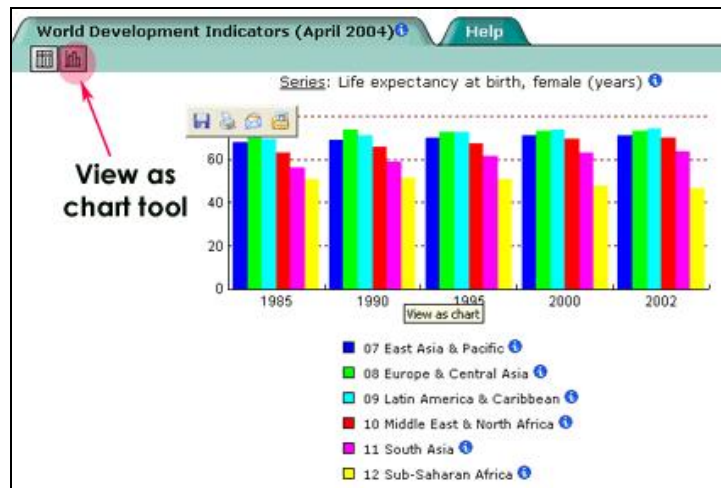
What has happened to life expectancy in Sub-Saharan Africa in the last ten years?

The male life expectancy in Europe and Central Asia has also fallen since 1987, although for different reasons, leaving Europe with the biggest difference in life expectancy between men and women.

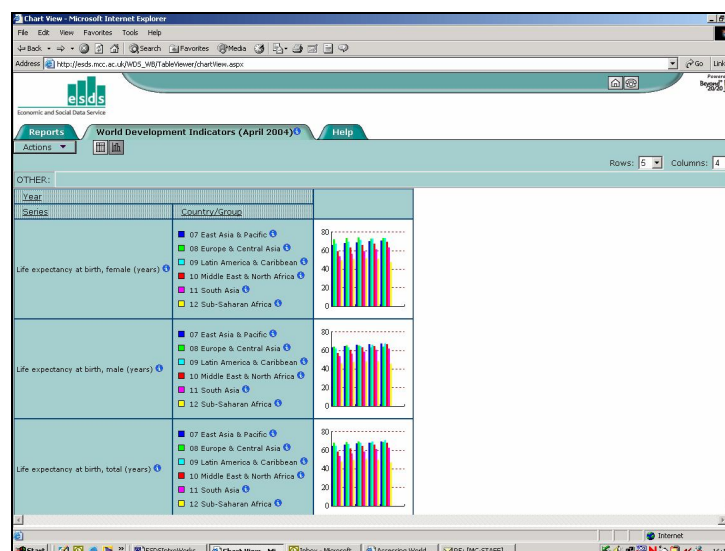
Why do you think this has happened?

Chart format

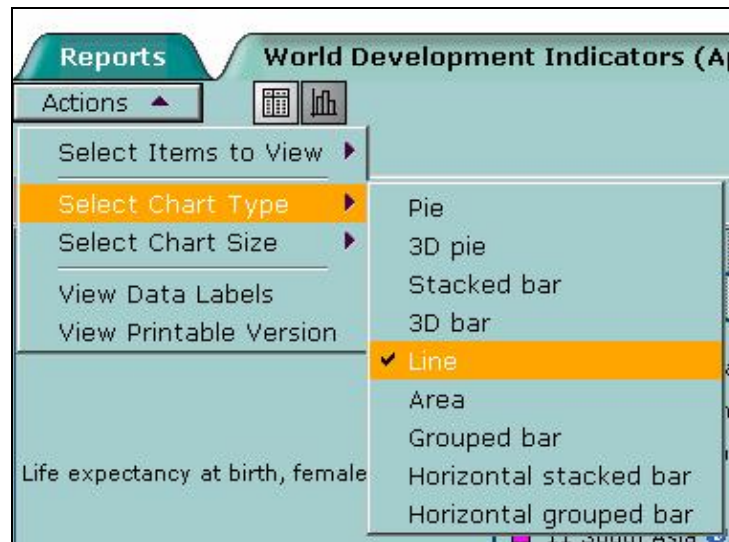
To now view the report as a chart, simply click on the **'View as Chart'** tool at the top of the screen.



