14. AIM/Common Database: A Tool for AIM Family Linkage

Go Hibino[1], Yuzuru Matsuoka[2], and Mikiko Kainuma[3]

Summary. Several teams have been developing AIM models in different countries to analyze their own policy options to mitigate greenhouse gas emissions. Although they need to collect their own data, there are a lot of data that can be shared. The AIM/Common Database, consisting of common formats and code system, has therefore been developed as a tool for AIM Family linkage. There are several advantages by using this database. First, as the formats and code system are common, one team can use data collected by the other teams without special data conversion. Thus the duplication of data collection and conversion can be greatly avoided. It also helps in setting up country specific data by referring to the data provided by other teams. Second, outputs of one model can be used by another model without having to check the data formats. It helps to share the relevant outputs of different models in AIM family. Third, as AIM/Common Database can provide a user-friendly interface based on the common data format, it facilitates users to access input and output data stored in the database.

14.1 Background

The AIM Family consists of a number of models, including AIM/Enduse, AIM/Trend, AIM/CGE, AIM/Material, AIM/Climate and AIM/Impact. Each model has been developed in various countries such as China, India, South Korea and Japan. As many teams in different countries are involved in model development, AIM/Common Database plays a very important role in the AIM Family by linking the different models and modelers in different locations.

It is necessary to collect and process a large amount of statistical data in the model development and application stages. There is a significant overlap in the data required by different models and modeling teams. To ensure efficient implementation of this work, there is a need to develop standard database system containing common formats and codes.

Within the AIM Family, the simulation results of one model are often utilized as input by other models. Additionally, data and results of one modeling team are often useful to other teams as well. Moreover, users of the simulation results are not limited to the AIM development teams, but also include researchers, government officials, NGOs, etc., who use them as a reference in their decision-

---

[1] Fuji Research Institute Corporation, Tokyo 101-8443, Japan
[2] Kyoto University, Kyoto 606-8501, Japan
[3] National Institute for Environmental Studies, Tsukuba 305-8506, Japan

making. With this large number of widely dispersed users, data sharing must be carried out under a unified code system and unified file format. Otherwise, each user would have to learn the data structure for each model independently, which would impair efficiency.

The AIM/Common Database was developed in order to solve these problems and construct linkages for the AIM Family.

14.2 Outline of AIM/Database

Figure 1 shows an outline of AIM/Database. AIM/Database contains various types of datasets including statistics, outputs by AIM Family modelers, outputs by other modelers, and estimates by international organizations and governments.

AIM/Database consists of the AIM/Primary Database and the AIM/Common Database. The AIM/Primary Database is made up of observed value and estimated value by international organizations, national governments and research institutions. These file format is the original state, so is not standardized. There are magnetic media including statistics edited by international organizations, the first processed statistics for the AIM/Common Database, worksheets input manually from book type statistics, and so on.

On the other hand, the AIM/Common Database is used directly by other models in the AIM Family. The code systems and data table formats are standardized so as to allow easy use of the data. The AIM/Common Database mainly stores datasets related to observed values quoted from statistics and estimated values obtained by the models and organizations with a specific format. The statistical data are stored automatically by a module from the AIM/Primary Database. The module makes it easy to update the statistical data. As modelers use datasets in the AIM/Common Database, the module developed by each modeler imports data from the database directly. The modelers do not necessarily have to use the statistics for all input data in the database. As for the output of the model, the modelers are required to store data in the AIM/Common Database in a specific format in view of data sharing. Both manual operation and the use of modules developed by modelers are acceptable as export methods. Users other than AIM development teams, such as administrators, other modelers, NGOs, and so on can therefore efficiently use the datasets though the interface of the database without knowing the detailed data structure of each model.

At the operational level, AIM/Database is a decentralized system. Specific databases and interfaces are maintained by different users in a decentralized way. However, in order for the work related to data gathering and data exchange to proceed smoothly, certain guidelines have been established regarding the development and management of the AIM/Common Database.

