



**DATA QUALITY  
FRAMEWORK**

# ACKNOWLEDGMENTS

This document was prepared by Karen Carter (Statistics Specialist – Administrative Data) and Toby Wicks (Chief – Data Use Unit); in close partnership with other colleagues in the Data Use Unit, Data and Analytics Section, Division of Data, Analytics, Planning and Monitoring.

The report was edited by Lisa Hiller-Garvey and designed by Ane Louise Gaudert (independent consultant).

## Copyright

© United Nations Children’s Fund (UNICEF), Division of Data, Analytics, Planning and Monitoring, December 2021

Permission is required to reproduce any part of this publication. Permission will be freely granted to educational or non-profit organizations. To request permission or for any other information on this publication, please contact:

UNICEF Data and Analytics Section  
Division of Data, Analytics, Planning and Monitoring  
3 United Nations Plaza, New York, NY 10017, USA  
Email: [data@unicef.org](mailto:data@unicef.org)  
Website: [data.unicef.org](http://data.unicef.org)

All reasonable precautions have been taken by UNICEF to verify the information contained in this publication.

United Nations Children’s Fund, Data Quality Framework,  
UNICEF, New York, 2021.

## Photo credits

Cover photo: © UNICEF/UNI326700/Tesfaye  
Page 2: © UNICEF/UN0465351/Pancic  
Page 3: © UNICEF/UN0441216/Shing  
Page 6: © UNICEF/UN0470727/Akhter VII Photo  
Page 9: © UNICEF/UN0378214/Panjwani  
Page 17: © UNICEF/UNI372362/Pouget  
Page 22: © UNICEF/UN0370045/Mussapp  
Page 23: © UNICEF/UN0458524/Mussapp  
Page 24: © UNICEF/UN0421355/Abdul

# CONTENTS

<b>1. INTRODUCTION</b>	<b>2</b>
1.1. Purpose	2
1.2. Key terms	4
1.3. Data context at UNICEF	4
1.4. Content and scope	6
<b>2. UNDERLYING STATISTICAL PRINCIPLES</b>	<b>7</b>
<b>3. QUALITY DIMENSIONS</b>	<b>9</b>
3.1. Access to data (Accessibility)	11
3.2. Output quality dimensions	12
3.3. Process quality dimensions	18
3.4. Institutional quality dimensions	21
<b>4. FURTHER INFORMATION AND NEXT STEPS</b>	<b>23</b>
<b>5. ANNEX</b>	<b>24</b>

# 1. INTRODUCTION



## 1.1. Purpose

In 2018, UNICEF released the Data for Children Strategy<sup>1</sup> with the aim of strengthening the data culture across the organization. The strategy recognizes that data in the right hands, at the right time, means that decisions can be better informed, more equitable, and more likely to protect children's rights. Effective use of data can help us not only track results for children, but also shape those results with better insights regarding the impact of our programming and a better understanding of who is being left behind. Improving the use of existing data is an underpinning concept of the data strategy and there is strong demand from programme and field offices for capacity building and support in how to do this effectively. While this demand was already growing, driven in part by the broader data revolution,<sup>2</sup> the 2030 Development Agenda,<sup>3</sup> and the changing landscape of national data systems (including increased digitization, data linkage, and use of administrative data and 'big data' sources), it has been further highlighted by the Covid-19 pandemic<sup>4</sup> and the recent release of the UN Data Strategy.<sup>5</sup>

As data are increasingly seen as a global commodity and the demand for it has grown, so too has the politicization of data. As such, the data presented by international agencies are increasingly being examined and questioned. It is no longer reasonable to expect our publications and data sets to be trusted based purely on our brand reputation. We must be able to demonstrate why UNICEF should be recognized as a global leader and trusted source of data related to children and development.

1. UNICEF, 'Data for Children Strategic Framework', May 2019, accessed at <<https://data.unicef.org/resources/data-children-strategic-framework/>>.

2. Data Revolution Group, 'A World That Counts', accessed at <<https://www.undatarevolution.org/>>.

3. United Nations Department of Economic and Social Affairs, 'Transforming Our World: The 2030 agenda for sustainable development', September 2015, accessed at <<https://sdgs.un.org/2030agenda>>.