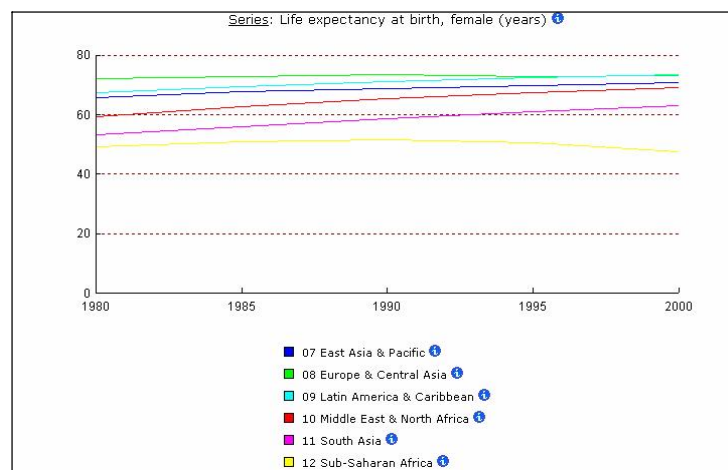
The default chart type is a bar chart:



Use the Actions menu to change chart type, chart size and to view a printable version of the chart.



Click on a chart image to enlarge it - in the below example the Female Life Expectancy at birth chart has been enlarged. The chart type has also been altered to display the data in a Line chart using the Chart Type drop down menu.



Answers to questions:

What has happened to life expectancy in Sub-Saharan Africa in the last ten years?

Life expectancy in sub-Saharan Africa has fallen over the last 10 years. The spread of HIV/AIDS across sub-Saharan Africa has resulted in a sharp fall in life expectancy, setting back the progress made over decades in

improvements in living standards. In the Botswana, for example, where one out of every three adults is infected with HIV, life expectancy at birth decreased from 65 to 37 years between 1987 and 2002 after it had been rising steadily for more than thirty years.

The male life expectancy in Europe and Central Asia has also fallen since 1987, although for different reasons, leaving Europe with the biggest difference in life expectancy between men and women.

Why do you think this has happened?

The fall in male life expectancy in Europe as a whole is due to falling life expectancy in the former Soviet republics since the break-up of the USSR. In Russia, between 1987 and 2002, male life expectancy fell by six years and that of women by 2 years. During this period, older men suffered more heart attacks and strokes, while younger men were involved in more accidents, suicides, substance abuse, and murder. These changes occurred while Russia was undergoing its transition to a market economy, and many people experienced severe economic hardship combined with deterioration in law, order and medical services.



Activity 2 : Locating Data Using the ESDS catalogue

In this task, you are going to search for data on poverty in Zambia. ESDS International has a search catalogue page at:

<http://www.esds.ac.uk/international/access/datasearch.asp> you can also reach this page by going to the **Data** tab then **Search International Datasets** link.

The screenshot shows the ESDS International website. The header includes a navigation bar with 'About', 'Data', 'Support', 'News', and 'ESDS International'. The 'Data' tab is selected. On the left, there is a sidebar with links: 'Introduction to data access', 'Search international datasets', 'New data releases', 'Macro data', 'Micro data', 'International data guide', and 'Visualisation interface'. The main content area is titled 'Search and browse international data'. It has a 'Search' section with a text input field containing 'zambia' and a 'GO' button. Below this is a 'Browse' section with radio buttons for 'Macro', 'Micro', and 'Both' (selected), and a 'Sort by' dropdown menu with a 'GO' button.

A search for 'Zambia' brings up all the databases that include Zambia as a country covered or in the description of the database. The default search will search both the macro (aggregate) and micro (survey) databases.

