

World Bank Open Data Initiative

The World Bank announced on April 20, 2010 that it will offer free access to more than 2,000 financial, business, health, economic and human development statistics that had mostly been available only to paying subscribers. An initial set of World Bank databases are now available at data.worldbank.org. This initiative is expected, among other objectives, to improve data access and quality.

Background

Data quality work at the World Bank covers many areas of data. The World Bank produces the World Development Indicators and Global Development Finance publications and databases. The World Bank actively participates in international statistical programs such as the Marrakech Action Plan for Statistics, Partnership in Statistics for Development in the 21st Century (PARIS21), International Household Survey Network (IHSN), Inter-Agency Task Force on Finance Statistics (TFFS), and the International Comparison Program (ICP). The World Bank participates in setting international statistical standards such as external debt, national accounts, remittances, and SDMX. And the World Bank contributes to improving statistical capacity in developing countries through the Trust Fund for Statistical Capacity Building (TFSCB), Lending programs (STATCAP), General Data Dissemination System (GDDS) work, and the Accelerated Data Program (ADP). For more details, see the Appendix – World Bank Statistical Programs.

The World Development Indicators and Global Development Finance databases require extensive data quality assurance work. Data for over 200 countries, going as far back as 1960, for over 1,000 indicators are maintained by the Development Data Group. Tools and procedures are employed to check all of the data for internal consistency, consistency with other data sources and previously published data, and for outliers. Recent data quality initiatives have resulted in the following list of recommendations.

1. Choose a set of priority countries and indicators on which to run detailed checks.
2. Have automated checks that flag issues to investigate.
3. Provide a way to annotate explanations for outliers, etc. (or this could be part of the metadata).
4. Have some mechanism that provides a sign-off or check-box for validation (and which would require re-validation with any subsequent data revision).
5. Follow a strict data update/validation schedule.
6. Compare data to multiple sources and the previously published version.
7. Compare data for similar countries and similar series.
8. Prioritize outputs to allocate resources effectively.
9. Have regular meetings for DCS (a data administration system) users to share best practices and provide information on data updates and DCS system issues.
10. Provide training for DCS and useful Excel functions (vlookup, etc.).
11. Have a selective audit trail for key series such as population and GDP.

4/30/2010

12. Before finalizing the database, do a check that runs a complete set of derivations and aggregations against what is actually in the database to check for partial derivations and aggregations.
13. Follow up on DCS system issues with zeroes for exclusion list indicators; system significant digits; gap-filled series transfers to databases beyond the initial transfer, and footnote transfers.
14. Provide sufficient notice when data have been updated so other database owners that "pull" the data know when to update their databases (e.g., GDF, MDG, IDA, Africa, etc.).

Open Data Initiative

A major new initiative at the World Bank is the Open Data initiative which could have significant impacts on data quality. In the announcement of Open Data, World Bank President Robert B. Zoellick said "Broader access to these data allow policymakers and advocacy groups to make better-informed decisions and measure improvements more accurately. They are also valuable tools to support research by journalists, academia and others, broadening understanding of global issues."

The initiative is based on three basic principles:

OPEN: All Bank data are public without any restrictions. There are no subscriptions, no registration—nothing between the user and the data.

ACCESSIBLE: The challenge is to make it easy for users to get to the data and easy for them to find what they are looking for. Data.worldbank.org is accessible to anyone—researchers, journalists, NGOs, entrepreneurs and school children alike are able to tap into the Bank's databases.

SEARCHABLE: A new search tool returns data directly from the database making it easier for people to find what they're looking for. Data.worldbank.org makes it easier to find, use and manipulate data. A catalog lists all available databases that users can search and retrieve data from. The Bank will continue to add databases in the coming months.

The Development Data Group (DECDG), the focal point for statistical activities at the World Bank, plays a leading role in the Open Data initiative. DECDG expects to see improvements in data quality as a result of broader access to data. Questions and feedback from our clients have always been valuable to DECDG in improving data quality. It is anticipated that broader access to the Bank's data will increase the discovery and correction of errors, raise quality issues from increased comparisons of data from different sources, allow the World Bank to better gauge client demand for different indicators and data sets, and bring more data producers and users together in a partnership for improved data quality and use.