4. UN Sustainable Development Goals, 'Global Data Community's Response to Covid-19', October 2020, accessed at <<https://unstats.un.org/sdgs/hlg/Global-data-communitys-response-to-COVID-19/>>; CCSA, 'How Covid-19 Is Changing the World: A statistical perspective', 2020, accessed at <<https://unstats.un.org/unsd/ccsa/documents/covid19-report-ccsa.pdf>>

5. United Nations, 'Secretary-General's Data Strategy', accessed at <<https://www.un.org/en/content/datastrategy/index.shtml>>

UNICEF collates, analyzes and publishes a broad range of global data sets that support our efforts to improve the lives of children. We also support, both directly and indirectly, a broad range of national and specialized surveys and administrative data systems. In addition, we are increasingly working with partners to ‘close the gaps’ on data for children through new and innovative approaches including ‘frontier data’.<sup>6</sup> The effective use of these data requires that the people making decisions, both within and external to the organization, can not only access data that meet their needs when they require it, but also are able to make thoughtful judgements about the meaning of the data they have received. Our ability to interpret data in a meaningful way begins with our ability to assess the quality of that data systematically.

While many documents across the organization address data quality in relation to specific projects or data sets, drawing these data quality principles together into an overarching framework will support UNICEF’s ongoing efforts to produce and publish high quality data. It will also facilitate appropriate data use by decision makers (both within and external to the organization) to improve outcomes for children and ensure public confidence in our data products. The document is also intended to support a coherent approach to understanding data quality across the vast range of data types, sectors, and products that comprise our work to encourage collaboration and organizational learning. In doing so, the document intends to create:

- **A shared language** – on the characteristics we use to define ‘good’ or ‘appropriate’ data use;
- **A shared vision and set of standards** – that describes what ‘good’ looks like in our own work; and
- **Transparency around how we meet the standards that we have set for our data work** – including the processes we use to evaluate the data and statistics we produce, how we correct or respond to errors, and how others can query our work.

The creation of this framework also brings UNICEF into line with other international agencies working in the data space and national statistics offices that have already adopted formal data quality frameworks. The work draws heavily from existing examples from these organizations; in particular, the Generic Statistical Quality Assurance Framework (Generic SQAF) template produced by the UNSD,<sup>7</sup> and international organizations such as the OECD,<sup>8</sup> FAO,<sup>9</sup> and World Bank, and draws much of the text from the UNSD document.



6. Frontier Technologies, ‘Releasing the Power of Digital Data for Development: A guide to new opportunities’, accessed at <<http://bit.ly/frontierdatastudy>>

7. UNSD, ‘Generic Data Quality Assurance Framework for a UN Agency’, 2015, accessed at

<<https://unstats.un.org/unsd/unsystem/Documents-Sept2015/GSQAF-GenericData-Sept2015.pdf>>

8. OECD, ‘Quality Framework for OECD Statistical Activities’, OECD, Paris, France, 2011, accessed at <<https://www.oecd.org/sdd/qualityframeworkforoecdstatisticalactivities.htm>>

9. FAO, ‘The FAO Statistics Quality Assurance Framework’, 2014, accessed at <[www.fao.org/docrep/019/i3664e/i3664e.pdf](http://www.fao.org/docrep/019/i3664e/i3664e.pdf)>

## 1.2. Key terms

### Data:

1. Systematic information about the attributes of an entity (person, thing, event). Data of this type may be referred to as “micro”, “unit-record”, or “individual-level” data. This data may be quantitative or qualitative but must be defined with a description of what has been recorded, ensuring all data has a qualitative element.
2. Numerical information derived from aggregation of an individual data set. Also known as “aggregate” or “tabular” data.
3. Quantitative information in general – including estimates, indicators, and statistics of all kinds.

- adapted from the UN Demographic Glossary.<sup>10</sup>

**Data set:** Organized collection of data defined by a common or defined data structure.

adapted from the SDMX data dictionary.<sup>11</sup>

**Official statistics:** Official statistics are those disseminated by a national statistical system, excepting those explicitly stated not to be official. NB: Countries may have processes that nominate specific line ministries or agencies as the responsible entity authorized to release specific data sets or types.

adapted from the OECD data dictionary.<sup>12</sup>

**Statistics:** The general term applied to statistical processes and a statistical output.

**Statistical output** – numerical data relating to an aggregate of individuals or entities;

**Statistical process** – the science of collecting, analyzing, and interpreting such data. Systematic processing or treatment of a data set.