Searching tips:

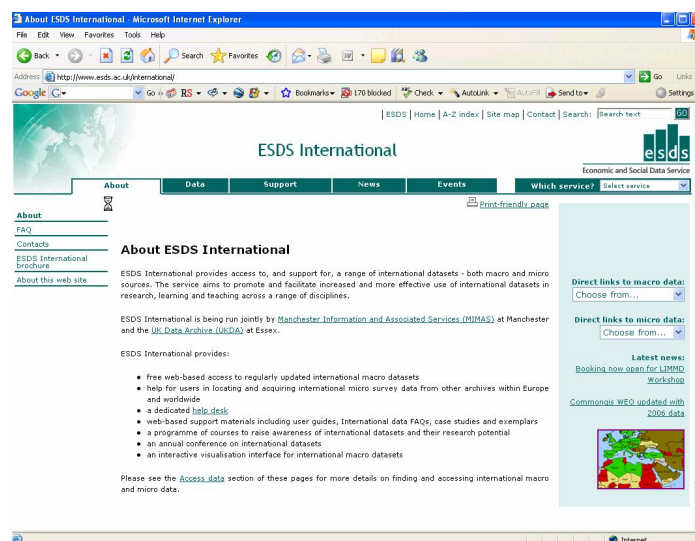
- use double quotes for an exact phrase search, e.g. "direction of trade"
 - use an asterisk, e.g. research* to find researcher, researching, researches, or researched
 - '?' wildcards are supported, e.g. teach??? to find teaching or teachers
- searches are case insensitive

A search for Zambia and poverty narrows the search to databases containing both those terms. The subject terms included in the catalogue are at a very broad level. For example, a search on 'aids' returns only survey datasets. It does not return the World Bank, United Nations Common Database, Eurostat or any of the international databanks that in fact contain very detailed time series data on HIV/Aids. This is because there is no common system of metadata creation for the international databanks and, because of the size and range of the databanks, metadata can only be produced at a very broad level, e.g. 'health' rather than 'HIV/AIDS'. Most of the data providers produce a country list in an attached file for each databank but there is very little searchable data on the subject covered and virtually none in any standard form. This is why to find the data you want we recommend that after locating the databases that include data on the country you are interested in, **use the dataset user guides rather than the catalogue for locating the data you need.**



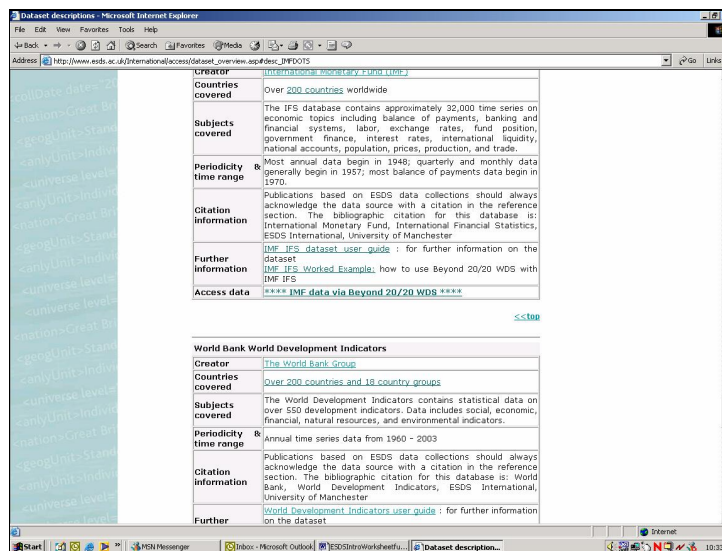
Activity 3 : Locating Data Using the ESDS International dataset user guides

The dataset user guides are the principle step in finding the data you need. Even before you access the data, the website provides a wealth of information about the databanks and the additional services ESDS International provide. Open a browser (the screen shots below show Internet Explorer) and go to www.esds.ac.uk/international to enter the ESDS International website:



