Policies, Technology, and Quality Returns from the World Development Indicators

Biokou Mathieu Djayeola [1] | *The World Bank, Washington D.C., United States of America*
Tony Fujs [2] | *The World Bank, Washington D.C., United States of America*

Abstract

The World Bank's World Development Indicators are the primary collection of development statistics including economic, social, and environment indicators compiled from recognized international sources. It presents the most current and accurate global development data available, and includes relevant national, regional and global aggregated estimates for the purposes of analytics and policy making in international development arena.

In this paper, we review major policy decisions, key investments in technological platforms, and best practices in quality management around the reputation of the World Development Indicators. These findings are presented with the prospects of assessing how international statistical organizations model high demand statistical products, gather resources for their production, and establish subsequent quality control frameworks – including customer relations – needed for their sustainability.[3]

Keywords	JEL code
Quality management, statistical products, international development, technology, customers' relations	*L15*

INTRODUCTION – BACKGROUND

Quality in the World Development Indicators context is presented as a multidimensional concept: from the systems to the product through the process followed to collect, analyze, and disseminate data, the World Development Indicators puts significant efforts in developing the highest standards for this flagship publication. While assessing data quality as part of the product component, this paper presents different tools including software applications as inherent parts of the WDI production systems.

For the past forty years, the World Bank has made quality an integrated part of the dissemination of the World Development Indicators. It started as an annex of the 1978 World Development Report[4] with a couple of dozens of indicators on about 120 countries in its early editions.[5] Demand for data

---

and other related information grew over time leading to the design of a database with a metadata component through the Stars[6] System.

In 1996,[7] the traditional annex to the World Development Report was replaced by an independent publication. In the Open Data era, the Bank first made its own development data available for free but moreover fostered the concept of Open Data and Open Government to its member countries with technical assistance where needed.[8]

The 2017 online edition covers over 1 500 indicators on more than 200 economies for over 80 years. The World Development Indicators is popular because of its convenience and credibility. It receives about 1.2 million unique visitors per month. The dissemination of WDI would not be possible without the day-to-day collaboration with national and international statistical agencies. The production unit strives constantly to improve the quality of data on a regular basis.

1 DATA QUALITY STANDARDS AND THE WORLD DEVELOPMENT INDICATORS

The World Development Indicators applies the IMF DQAF as a major international framework on data quality. The DQAF was first introduced at the Fifth Review of the Fund's Data Standards Initiatives and serves as an umbrella for several dataset-specific frameworks.[9]

The IMF's Data Quality Assessment Framework (DQAF) was developed as an assessment methodology that aims to provide structure and a common language for the assessment of development data quality. It facilitates dialogue with national statistical agencies and country authorities, as well as a more homogenous approach to assessing data quality.[10]

In addition to adhering to DQAF principles, the World Development Indicators documents its data compilation methodology as well as issues related to data quality and effectiveness in relevant sections of its *knowledge base*. This resource also presents a wealth of information related to how the World Bank classifies world economies, national currencies, how often WDI is updated, metadata coverage, and WDI related products.

The World Bank policies towards data in general and the World Development Indicators have evolved over time. The following section will provide a set of institutional reforms that shaped the WDI ecosystem and cumulated to the current corporate product.

2 WDI DISSEMINATION PROCESS: PRACTICES, METHODOLOGY, AND QUALITY

The World Bank Development Data Group collects data from internal departments, national statistical agencies, and international sources for dissemination in the World Development Indicators. Content experts are responsible for the collection of a subset of sector statistics. The data is first processed in satellite databases before it converges to a working environment. Once assessed for quality and relevant operations, the data is moved to a production database where it is stored and published to various outlets. This allows the production team to validate the data at relevant stages in the ingestion process. The following graph provides a snapshot of the WDI data management process.

