Framework for Assessing Indicators

Several international organizations and national statistical offices have produced data quality frameworks to ensure the availability of high-quality and relevant data for users. For example, the UK Office for National Statistics released A Government Data Quality Framework (UK ONS 2020) based on five principles: commit to data quality, understand user needs, assess quality throughout the data lifecycle, communicate data quality clearly and effectively, and anticipate changes affecting data quality. These principles are broadly consistent with other frameworks such as Eurostat (2017), the United Nations (2019), Statistics Canada (2017), the OECD (Organisation for Economic Cooperation and Development) (2011), Biemer (2010), and Jolliffe et al. (2023) (see Table A1 for a comparison of these frameworks).

The frameworks for data governance are divided into two distinct categories: one that provides guidelines for data producers to ensure the creation of high-quality outputs, and another that delineates the responsibilities of data providers to guarantee user access to data and metadata that satisfy their needs.[1] This note focuses on the latter, underscoring the importance of meeting user data requirements. In shaping the criteria for the World Development Indicators, a comprehensive set of factors was considered, including relevance, accuracy, coherence, clarity, comparability, completeness, confidentiality, timeliness, accessibility, and the extent of detail. These criteria are integral to ensuring that the data is not only of high quality and user-friendly but also come from a trusted source, have development relevance, and characterized by high geographical coverage. This framework for the WDI is presented in Table 1.[2] To be considered a good fit for inclusion in the WDI, an indicator should perform well across all four dimensions.

Table 1: Framework for Indicator Selection in the WDI: Adapted from Jolliffe et al. (2023)

Area	Dimension	Definition
Easy to Use	Accessible	Data is machine-readable and openly licensed, facilitating ease of access for users.
	Understandable	Data is accompanied by clear metadata, enhancing user comprehension.
	Interoperable	Data can be easily integrated with other datasets via common identifiers and standards.
Trusted & Relevant	Impartial	Data is unbiased, free from stakeholder influence that could compromise its integrity.

[1] The World Bank's Policy on Development Data Quality covers the policy for data producers at the World Bank to maintain high-quality products.

[2] "Trusted & Relevant" was used in place of "safe to use," because "safe to use," as expressed in Jolliffe et al. (2023), covered data other than the cross-country time series data found in the WDI, for instance survey data. For survey data, issues such as confidentiality are much more relevant than is the case of the WDI, which is based on country level data.

Area	Dimension	Definition
	Confidentiality Protected	Sensitive and personal data is securely protected against unauthorized access.
	Development Relevance	Data aligns with and supports internationally adopted development goals, priorities, and frameworks.
Adequate Coverage	Complete	Data comprehensively represents the target population or area of interest.
	Frequent	Data updates occur at frequent intervals, reflecting the dynamic nature of the information.
	Timely	Data is made available promptly following its collection or the occurrence of relevant events.
High Quality	Accurate	Data is precise, capturing the intended concepts with minimal error.
	Comparable	Data maintains consistency with standards, enabling comparison across geography and time.
	Granular	Data is sufficiently detailed, allowing for disaggregation where appropriate or necessary.
	Not Redundant	Data is unique and does not duplicate other available data, ensuring efficiency and clarity.

This structured approach ensures that indicators selected for the WDI are not only high in quality but also practical, reliable, and relevant for users' needs.

The framework outlined in Table 1 outlines the set of desired attributes for data to be fit for inclusion in the WDI. However, measuring whether an indicator meets all these criteria in practice can be challenging, as some attributes like impartiality or development relevance are difficult to pin down or specify unambiguously.

Table 2 presents a set of commonly available metrics that can serve as useful proxies to assess various aspects of the framework empirically. While these metrics do not capture the full extent of the framework, they provide a starting point for benchmarking key data attributes like coverage, timeliness, accessibility, and granularity. It is important to interpret these metrics as indicative rather than conclusive measures of whether an indicator meets the standards laid out in the conceptual framework.

Table 2: Metrics for Assessing Indicator Suitability for the WDI

Area	Qualitative Metrics	Quantitative Metrics
Easy to Use	Data is produced using transparent and clear methodology. Metadata is comprehensive and readily accessible.	Availability of an open license.

Area	Qualitative Metrics	Quantitative Metrics
Trusted & Relevant	Data aligns with and informs global development goals, such as the UN Sustainable Development Goals, World Bank objectives, and sector-specific priorities. Data production ensures confidentiality and impartiality. Data is sourced from reputable, established sources.	Engagement metrics, such as the number of unique visitors and frequency of citations.
Adequate Coverage	Data values are regularly updated and are expected to continue being updated. Data includes relevant subgroup details for comprehensive analysis.	Coverage metrics, including the number of economies covered, the proportion of low- and middle-income economies, the range of years data spans, the most recent year data is available, and the presence of non-missing data points.
High Quality	Data complies with domain-relevant international standards, when relevant, measures of accuracy or precision are considered. Consistency in methodology over time, ensuring comparability. Data is unique and does not duplicate other available data, ensuring efficiency and clarity.	