Currently, the relationship between data quality and openness occurs at many levels. Within the World Bank, the consultation process with operational staff, our country economists brings much added value to our data. Our country economists collect macroeconomic data for most developing countries for their own monitoring and analysis as well as for the institution's knowledge base. They, along with topical experts in our network units—Human Development, Poverty Reduction and Economic Management,

4/30/2010

Financial and Private Sector Development, and Sustainable Development—review our data during a formal process as well as ad-hoc questions that arise when they are using the data for research and analysis. This review covers not only macroeconomic data but other data compiled from UN agencies such as UIS, UNCTAD, WHO, and FAO; other international organizations such as IMF and OECD; and private firms such as Standard and Poors, KPMG, and PricewaterhouseCoopers. And finally, our clients outside of the Bank also sometimes find problems in the data and contact us through our helpdesk.

The consultation process can uncover data problems which we report back to our partners or it leads to better documentation of our database. Footnotes are often added to report data revisions due to changes in methodology, or to significant events which result in unusual or volatile data points, e.g. natural disasters or conflict.

The Open Data initiative is ongoing. Over the coming weeks and months, the available content will be expanded to provide improved tools for data users and developers, and to issue an "Apps for Development" challenge to developers to find innovative and useful ways to present and visualize the data. Up to now, the Bank's major databases were available only through subscription and provided needed revenue. The loss of this revenue will be offset by an increase in innovations and partnerships with application providers and will likely stimulate the demand and use of statistics by more than the traditional data users. The innovation and higher use will lead to more support for production of better statistics in countries and international agencies which, in turn, will help with monitoring and facilitating development progress.

An exciting possibility of the Open Data initiative is the possibility of collecting data through "crowd sourcing". It is envisioned that as the Open Data initiative advances, users will be able to provide data to the World Bank Open Data website. Although one can imagine much more data becoming publicly available, quality control and proper documentation will be a major challenge. At a recent meeting at the World Bank on Government 2.0, Andrew McLaughlin, Deputy Chief Technology Officer of the White House addressed this topic and suggested that data need to be categorized to distinguish between data reviewed and disseminated by the World Bank from other data. In fact, much more metadata will be required to make the data searchable and to flag data quality issues beyond just the source of data. Many more issues will arise of course but the possibility of crowdsourcing in data collection is promising. The Development Data Group is currently initiating a project to develop a data center to host data collected from a variety of sources to support a research program on economic development and structural change. The data center will not be a pilot for crowd sourcing—it will only be open to invited participants. Still, the project is expected to provide some lessons related to data crowd sourcing in terms of differentiating, documenting, and putting to use data from different sources with disparate levels of data quality.

Appendix – World Bank Statistical Programs

International Statistical Programs

The World Bank works closely with the international statistical community, including the United Nations (UN), the Organisation for Economic Co-Operation and Development (OECD), the International Monetary Fund (IMF), regional development banks and donors by:

- Participating in the UN Statistical Commission and other statistical forums to develop appropriate frameworks, guidance and standards of good practice for statistics.
- Building consensus and defining internationally agreed indicators, such as indicators for the MDGs.
- Establishing data exchange processes and methods.
- Assembling, analyzing, and disseminating data in user-friendly online print publications.
- Working together to help countries improve statistical capacity.

In addition to compiling international data sets, which are generally based on data generated by national statistical systems, the World Bank supports a number of programs to collect transnational data. This specific data can only be collected by a globally coordinated program because the phenomena they are concerned with cross national boundaries.

Marrakech Action Plan for Statistics

DECDG works with developing countries to increase the capacity, efficiency, and effectiveness of their statistical systems. The World Bank has played a critical role in establishing and implementing a global action plan for statistics in response to the need for better data as expressed through the series of International Roundtables on Managing for Development Results.

