

## **Quality Assurance, Arrangements With Common Questionnaires and Conflicting Statistics**

### Note by the World Bank Development Data Group

The Development Data Group manages over 40 databases of interest to the Bank ranging from financial, economic, social, and environmental statistics. This large collection is made available internally through a web-based system called SIMA. It is used for research/analysis, official internal documents as well as production of flagship publications such as the *World Development Report* and *Global Development Finance* and the *World Development Indicators*. The majority of these databases are collected by partner agencies (FAO, IMF, UN, UNESCO, for example) and other private or commercial data providers. The data posted on SIMA are mainly data reported by official sources of data.

Some level of quality assurance and review is performed on all data used in key Bank products. However, there are some databases for which the Bank takes greater care and responsibility for quality assurance. Two categories of special importance are: financial data collected through the Debtor Reporting System (DRS); and a core set of macro and social indicators used in key Bank strategy and project documents.

### **I. Quality Assurance for Bank's Key Data Sets:**

*Financial Statistics* The Bank's general conditions require a borrowing or guaranteeing member country to furnish reliable and timely external debt information to assess a borrowing country's foreign debt situation, and creditworthiness and economic management and to conduct its country economic work. The external debt that the country must report to the Bank includes all public debt and private nonguaranteed debt. The basic reporting consists of the following:

1. Annual summary reports showing public or publicly guaranteed debt and private non-guaranteed debt aggregates. These reports are due within three months after the reporting period.
2. Quarterly reports showing detailed information on each new commitment of a public or publicly guaranteed debt received during the period. These reports are due within one month after the end of the quarter.

As a condition for board presentation of loans and credits, complete reports must be submitted or an acceptable plan of action for complying with this requirement.

The DRS data is used in combination with information obtained from creditor sources through the debt data collection systems of other agencies such as the Bank for

international Settlements (BIS), Organization for Economic Co-Operation (OECD), African Development Bank (AFDB), Asian Development Bank (ADB), Central Bank for Economic Integration, IMF, and Inter-American Development Bank. The information obtained from these organization is used to fill in gaps in reporting and improve the quality of the data reported.

Data validation includes checks for accuracy and completeness of the information through the following steps:

1. Improving coverage and completeness of the statistics through using creditor data .
2. Performing validation exercises on the reported data by comparing the data on stocks of a given year to the flows and the stocks of the following year using the available data on general trends in lending. And checking the consistency of the data from one period to the other.
3. Incorporating information on bilateral and multilateral rescheduling and refinancing agreements once agreements are signed. And verifying the amounts reported under the DRS by the member countries with other sources.
4. Estimating missing data by using terms of commitments, the international market situation.
5. Providing technical assistance for member government on computer system, debt data collection and debt management to improve debt data collection, manipulation and coverage. This takes the form of mission, seminars, and training courses.

*Core Macroeconomic and Social Data* The Bank's country and sector units are responsible for managing a selected set of key macroeconomic and social indicators most frequently used in official country assistance strategy, poverty reduction strategy, and dialogues with the member countries. This include a set of data from national accounts, BOP, GFS, prices, and a set of social indicators including population, education, health, nutrition, poverty, environmental, and so on. To establish a standard platform for specialists from around the Bank who work with this data (collect, validate, analyze, and produce reports), we have come up with a data management tool called the Live Database (LDB). A copy of this Excel-based data management tool has been distributed to each country management unit in the World Bank headquarter and field offices. A similar tool is also set up for the sectoral data users and data managers in the education, health, nutrition, population, poverty, etc. The LDB tool provides a "refresh" function to download available data from official sources, derive indicators, allow data updating, and perform a number of validation functions. On a regular basis, data are downloaded to the central files from country and sectoral LDBs, where an extensive set of validations are performed.

The focus of the mechanical validation processes are primarily to ensure internal consistency. This is achieved through a validation report which is built into the LDB tool. This report uses a graphing and reporting feature to check the data across all topics in the LDB. The reporting feature includes checking the data on a vertical basis against built in additivity rules and on a horizontal basis using growth rates. Unreasonable growth rates can reveal data problems to be further investigated. Another method of validation is to compare data against the previous year. This is done through graphs of the time series to quickly see any discrepancies. Finally, there are periodic validations to cross check data

with statistics published by other agencies. The unresolved issues found in all validations process are used in preparing notes on data, which are included in the main databases and published in statistical publications such as the *World Development Indicators* along with data tables, source, and definitions of terms.

## **II. Consistency with Data Disseminated by Other International Organizations**

As noted above, we make extensive use of other international organizations' data, noting the appropriate source and definitional information. Therefore, ideally there should not be any major inconsistencies between Bank published data and other international organizations. However, for data sets such as national accounts, Bank economists and analysts collect additional data (not always from an official source) to fill-in gaps particularly for most recent periods. To minimize inconsistencies in such cases, we are trying to cross-check data with primary data sources to detect discrepancies to be reported to and sorted out with responsible country or sector units in the Bank. Another source of inconsistency between Bank reported data and other international organizations is due to revisions made to the official UN population data . This is a subject of a discussion at the ACC Subcommittee on Demographic Estimates and Projections, with a paper from the Bank explaining the process and rational in details.

## **III. Conflicts between Nationally Supplied Data and Internationally Comparable Data**

In critical data areas such as poverty data, PPPs, and various social indicators, there is a growing problem with conflicts between nationally supplied data and internally produced cross-comparable data. There is no easy way to correct this as more research and work is needed to establish internationally recognized standards for producing cross-country comparable data in all development economics, social, and environmental sectors. However, a number of activities in the Bank are focused at better understanding sources of discrepancies -- to answer, for example, why there is a growing divergence between nationally produced sample surveys and national accounts data, often causing large discrepancies between indicators that use one or the other source.

Another source of conflict stems from the fact that large organizations with diverse needs have a tendency to create at least two sets of official files on key topics -- one specific for single-country analysis and another for cross-country comparable or for regional/global work. At the Bank, there are a few months during the annual updating cycle, when the two sets don't match. During such periods, we try to clearly label files which are specifically for one or the other purpose.

DECDG  
September 19, 2000