The AIM/Common Database is distributed via a Web site, and the AIM project teams in individual countries download the database files from this Web site. The Web site has an area where access is restricted to the development teams only, and an area accessible to the public. The development data are published only in the
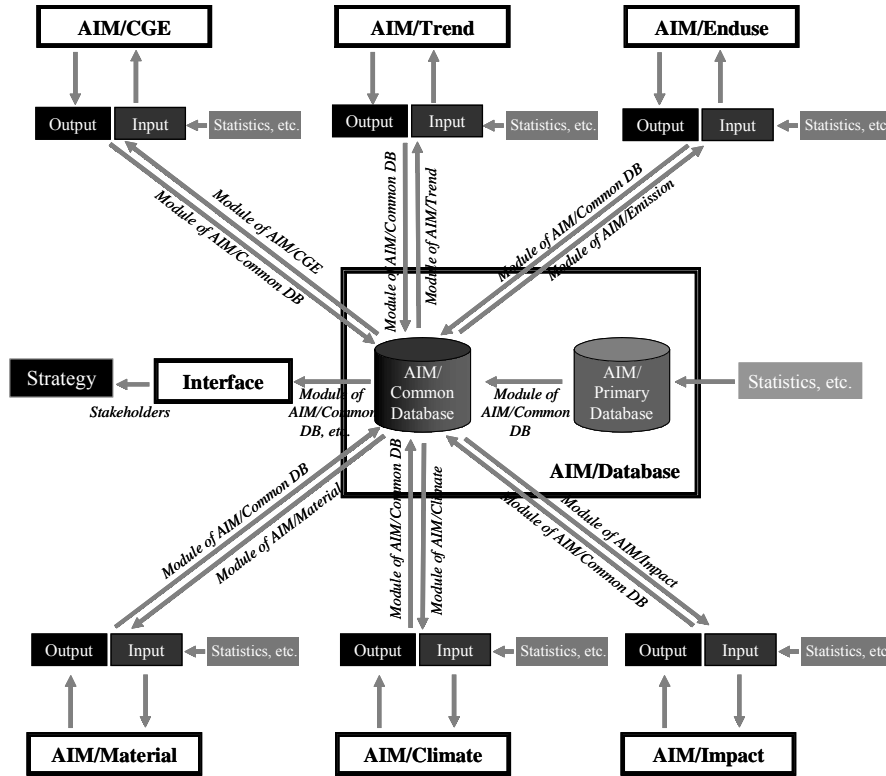
Fig. 1. Outline of AIM/Database

area restricted to the development teams. Users other than AIM development teams, such as researchers, policy makers, and NGOs, can obtain the output of the AIM Family from the site through the area accessible to the public.

14.3 AIM/Primary Database

The AIM/Primary Database is made up of observed value and estimated value by international organizations, national governments and research institutions. These file format is the original state, so is not standardized. The formats vary among the different materials; e.g., magnetic media, books, Web sites, copies of theses, and so on. Magnetic media include not only media such as CD-ROMs and so on issued by statistics institutions, research institutes, etc., but also worksheet format files of data from printed matter input by AIM development teams, and files of data contained in CD-ROMs, etc. that have been converted or processed into a format allowing them to be easily imported into the AIM/Common Database.

With regard to the components of the AIM/Primary Database, lists describing the items such as Media, Language, Data format, Abstract, Data collection period, Published year, Price, etc. have been prepared in order to maintain and manage the components.

14.4 AIM/Common Database

14.4.1 Structure of AIM/Common Database

The AIM/Common Database contains a large number of database files in Microsoft Access format. One of these is a database file that manages the code system of the AIM/Common Database, Code.mdb. The others are database files that mainly manage the statistical datasets and output of AIM Family models. These datasets are stored in the data table of the database file. Table 1 shows the data table in Energy_IEA.mdb. The table stores the country-wise, sector-wise, and energy-wise energy flows quoted in IEA (2001a, 2001b). The data table format in the other database files is almost the same as this data table. The codes ITEM, COMMODITY, SECTOR, COUNTRY, and REFERENCE in the fields correspond to the codes managed in Code.mdb.

Database file to manage code system, Code.mdb

Code.mdb is a database file that manages the code system of the AIM/Common Database. The database file contains tables listing the codes for items, commodities, sectors, countries, districts, and references. Each table is explained in turn below.

Code_Item. Code_Item is the code table for items (Table 2). An item shows what value the data indicate. An item code with the same unit is used in all the tables. The unit should always be listed in the NAME field. Input of the item is indispensable in the data tables.

Code_Commodity. Code_Commodity is the code table for commodities (Table 3). The code is used as the data table stores commodity-wise data; e.g., industrial commodity production, energy consumption (energy-wise), or agricultural commodity production. When the commodity code is not required, it can be omitted in the data table.