---

[5]   The 1981 World Development Report also includes the World Development Indicators, a set of 25 tables of economic and social indicators for 124 countries.
[6]   World Bank Socio-economic Time Series Access and Retrieval System (March & July 1990).
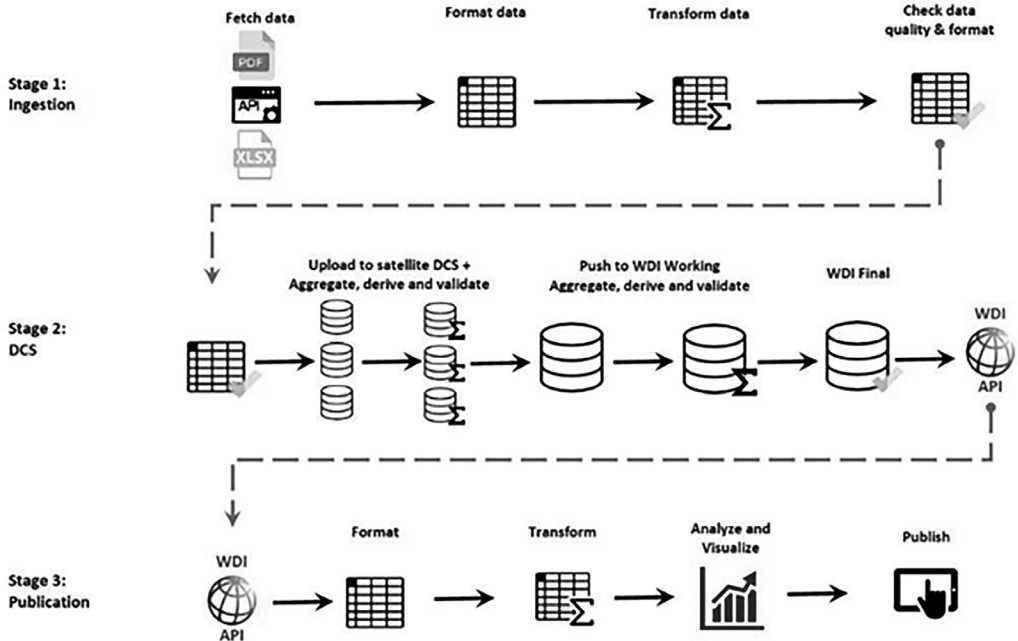[7]   World Bank, World Development Report 1996 – From Plan to Market, 1996.
[8]   World Bank Group, World Bank Support for Open Data 2012–2017, June 2017.
[9]   International Monetary Founds Data Quality Reference Site on the Dissemination Standards Bulletin Board.
[10] United Nations Statistics Division UNSTAT.

Figure 1 Description of WDI Work Flow

Source: Own construction, World Bank

In the following section, we to explore major principles governing official statistics and data quality in the international environment.

a) Principles governing statistical data quality

Through the World Development Indicators, the World Bank aims at providing high-quality data and data services by 1. establishing and maintaining the highest professional data standards, 2. supporting the improvement of national statistical systems, 3. actively participating in the international statistical community, and 4. developing and maintaining the tools to use data effectively through the quality principles of integrity, client focus, and fiscal responsibility.

These principles apply to all team members, regardless of terms of their employment, grade, line of business, or field of expertise. They are consistent with the Fundamental Principles of Official Statistics and the Principles Governing International Statistical Activities of the United Nations Statistical Division (UNSD).[11]

Moreover, in adherence to the Data Quality Assurance Framework, any indicator disseminated through the WDI output ought to be accurate, reliable, and accessible. Extensive metadata is provided to document methodology used to produce the indicator, limitations, as well as sources and revision policies.

b) Methods used to calculate aggregates for groups of countries

One of the key value added that make WDI unique in international development data arena is its well-maintained derivation and aggregation formulas and procedures that create new data out of what is

---

[11] Source: World Bank Data HelpDesk.

collected from sources. What are the methods and techniques used to calculate aggregates for country groups and categories?

In general, when indicators are ratios, aggregates are computed using weights corresponding to the denominator of the ratio. Other methods of aggregation used in the World Development Indicators are presented as follows:

- For group and world totals denoted in the indicator metadata as *gap-filled total*, missing data is imputed based on the relationship of the sum of available data to the total in the year of the previous estimate.
- For aggregates denoted in the indicator metadata as *sums*, missing values are not imputed. Sums are computed if less than a third of the observations in the series or a proxy for the series are missing in each year.
- For aggregates of ratios denoted as *weighted averages* of a ratios (using the value of the denominator or, in some cases, another indicator as a weight), the aggregate ratios are based on available data.
- For aggregates of ratios denoted as *unweighted averages*, the aggregate ratios are based on available data. Missing values are assumed to have the same average value as the available data.
- Aggregates denoted as *medians* in the indicator metadata are medians of the values shown in the table. No median aggregate value is shown if more than half the observations for countries with a population of more than one (1) million are missing.[12]
- When aggregating relevant indicators, one of the main requirements is to insure international comparability of national inputs. For instance, no aggregation shall be done for data in national currency. This must be converted to a common currency (e.g. USD) using exchange rate or purchasing power parities. In some cases, methodological adjustments are required before aggregation (e.g. Atlas method for Gross National Income).

3 TOOLS, SYSTEMS, AND INFORMATION TECHNOLOGY PLATFORM FOR QUALITY MANAGEMENT

In this chapter, we present various tools and techniques used to ensure quality data during collection, processing, and dissemination of World Development Indicators.

The World Bank **Data Collection System** is an in-house SQL system built to gather data from various sources for processing and dissemination purposes. The Data Collection System is organized into multiple databases. These are thematic (satellite) databases, one "work" or live database (WDI Working), and one dissemination database (WDI Final).