At the Second International Roundtable on Managing for Development Results (Marrakech, 2004), the international statistical community put forward a six-point program to address statistical capacity building needs at both the international and national levels and to provide the basis for a shared work program among development partners.

The key achievements of the Marrakech Action Plan for Statistics (MAPS) include:

- stimulating increased investment in statistical systems;
- establishing the International Household Survey Network;
- establishing a global program for the 2010 round of censuses with partial funding provided through the Development Grant Facility, in cooperation with the United Nations;
- providing support to PARIS21 to scale up national strategic statistical planning for developing countries; and,
- initiating an accelerated program for improving the use of household survey data, with a focus on Africa.

The Third Roundtable on Managing for Development Results in Hanoi (2007), recognized that improvements in statistical capacity need to be given higher priority in the policy dialogue between

4/30/2010

donors and countries. Recognizing that statistical capacity weakness has not always been adequately addressed in aid programs and national budgets, development partners agreed to create a new program that:

- provides ways to better coordinate donors work in this area;
- recommends a program-based approach; and,
- suggests sharing expertise between more advanced and developing national statistical offices.

More information is available at www.worldbank.org/data/action.

Partnership in Statistics for Development in the 21st Century (PARIS21)

The World Bank is a founding member of the PARIS21 consortium, which was set up to raise awareness among users and producers, to promote statistical capacity building in developing countries, and to develop a culture of evidence-based decision making. The consortium, which is managed by a small secretariat based at OECD-DAC headquarters in Paris, holds regional and country statistical workshops, advocates better statistics, promotes donor coordination, and provides advice and support to data providers and users world-wide. The Bank participates in the governance of the consortium, provides financial and technical support, and works closely with other partners, especially through the operations of the Trust Fund for Statistical Capacity Building. The World Bank has provided funding through the Development Grant Facility to expand PARIS21's support for the development of national strategies for the development of statistics and implementation of the accelerated program for basic statistics.

More information is available at www.PARIS21.org.

International Household Survey Network (IHSN)

In developing countries, sample surveys are a key source of socioeconomic indicators for policy design and evaluation, and for monitoring key indicators (including 25 out of the 48 MDG indicators). But past survey programs have often failed to provide the necessary flow of reliable and comparable data. International survey programs are not harmonized and their implementation is poorly coordinated. This has led to conflicting or duplicating survey activities and to the production of nonharmonized and non-comparable data. Also, existing microdata are in many cases poorly documented, difficult to access, and therefore remain largely unexploited. As recommended by the Marrakech Action Plan for Statistics, the International Household Survey Network (IHSN) was established in 2004 to address these issues. The IHSN Secretariat, hosted by DECDG at the World Bank, seeks to coordinate international survey programs and harmonize survey methods. It also maintains a central survey inventory, and develops tools and guidelines to foster accessibility and use of survey data. All major international survey sponsors have joined the IHSN (UN agencies, regional development banks, and others). At the request of some members, the IHSN is now expanding its activities to provide expertise and solutions for the management of routine data collection activities, in particular in the health sector (in cooperation with the Health Metrics Network and WHO). The IHSN core program is funded by the World Bank Development Grant Facility, and administered by the PARIS21 Secretariat at OECD.

More information is available at www.internationalsurveynetwork.org.

Inter-Agency Task Force on Finance Statistics (TFFS)

The Inter-Agency Task Force on Finance Statistics (TFFS) is comprised of the Commonwealth Secretariat, the Bank for International Settlements (BIS), the European Central Bank, Eurostat, the International Monetary Fund (IMF), the Organisation for Economic Co-Operation and Development (OECD), the Paris Secretariat, the United Nations Conference on Trade and Development (UNCTAD), and the World Bank. It was established in 1992 as one of the interagency task forces formed under the aegis of the United Nations Statistical Commission and the Administrative Committee on Coordination-Sub-Committee on Statistical Activities. It was reconvened in 1998 to coordinate work among the participating agencies to improve the methodological soundness, transparency, timeliness, and availability of statistics. The TFFS is a useful umbrella for greater coordination of work on external debt statistics. The TFFS produced the External Debt Statistics: Guide for Compilers and Users in 2001 (Debt Guide) to support the work of compiling external debt statistics worldwide and launched the Quarterly External Debt Statistics database.