Table 1. Data table in Energy_IEA.mdb

ITEM	COMMODITY	SECTOR	COUNTRY	YEAR	VALUE	REFERENCE
EN_FLW	EN_COL	TPES	AUS	1961	16107	IEA_99
EN_FLW	EN_COL	TPES	AUS	1962	16394	IEA_99
....	....	....	....	......	....	....

Table 2. Code_Item code table in Code.mdb

ID	CODE	NAME	NOTE
101	EN_FLW	Energy flow (kTOE)	
102	EN_ELE	Electricity output (GWh)	
...	......	......	......
201	PP_TTL	Female population (thousand)	
202	PP_MAL	Male population (thousand)	
...	......	......	......
401	IN_PRD	Industrial commodity production (tons)	
402	IN_PRD_L	Industrial commodity production (meters)	
...	......	......	......
501	NA_GDP_DF	GDP deflator	
502	NA_GDP_CD	GDP (current million US$)	
503	NA_GDP_CN	GDP (current million LCU)	
...	......	......	......

Table 3. Code_Commodity code table in Code.mdb

ID	CODE	NAME	NOTE
...	......	......	......
203	EN_OLK	Kerosene	
204	EN_OLG	Gasoline	
205	EN_OLD	Diesel oil	
...	......	......	......
401	MN_IRN	Iron ore	
402	MN_LMS	Limestone	
...	......	......	....

Table 4. Code_Sector code table in Code.mdb

ID	CODE	NAME	NOTE
...	......	......	......
201	IN_STL	Iron and steel	
202	IN_CHM	Chemical and petrochemicals	
203	IN_NMT	Non-metallic minerals	
...	......	......	......
401	OT_AGR	Agriculture	
402	OT_CMM	Commercial and public services	
403	OT_RSD	Residential	
...	......	......	....

Code_Sector. Code_Sector is the code table for sectors (Table 4). The code is used as the data table stores sector-wise data; e.g., energy consumption or value added. When the sector code is not required, it can be omitted in the data table.

Code_Country. Code_Country is the code table for countries (Table 5). The relationships between the codes and major organizations' codes are shown in the table. The definition of a country, such as changes resulting from separation or amalgamation, is listed in the NOTE field. Input of the country is indispensable in the data tables.

Code_District. Code_District is the code table for districts in a country (Table 6). The code is used as the data table stores district-wise data. When the district code is not required, it can be omitted in the data table.

Code_Reference. Code_Reference is the code table for references (Table 7). The data sources of values in the data table are listed. Input of the reference is indispensable in the data tables.

Table 5. Code_Country code table in Code.mdb

ID	CODE	NAME	UN	WB	ISO1	ISO2	....	NOTE
4	AFG	Afghanistan	AFG	AFG	AFG	4		
8	ALB	Albania	ALB	ALB	ALB	8		
12	DZA	Algeria	DZA	DZA	DZA	12		
16	ASM	American Samoa	ASM	ASM	ASM	16		
20	AND	Andorra	AND	ADO	AND	20		
24	AGO	Angola	AGO	AGO	AGO	24		
28	ATG	Antigua and Barbuda	ATG	ATG	ATG	28		
31	AZE	Azerbaijan	AZE	AZE	AZE	31		
32	ARG	Argentina	ARG	ARG	ARG	32		
36	AUS	Australia	AUS	AUS	AUS	36		
40	AUT	Austria	AUT	AUT	AUT	40		
......	......		......	......	......	......		

* UN: Country code used in UN statistics (2001), WB: Country code in World Bank (2002), ISO1: ISO3166 code A3, ISO2: ISO3166 code number

Table 6. Code_District code table in Code.mdb

ID	COUNTRY	CODE	NAME	NOTE
1	IND	101	Adiladbad	
2	IND	102	Anantapur	
3	IND	104	Cuddapah	
......	......	......	......	......