- adapted from the OECD data dictionary.<sup>13</sup>

## 1.3. Data context at UNICEF

UNICEF works in 190 countries and territories to improve children’s lives, to defend their rights, and to help them fulfil their potential from early childhood through adolescence. Grounded in the Convention of the Rights of the Child,<sup>14</sup> the organization works across a broad range of programmatic sectors including health; nutrition; social protection; education; and child protection, in both humanitarian and development settings. Our work includes direct service to communities; engagement with national governments to build and improve data systems; and collaboration with partners, including academia, international agencies, non-government organizations, and local service organizations.

10. UNData, UN Demographic Glossary, 2019, accessed at <<http://data.un.org/Glossary.aspx?q=data>>

11. SDMX, SDMX Glossary Version 2.0, 2018, accessed at <[https://sdmx.org/?sdmx\\_news=final-version-2-0-of-sdmx-glossary-now-available](https://sdmx.org/?sdmx_news=final-version-2-0-of-sdmx-glossary-now-available)>.

12. OECD, ‘Glossary of Statistical Terms’, 2004, accessed at <<https://stats.oecd.org/glossary/detail.asp?ID=4350>>.

13. OECD, ‘Glossary of Statistical Terms’, 2005, accessed at <<https://stats.oecd.org/glossary/detail.asp?ID=3847>>.

14. UNICEF, ‘Convention on the Rights of the Child’, accessed at <[www.unicef.org/child-rights-convention](http://www.unicef.org/child-rights-convention)>

Data are a central part of the development work at UNICEF and the Data for Children Strategy outlines five key principles around our data-related activities:

1. Data demand, supply, and use are equally important;
2. Data investments must support government data systems, not supplant them;
3. Effective data systems must function both within and across sectors;
4. Different data are appropriate for different uses and contexts;
5. Data for children is a team sport and working with partners is essential to create value.

The organization supports member states to collect, analyze, and use data to support development outcomes for children through several headline projects. These include Multiple Indicator Cluster Surveys (MICS); innovation projects, and a broad scope of engagement with administrative data systems including health information systems (HIS), education management information systems (EMIS), and birth registration among many others. UNICEF also engages directly with the operation of data systems, or collection of data on individuals and families, through cash-transfer programmes, humanitarian support, or specific research studies or collections. Our work in these fields may include technical support, software and system development, funding, and/or the direct collection of data. Tools such as U-report have been developed and are used extensively across the organization to collect data directly from young people – empowering and connecting young people around the world to engage with and speak out on issues that matter to them.<sup>15</sup>

UNICEF also collects and collates data from national governments (line ministries and NSOs), partner agencies, and other service organizations, and produces and disseminates statistical data on a range of key issues related to children. These data inform the organization's analytical reports, programmatic needs, and key advocacy campaigns. Data sharing with the organization occurs under the framework of both international

conventions and commitments and negotiated agreements with government and partners. The collected data cover many countries and span long periods. UNICEF adds value to the data it obtains from national governments and other organizations within the national statistical systems (NSS) in several ways. These include by editing the data and eliminating erroneous values; by harmonizing data across countries, time periods, sources, and methodologies; by adjusting for series breaks; and by estimating or modelling data for countries for which no data have been received, or the data received are of poor quality.

UNICEF disseminates statistics externally in four distinct ways. Statistical data are published:

1. in the form of databases that can be interrogated;
2. as free-standing statistical tables in electronic and printed publications;
3. in data visualization tools and dashboards;
4. as tables, graphs, and other figures embedded in reports and studies.

The organization is the custodian of seven global indicators under the Sustainable Development Goals and supports, or is 'co-custodian', for another 12.<sup>16</sup> Overall, the organization has identified 44 indicators that most directly concern children and will be the major focus of UNICEF's efforts to monitor and report on 'progress for every child' during the SDG era.<sup>17</sup> The current strategic plan remains closely aligned with the 2030 Sustainable Development Agenda envisaging a world in which all children are healthy and protected; live in a clean environment; learn; and have a fair chance to succeed. We are committed to supporting member states to nationalize the SDGs by setting ambitious national targets relating to children, establishing systems to track progress in reducing inequality, and ensuring that 'no child is left behind'. We will ensure a strong focus on the production, collation, and dissemination of high quality data and effective data use to achieve this end.