International Comparison Program (ICP)

The International Comparison Program (ICP) is a global statistics initiative to collect internationally comparable price data. These data are used to compute purchasing power parities (PPPs), which allow standards of living to be compared across countries. At the request of the international statistical community, the World Bank coordinates and manages the ICP in developing countries and works closely with OECD and Eurostat, who coordinate the program in the developed world. The completion of the 2005 round covers 146 economies, making it the largest international data collection exercise in the world. These new PPP numbers replace previous benchmark estimates, some of which date back to the 1980's.

More information is available at www.worldbank.org/data/icp.

Statistical Capacity

Developing countries face a number of problems in providing statistics that are reliable, relevant and timely. They find themselves caught in a vicious cycle – under-investment in national statistical systems which hinders activities and results in poor quality of data which policy makers cannot use. As a result, weak demand for data occurs and fewer resources are made available for their production.

The World Bank is committed to helping developing countries break out of this cycle. Our work includes helping countries develop National Strategies for the Development of Statistics (NSDS), helping countries invest in statistical capacity improvement, and helping mobilize the expertise of the international statistical system. We also monitor developments in statistical capacity and maintain information about the statistical capacity of countries.

More information is available at www.worldbank.org/data/statcap.

Trust Fund for Statistical Capacity Building (TFSCB)

The Trust Fund for Statistical Capacity Building (TFSCB) was established in 1999 by the Development Data Group to provide grants to developing countries undertaking improvements in their statistical systems. Funded by several donors, the TFSCB provides small grants with a maximum implementation

period of three years. Recently, emphasis has been placed on supporting the preparations of national strategies for the development of statistics. As of today, the TFSCB has approved funding for more than a hundred statistical projects, which represents an investment in statistics of close to \$22 million.

Lending programs (STATCAP)

The World Bank's STATCAP is a lending program to assist member countries finance investments to upgrade the capacity of their statistical systems. It is designed as a flexible instrument to meet the needs of countries at different levels of development. A key requirement for participation is that countries should have prepared a national strategy for the development of their statistical system.

General Data Dissemination System (GDDS)

The General Data Dissemination System (GDDS) is a development framework of national statistical systems. It emphasizes the publication of metadata or detailed descriptions of methods and procedures used for the most important data sets. It also encourages countries to publish plans for improving data quality and coverage. The Data Group manages a project in anglophone Africa, funded by the UK Department for International Development (DFID), to help countries implement their plans to improve social and demographic statistics. The GDDS covers economic, financial, and socio-demographic data and is jointly supported by the IMF and World Bank. Participation has rapidly increased in recent years. By the end of May 2006, 83 countries had published metadata on the IMF Data Standards Bulletin Board, including more than 80 per cent of the countries in Sub-Saharan Africa.

More information is available at www.worldbank.org/data/gdds.

Accelerated Data Program (ADP)

Household survey programs in developing countries provide crucial data for policy design, monitoring and evaluation. But various issues of reliability, timeliness, comparability and accessibility reduce the relevance and use of these data. Unless urgent action is taken to improve survey programs, many countries will not be in a position to report satisfactorily on the achievement of the MDGs by 2015. To complement the activities undertaken at the international level by the IHSN, a pilot Accelerated Data Program (ADP) was launched in 2006 to address issues at the national level. The ADP works with national agencies in selected countries on:

- Establishing national survey databanks following international standards and best practices, to make existing survey data more usable and accessible.
- Analyzing and assessing the relevance, comparability, and reliability of existing survey data, to increase the future data collection.
- Implementing improved, more cost-effective, and sustainable national survey programs.