Table 7. Code_Reference code table in Code.mdb

ID	CODE	NAME	NOTE
101	IEA_99	IEA Energy Balance 1999	IEA Energy Balances of OECD Countries 1960-1999 (CD-ROM), IEA Energy Balances of Non-OECD Countries 1971-1999 (CD-ROM)
200	UN_STY_45	UN Statistical Yearbook Forty-fifth issue	UN Statistical Yearbook Forty-fifth issue (CD-ROM)
201	UN_CMM_01	UN Industrial Commodity Production 1999	UN Industrial Commodity Production Statistics Database 1950-1999 (CD-ROM)
202	UN_POP_01	UN World Population Prospects 2000	UN World Population Prospects 2000 (CD-ROM)
......	......	......	......

Database files to manage datasets

The database files other than Code.mdb mainly manage statistical datasets and output of AIM Family models. The structure of the files is shown in Fig. 2. The two upper tables manage datasets, while two tables, one form, and one module play the role of importing data from the AIM/Primary Database and the output of other models. The tables, form, and module are described below.

Data table. This table consists of data imported from the statistical data in the AIM/Primary Database and the output of models using the unified code system. The table format is standardized and is composed of the Item, Commodity, Sector, Country, District, Year, Value, and Reference fields.

Code table. This table shows explanations concerning the codes used in the Item, Commodity, Sector, Country, District, and Reference fields in the Data Table. It is linked with the table in Code.mdb.

Import file table. This table is used when the module imports statistical data from the Primary Database or the output of models to the Common Database. The table indicates from which item and which file in the Primary Database to import the data and to which item in the Common Database.
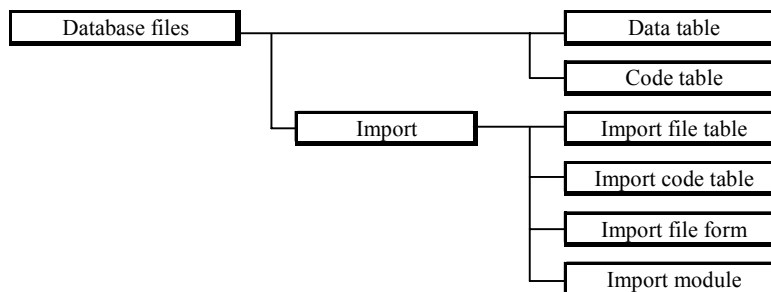
Fig. 2. Structure of database files

Import code table. This table is used when the module imports statistical data from the Primary Database or the output of models to the Common Database. The table indicates which code in the Common Database corresponds to the code in the Primary Database.

Import form. This form is used when the module imports statistical data from the Primary Database or the output of models to the Common Database. It is the interface for importing data from the Primary Database and users can execute the module on the form.

Import module. This module is used when the module imports statistical data from the Primary Database or the output of models to the Common Database. It is written with Access VBA.

14.5 Development and Management of AIM/Common Database

Maintenance and management of the AIM/Common Database is not carried out by a specific team in a centralized way, but in a decentralized way by the model development teams of each country. However, in order for the work related to data gathering and data exchange to proceed smoothly, certain guidelines have been established within the AIM development team regarding the development and management of the AIM/Common Database. These guidelines are introduced below.

14.5.1 Planning phase

The guidelines on planning phase of AIM/Common Database are described below.

- AIM/Common Database is to be developed with using MS Access 2000 or later.
- Data table is to be made up of Item, Commodity, Sector, Country, District, Year, Value and Reference fields. Data in Item, Country, Year, Value, and Reference are indispensable.
- The code system (Code.mdb) of the AIM/Common Database is to be common to each development team in each country, and the codes in the data files must conform to this system.

14.5.2 Accumulation phase

The guidelines on accumulation phase of AIM/Common Database are described below.

- Coordination and management of the code system is to be carried out by AIM/Common Database coordinators (one to two persons for the overall AIM project).
- When creating a database file for the AIM/Common Database, if a new data source is used, the relevant information is to be added to the data source lists of the Primary Database following to the prescribed format.
- When a database file of the AIM/Common Database is created or updated, the document is to be prepared following the prescribed format for the contents of the file.
- When a database file of the AIM/Common Database is created or updated, notification to that effect is to be given on the AIM Development Team Web Site.
- When the creation of new data is desired, notification to that effect is to be given on the AIM Development Team Web Site.

14.5.3 Communication phase with AIM family models

The guidelines on communication phase with AIM family models via AIM/Common Database described below.