While it is uncommon for indicators to meet all the criteria completely, the World Development Indicators (WDI) team utilizes a scoring system to initially gauge an indicator's suitability. This system quantitatively assesses each indicator against the established criteria, providing a preliminary measure of its appropriateness for inclusion. However, this is just the first step. The team also carefully considers additional qualitative factors—those that are not as easily measured—to ensure a well-rounded evaluation. Together, these quantitative scores and qualitative assessments form a robust framework for deciding on the inclusion or exclusion of indicators in the WDI.[3]

Quantitative Criteria for Inclusion

Easy to Use

[3]For example, while adequate global coverage is good as an overall principle there are situations where issues are only encountered in certain regions and where monitoring and policy is a high priority. Additionally, new series may only have one or few years of data.

- **Metadata Availability**: Does the indicator include essential metadata? This encompasses a clear indicator name, description, definition, relevance to development, measurement units, statistical concepts, methodology, aggregation method, and data sources, as specified in annex Table A2.
- **Open License**: Is the data distributed under an open license, such as CC BY 4.0?

Trusted & Relevant

- **User Metrics**: What is the user interaction with the data over a year, including searches, browsing, downloads, or citations?

Adequate Coverage

- **Number of Economies**: For how many economies is data available for the indicator?
- **Percent of Low- and Middle-Income Economies**: What is the percentage of data coverage for low- and middle-income economies (LMICs), which are a focus of the World Bank's mission?
- **Span of Years**: What is the range of years for which data is available, and how many years does this span cover, as determined by the difference between the earliest and latest years with available data?
- **Timeliness of Data**: What is the timeliness of the data? This is measured using two metrics:
  - **Absolute Most Recent Year**: What is the latest year for which the indicator's data is available across any economy?
  - **Median Most Recent Year**: Across economies, what is the median of the most recent year of data available.
- **Periodicity of the Data**: How often are values in the time series available? This metric **(non-missing data, share)** evaluates the proportion of available data within the time span and country coverage previously calculated for the indicator, not the span and coverage of the WDI itself.

Qualitative Criteria for Inclusion

Evaluating the quality, relevance, accuracy, and suitability of a data source for inclusion in the WDI involves a careful and nuanced assessment of several factors. While quantitative criteria can be objectively measured, qualitative aspects often require the judgment of the WDI team. The World Development Indicators (WDI) team employs a qualitative assessment process that leverages the collective expertise of its members and the broader knowledge base within the World Bank. This process includes a thorough evaluation of several key factors:

1. **Data Provider's Reputation and Credibility**: The team scrutinizes the standing and reliability of the data source, ensuring that the provider is recognized for their integrity and the quality of their data. The team may consult with sector specialists at the World Bank.
2. **Methodological Transparency and Robustness**: The team examines the clarity and strength of the data collection process, and the methodologies applied, affirming that these practices meet high standards of transparency and robustness. Again, this may involve a consultation with sector specialists at the World Bank or an examination of published reviews.
3. **Alignment with Development Frameworks**: The data is evaluated for its consistency with globally recognized development frameworks, such as the Sustainable Development Goals (SDGs), the World Bank's mission and corporate objectives, and sector-specific goals and priorities, including those related to climate agreements.
4. **Not Redundant**: Data is unique and does not duplicate other data already included on the WDI, ensuring efficiency and clarity.

This qualitative assessment is underpinned by extensive background research and thoughtful discussions among WDI team members and World Bank domain experts. The evaluation capitalizes on their deep expertise and nuanced understanding of how indicators can be utilized effectively in various development contexts.

Moreover, the WDI team assesses the impartiality and suitability of the data to ensure it is free from bias and appropriate for its intended analytical purposes. The quality of the metadata is also a critical consideration, as it enhances the data's utility and interpretability. This comprehensive approach ensures that the data curated for the WDI is not only of high quality but also ethically sourced and relevant to the Bank's mission and the needs of its stakeholders.

The World Bank's World Development Indicators (WDI) team employs a framework that incorporates both quantitative and qualitative criteria to ensure the selection of the most appropriate indicators. The quantitative criteria provide objective measures of the data's accessibility, timeliness, and coverage, offering a clear view of its practical utility. The qualitative criteria, on the other hand, assess the data's relevance, accuracy, and reliability, ensuring that it aligns with the World Bank's mission and the broader development agenda. By integrating these criteria, the WDI team can make informed decisions, selecting indicators that best support the needs of World Bank clients and staff, researchers, policymakers, and development practitioners across the globe.