Each database has dedicated process(es) administered either by the satellite database owner or a central administrator. These processes are used to collect, view/edit, transfer, import/export, process (aggregate/derive), and publish the data. In addition to these functions, the central administrator is responsible for defining variables (countries, series, and time).

The WDI also leverages existing metadata standards to facilitate the data ingestion process. Select WDI structured data is collected using the **Statistical Data and Metadata eXchange** also known as **SDMX**. This platform stems from an international initiative aiming at standardizing and modernizing the mechanisms and processes for the exchange of statistical data and metadata among international organizations and their member countries. The World Bank is an official sponsor of the SDMX together with six other international organizations. The SDMX is used to collect official statistics in member countries of these organizations in different areas including agriculture, environment, social, economic, and financial statistics.[13]

---

[12] Source: World Bank Data HelpDesk.
[13] Source: Statistical Data and Metadata eXchange.

The Data Exchange works along other **specialized tools** to collect and prepare data for ingestion in the World Development Indicators. These tools include but are not limited to the External Debt Reporting System, the Joint External Debt Hub, and the International Comparison Program for purchasing power parities.[14] The World Bank Data Catalogue is an umbrella tool for the dissemination of various datasets including WDI and its interactive query tool known as DataBank. The main gateway to access WDI Data is the Open Data website for indicator level search, visualization, and extraction.

CONCLUSION – TOWARDS THE NEXT GENERATIONS OF THE WORLD DEVELOPMENT INDICATORS

While recognizing the importance of the World Development Indicators, with its high profile and exceptional exposure, it remains clear that this flagship product does not cover all the needs and/or every aspect of development data needs expressed by the users' community.

The team is leveraging recent technologies and data science techniques to automate the production of the World Development Indicators. It has recently migrated from its popular online data query tools to a mobile responsive web technology. What plans are foreseen for the next generation of the World Development Indicators?

On data quality side, the World Development Indicators processes focus mostly on ad-hoc data checking with abilities to perform select comparisons on a rather small subset of the whole database. Improving data quality in the World Development Indicators will require a comprehensive strategy with short, medium, and long-term perspectives.

In the *short term*, the team is working on plans to identify areas of concern and mitigate potential risks including data edit, sensitivity for identifying errors, and what discrepancies could be tolerated.

In the *medium term*, we are reducing the time needed to make the data available to users from when it is accessed at sources. We are leveraging on technology to improve the process while keeping the integrity of the data.

As indicated earlier, one of the greatest value added to original data received by the Bank through WDI is the aggregation of – and derivation from – select indicators. The team is working to determine the soundness of current methods and establish protocols for aggregation and indicators selection - which indicators to keep, which ones to drop, and what new indicators could be explored?

In the longer term, the data quality processes will consider expansion to two categories namely sub-national and high frequency time series.

Over the years, the World Bank operations have come across projects for which there was a need to disaggregate data within country boundaries. Multiple sectors were covered including environment, population and even economic activities. Some of the phenomenon being observed even have a time dimension making their structure a bit more complex than the current WDI records system.

Also, some areas of global development require data with frequencies higher than the annual data collected in WDI. Commodity prices, exchange rates, environmental measures (e.g. weather), and market indexes are crucial to knowledge about development matters, yet the current infrastructure of WDI is not covering those.

Moreover, other kinds of data are collected by the development community including the Bank and do not fit to any of these categories though they must be of equal importance. A significant amount of resources is devoted to collecting geospatial data, shapefiles, raster, maps as well as survey data of all kinds from single observations to big data and longitudinal datasets sometimes referred to as panel data, tracking the same sample at different points in time.

---

[14] Reference: World Bank Data Programs.

DISCLAIMER – ACKNOWLEDGEMENT

References

AUSTRALIAN BUREAU OF STATISTICS. *Data Quality Framework*. Canberra, 2009.

BAILEY, D. Why outliers are good for science? *Significance*, Royal Statistical Society, London, 2018, Vol. 15, Iss. 1.

EUROSTAT. *European Statistics Code of Practice.* Luxembourg, 2011.

EUROSTAT. *Quality and Performance Indicators.* Luxembourg, 2014.

U. S. GEOLOGICAL SURVEY. *Introduction to the Quality-Assurance Plan for Water-Quality Activities Workbook.* Reston VA, 2013.

INTERNATIONAL STANDARDS ORGANIZATION. *ISO 19011 Guidelines for auditing management systems.* Geneva, 2011.

OFFICE FOR NATIONAL STATISTICS. *Guidelines for Measuring Statistical Quality.* London, 2013.

WORLD BANK. *World Development Report (1978–2017).* Washington, DC.

WORLD BANK. *World Development Indicators (1996–2017).* Washington, DC.

STATISTICS CANADA. *Statistics Canada Quality Guidelines (2009).* Ottawa.