- Users within the AIM development teams need to have basic knowledge of MS Access (database structure, how to view data, copying to MS Excel, etc.).
- A precondition for the development of AIM Family models is that they conform to the data format and code system of the AIM/Common Database.
- The output of AIM Family models is to be provided following the data format and code system of the AIM/Common Database so as to enable smooth exchanges of data between different teams.
- Since the database files of the AIM/Common Database are in Microsoft Access format, modelers need to learn how to extract data files in MS Access.

14.5.4 Dissemination phase

The guidelines on dissemination phase of AIM/Common Database are described below.

- AIM/Common Database is to be disseminated for external uses via Web site.
- For external users of data accessible to the public, an interface is to be provided that enables use of the data without knowledge of MS Access. Therefore, modelers of AIM Family are to provide the aggregated results in the html format for external users.
- Concerning detailed data, the files including simulation results are to be uploaded in MS Access format. External users download the file and use it on their end. Modelers are to provide not only data table but also pivot table and pivot chart in the file. The table and chart make it easy for external users without knowledge of MS Access to use data.

14.6 An Example: How to Interact with AIM Family Models

Linkage between AIM/Enduse and AIM/Trend through the AIM/Common Database is described here as an example of AIM Family linkage.

The AIM/Enduse model is the tool to estimate future greenhouse gases emission with using detailed, sector-by-sector information about future driving forces given externally. For example, the production volume of each industrial product in the industrial sector and the demand volume of each application (e.g., heating and air conditioning) in the home sector must be given. As most of the estimates provided by governments and international research institutes are macro indicators such as economic growth rates and population, each analyst is often required to individually estimate the driving forces on a segment-by-segment basis. Analyses previously conducted in Japan, China, India, and Korea, based on end-use models, also used estimated segment-by-segment future driving forces, although they were based on the GDP, population, and other macro indicators predicted and planned by governments.

The prediction of these future values is based on a very simple methodology. First, regression equations are derived from the actual driving force volumes and GDP measurements. Then, the future driving force volumes are estimated using the regression values and future GDP measurements predicted by the governments or research institutes. Finally, the values obtained are converted into the form that the model can use for simulation. Despite the simplicity of the methodology, this series of actions is still not systematically documented, which makes it difficult to update the estimated values and transfer the process of estimate to others.

Under the circumstances, new capabilities were added to the AIM Family models and databases to construct a system that can systemize and automate the above series of actions. In addition, using the AIM/Trend Program Language
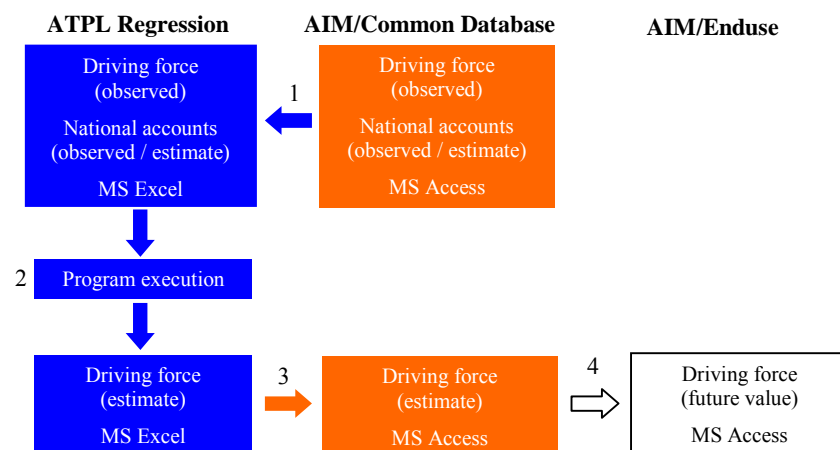

Fig. 3. Work flow: Linkage between AIM/Enduse and AIM/Trend through AIM/Common Database

(ATPL) developed by AIM Project Team (2001), ATPL Regression was constructed as a new program that derives future driving force volumes from the regression equations for driving forces and GDP measurements (Fig. 3 and Fig. 4).