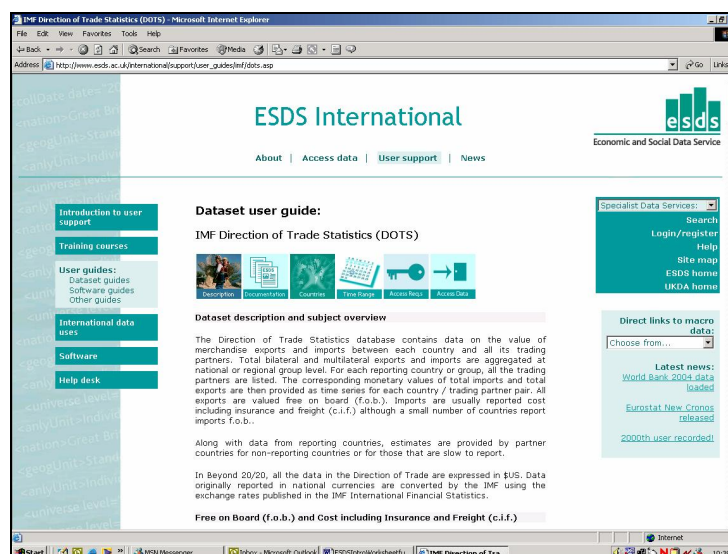
Select **Access Data** from the navigation bar at the top of the green ESDS International screen. The Access Data page contains a list of all the databanks available through ESDS International (you may have to scroll down to see the list of micro datasets).

The ESDS International website provides a full description of each dataset. As an example, we will now look at the documentation for one of the International Monetary Fund databanks. Select **IMF Direction of Trade Statistics** from the list.



This takes you to a summary table containing a brief overview of the dataset. Each databank has its own table on this page. The table has links to more information about the data provider and tells you how to cite the data. If you use data from ESDS International, the data source always needs to be acknowledged with a citation in the references. Citations also enable the ESRC to see how the data is being used and help us develop the data portfolio. The table also contains a link to the dataset user guide.

Select the **IMF DOTS dataset user guide** link.



Use the ESDS International dataset user guide for the Direction of Trade to help you **answer the following questions:**

- **How many countries and regional groups are in the Direction of Trade?**
- **Give an example of a regional group**
- **Can you think of any countries which are not there? Why do you think that is?**
- **What years does the database cover?**
- **What do you do if you want earlier data?**

Each dataset has its own detailed user guide. The dataset description and subject overview contains a list of the main subject groups in a dataset and where possible a link to a full list of indicators. The documentation section contains links to publications accompanying the data and any relevant guides or manuals. In this section you can find any available information on sources and definitions from the data producer. For example, in this case, you can find the IMF's own guide to the Direction of Trade describing the compilation, concepts, methodology, coverage and reliability of the databank.

The dataset user guide also contains a list of the **countries** covered by the database and the **periodicity and time range**.

Once you are inside a database, you can use the Search functionality of Beyond 20/20 to locate the data you need at item level.

Answers to questions:

The IMF Direction of Trade covers around 186 countries and 12 regional groups, an example of which is "Africa". Countries not included in the data include Palestine and Taiwan because of their ambiguous status. (IMF says "Because of political reason, Taiwan is not included in the DOTS. The data for Taiwan is not included as part of the mainland China data either. However, Taiwan is included in the DOTS yearbook.") Countries that have declared independence but not been formally recognised such as Western Sahara, Somalialand are also not found in the database. Other countries missing from the Direction of Trade include Lesotho and Bhutan. A look at the country notes tells you that the data for these countries are included with data for larger neighbours.