Although the pilot ADP initially planned to cover up to 12 countries, more than 20 have already joined the program. International partners are contributing to its implementation, such as ESCAP in Asia and Inter-American Development Bank in Latin America. The pilot ADP is funded by the World Bank Development Grant Facility, and implemented with the PARIS21 Secretariat at OECD. A plan for a scaling-up the ADP to all African IDA countries has already been proposed. Funding has not been secured yet.

4/30/2010

More information is available at www.surveynetwork.org/adp.

Data Standards

To ensure quality, national and international statistics should follow well-defined and published standards. The World Bank is involved in several initiatives to improve and update data standards, such as work on external debt, SDMX, and sectoral data work in education and infrastructure described earlier.

Statistical Manual

The Statistical Manual brings together definitions and explanations of the relationships between concepts used in economics statistics such as the balance of payments, the national accounts, and government finance. The manual further discusses specific measures and methodologies used by the Bank, such as domestic savings and the Atlas methodology, for which there are no references in other international guidelines and manuals. Because the manual was written with the intention of covering statistical concepts most used by Bank economists, and because it is easily accessible, it is a resource for everyone in the Bank working with macroeconomic statistics, in particular those working with the Live Database.

Remittances

International remittances are an important source of finance in developing countries, especially for the poor. For a number of reasons – including the lack of clarity in the statistical definition of remittances and difficulty in recording remittances made through informal channels – the quality of remittances data needed to formulate policy options in this area, has been weak. Following a request from G7 Finance Ministers to the World Bank, DECDG has been leading an International Working Group on Improving Remittance Statistics, in close collaboration with the IMF and the UN. Working with existing international statistical forums wherever possible, such as the IMF Committee on Balance of Payments Statistics, improved definitions have been agreed, for reporting remittances within the Balance of Payments Framework. Interested Balance of Payments compilers from various countries have formed the “Luxembourg Group,” which has produced a comprehensive guide on estimating remittance flows, to help countries improve their estimates and their data sources. Recently, the Conference of European Statisticians also initiated a work plan to provide better guidance to countries on the use of household surveys to estimate remittance data. More information is available at econ.worldbank.org — search for “remittances”.

Atlas Method

The Atlas method is a procedure for adjusting exchange rates developed at the World Bank. The Atlas method is used to calculate gross national income per capita used in the Bank’s Operational Guidelines exercise, which is used in determining the Bank’s lending guidelines at the beginning of each fiscal year. The data are collected with the cooperation and input from country management units and regions.

System of National Accounts

The System of National Accounts is a comprehensive, consistent, and flexible set of macroeconomic accounts intended to meet the needs of government and private-sector analysts, policy makers and

4/30/2010

decision takers. The System of National Accounts 1993 (1993 SNA) was the joint responsibility of Eurostat, the IMF, the OECD, the United Nations and the World Bank. To maintain the validity of the System of National Accounts and to respond to changes in the economic environment, a joint team is developing an updated version of the 1993 SNA Revision 1. The World Bank is actively involved in the updating process and manages a trust fund to support the process.

External Debt Manual

The World Bank, along with the IMF and other international agencies, has produced a guide to external debt statistics for compilers and users. The guide provides comprehensive instructions for measurement and presentation of external debt statistics, replacing the old Debt Statistical Manual.

Statistical data and metadata exchange (SDMX)

The Bank for International Settlements, The European Central Bank, EUROSTAT, the IMF, OECD, the UN, and the World Bank have joined together to develop more efficient processes for the exchange and sharing of data and metadata through the use of e-standards. As part of this effort, the Bank is involved in a pilot project to rebuild the Joint External Debt Statistics using SDMX standards. These statistics are a product of the Inter-Agency Task Force on Finance Statistics. They bring together data currently compiled and published separately by the contributing international agencies, on components of countries' external debt and international reserve assets. This should facilitate timely and frequent access by a broad range of users to a single data set. These data are mostly from creditor and market sources but also include information provided by debtor countries using the Debtor Reporting System (DRS) of the World Bank.