1. The Import module of ATPL Regression imports the information about the actual driving force volumes and actual/estimated future value added figures contained in the AIM/Common Database into an Excel file.
2. The Regression/Projection module of ATPL Regression derives regression equations from the actual driving force volumes and actual value added figures. Then, the derived regression equations and future value added figures are used to estimate the future driving force volumes. The computation results are output as Excel data.
3. The Import module of AIM/Common Database imports the computation results provided by AIM/Trend into the AIM/Common Database as MS Access data.
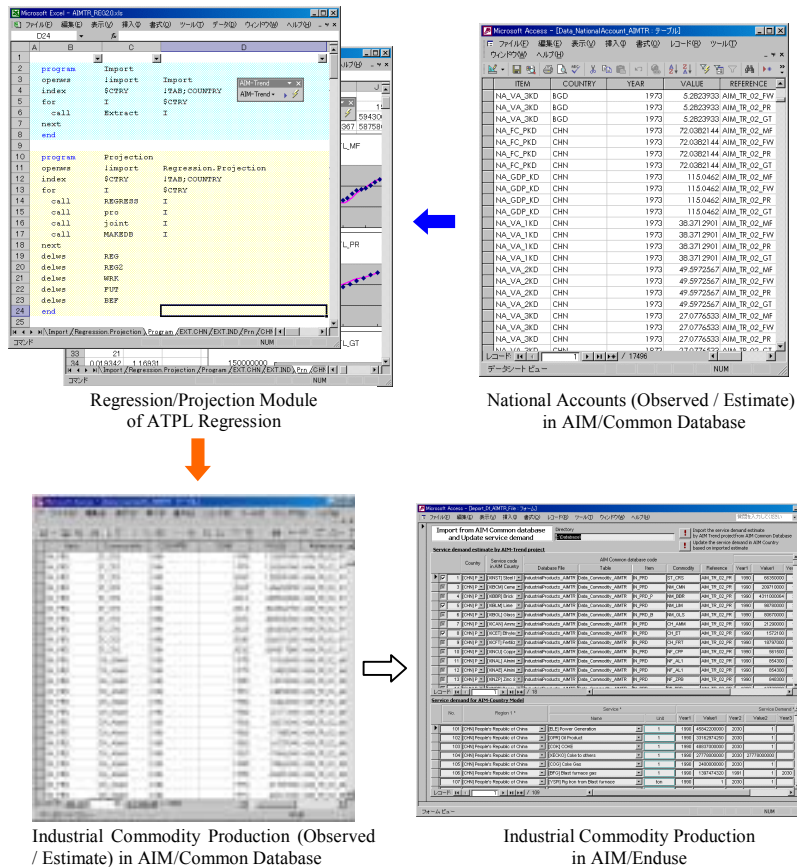

Regression/Projection Module
of ATPL Regression

National Accounts (Observed / Estimate)
in AIM/Common Database

Industrial Commodity Production (Observed
/ Estimate) in AIM/Common Database

Industrial Commodity Production
in AIM/Enduse

Fig. 4. Examples of model and database displays: Linkage between AIM/Enduse and AIM/Trend through AIM/Common Database

4. The Import module of AIM/Enduse imports the future driving force volumes contained in the AIM/Common Database into AIM/Enduse as MS Access data.

14.7 Concluding Remarks

There are several AIM modeling teams in Asian countries such as China, India, South Korea and Japan. Although they need to collect their own data, there are a lot of data that can be shared. If an efficient database system is developed to facilitate this sharing, data collection work can be reduced. The AIM/Common Database, consisting of common formats and code system, has therefore been developed as a tool for AIM Family linkage. Maintenance and management of the AIM/Common Database is not carried out by a specific team in a centralized way, but in a decentralized way by the model development teams of each country based on certain guidelines.

The following effects have been realized through application of the AIM/Common Database.

- A team is able to easily utilize statistical data gathered or processed by another teams, and overlapping of work is greatly reduced.
- Results of other models can be easily utilized as input data or reference data for a model, without having to check the data formats of the other models.
- The output of models is in a common format, allowing it to be readily used by stakeholders and promoting the efficient utilization of simulation results.

The AIM/Common Database should play an important role to link the models in AIM Family that are used for integrated assessment. In order to maintain the efficiency, it is essential for each modeling team to observe the guidelines for its proper maintenance and management.

References

AIM Project Team (2001) AIM Trend model user's manual. AIM Project Team, AIM/Trend Model distribution CD

IEA (2001a) Energy balances of OECD countries 1960-1999. International Energy Agency, CD-ROM

IEA (2001b) Energy balances of non-OECD countries 1970-1999. International Energy Agency, CD-ROM

UN (2001) Statistical yearbook. United Nations, CD-ROM

World Bank (2002) World development indicators. World Bank, CD-ROM