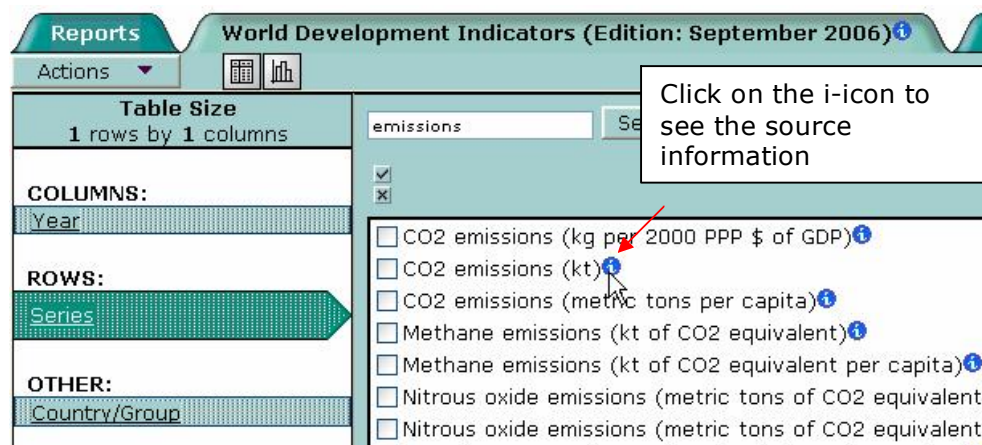


Activity 4 : Looking at sources

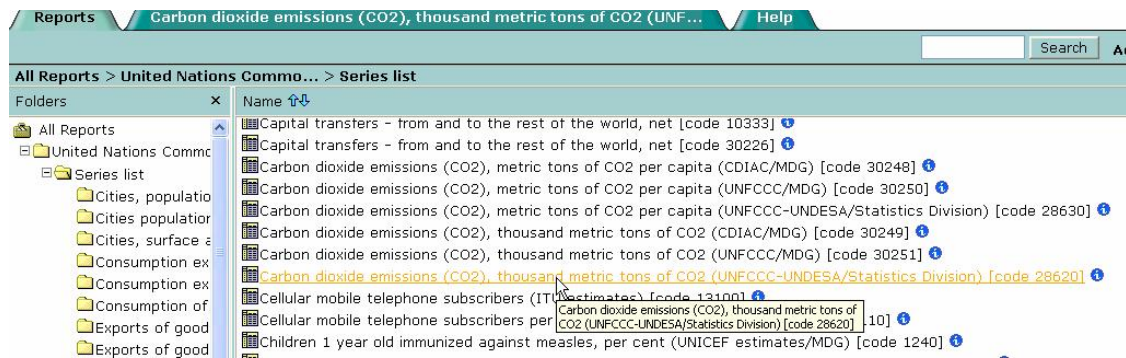
In this task you will compare the data and sources for theoretically identical indicators from different databanks. You will need your Athens username and password to complete this task.

Select one indicator, CO₂ emissions per capita for Chile, say, from the UN Common Database and the World Bank World Development Indicators.

To see the source for the World Bank data, use the skills you learnt in Activity 1 to open the World Bank World Development Indicators and locate the CO₂ emissions series. Then click on the i-icon next to the series name to see the source information.



The source information for the UN Common database can be accessed from the dataset user guide. The documentation section has links to the [UN's Series Notes page](#) which provides a list of series presented in the UN Common Database. By drilling down into a selected series you can see the source information for each series, a list of the countries for which this data is available and definitions of the terms used within the series. You can also access the database through the **Access Data** button in the dataset user guide. Once open, click on **Series list** to see the list of indicators, then scroll down to **Carbon dioxide emissions (CO₂), thousand metric tons of CO₂ (UNFCCC-UNDESA/Statistics Division) [code 28620]** to see the carbon dioxide emissions data for Chile.



Answer the following questions (a few concise sentences).

- **What is the database's source for this indicator?**
- **Is the value the same for the same years in both databases?**

Answers to questions:

Carbon dioxide emissions are those stemming from the burning of fossil fuels and other man-made sources and are measured in metric tons. The UN Common database contains 3 tables each of data for CO2 emissions per capita and CO2 emissions in total. Each pair of tables reports data for most countries to a different standard. However data for Chile is only reported from one source, the Carbon Dioxide Information Analysis Center (CDIAC), <http://cdiac.esd.ornl.gov/home.html>, which is the global-change data and information analysis center of the U.S. Department of Energy. The World Bank uses the Carbon Dioxide Information Analysis Center at the Oak Ridge National Laboratory in Tennessee as the source of its CO2 emissions data.

The World Bank data for total CO2 emissions for Chile starts in 1960, whereas the UN data starts in 1980. The actual data values vary by a tiny amount between the two databases (under 0.2%), but the trends are the same.

CO2 Emissions in Chile

Source: World Bank and UN