15. UNICEF, 'U-Report', June 2020, accessed at <[www.unicef.org/innovation/U-Report](http://www.unicef.org/innovation/U-Report)>

16. UNICEF, 'Using Data to Achieve the Sustainable Development Goals for Children', accessed at <<https://data.unicef.org/sdgs/>>

17. Starting in 2018, UNICEF used 35 indicators (or 44 if components of some of these indicators are counted) to assess progress towards child-related SDG targets. The same set of indicators was used in subsequent assessments in 2019, 2020, and 2021, <<https://data.unicef.org/resources/progress-for-every-child-2018/>>.

## 1.4. Content and scope



The framework includes the following major components:

- **Commonly agreed statistical principles and principles adopted by UNICEF;**
- **Agreed language and definitions of what we mean by quality and its dimensions;**
- **Guidance on what 'quality' looks like at UNICEF.**

This will subsequently be supported by the development of an overview of how data quality is ensured for the major data collections and systems at UNICEF. Additionally, a quality assessment programme will enable staff to evaluate their implementation of these standards and provide evidence for data users to give confidence in our work.

The term 'quality' is interpreted in this document in a broad sense, encompassing all aspects of how well statistical outputs fulfil user and stakeholder needs and expectations, and the efficiency and rigor of the processes used to obtain, collate, analyze, and present these data. Good quality outputs are statistics fit for purpose from the user perspective; more specifically, meaning they are relevant, accurate, reliable, coherent, timely, accessible, and interpretable enough. They are produced by processes relying on sound methodology and systems and are cost-effective within an institutional environment that recognizes the need for objectivity, impartiality, and transparency for statistical coordination.

The framework is intended to cover all statistical development and production activities within UNICEF, both at headquarters, and elsewhere, whether conducted by persons with a formal role of statistician or any other role, by employment status (fixed term versus contractor), and whether on a full-time or part-time basis. This includes programme data on or for children and families that is used to inform our programmes, evidence, or advocacy. It does not include human resource, or financial and ICT infrastructure management (operational data related to running the organization) as these are not specific to UNICEF's statistical activities. For the purposes of the framework, a statistical activity is one in which the primary focus is acquiring, processing, storing, analyzing, and disseminating statistical data. It does not include the use of statistics.

## 2. UNDERLYING STATISTICAL PRINCIPLES

The underlying Principles Governing International Statistical Activities, upon which the QAF is based, were formulated by the Committee for the Coordination of Statistical Activities and endorsed by chief statisticians/coordinators of statistical activities of UN agencies in 2005.<sup>18</sup> They build on the fundamental principles of official statistics and are as follows:

1. High quality international statistics, accessible for all, are a fundamental element of global information systems
2. To maintain the trust in international statistics, their production is to be impartial and strictly based on the highest professional standards
3. The public has a right to be informed about the mandates for the statistical work of the organizations
4. Concepts, definitions, classifications, sources, methods and procedures employed in the production of international statistics are chosen to meet professional scientific standards and are made transparent for the users
5. Sources and methods for data collection are appropriately chosen to ensure timeliness and other aspects of quality, to be cost-efficient and to minimize the reporting burden for data providers
6. Individual data collected about natural persons and legal entities, or about small aggregates that are subject to national confidentiality rules, are to be kept strictly confidential and are to be used exclusively for statistical purposes or for purposes mandated by legislation

7. Erroneous interpretation and misuse of statistics are to be immediately appropriately addressed
8. Standards for national and international statistics are to be developed on the basis of sound professional criteria, while also meeting the test of practical utility and feasibility
9. Coordination of international statistical programmes is essential to strengthen the quality, coherence and governance of international statistics, and avoiding duplication of work
10. Bilateral and multilateral cooperation in statistics contribute to the professional growth of the statisticians involved and to the improvement of statistics in the organizations and in countries

In addition to the principles adopted by the UN Statistics Commission, UNICEF – as an organization grounded in a rights-based approach for children – has adopted the principles of the document ‘A Human Rights Approach to Data’, published by the UN Office of the High Commissioner for Human Rights.<sup>19</sup>

These principles are as follows:

1. **Participation:** Participation of relevant population groups in data collection exercises, including planning, data collection, dissemination, and analysis of data.
2. **Data disaggregation:** Disaggregation of data allows data users to compare population groups, and to understand the situations of specific groups. Disaggregation requires that data on relevant characteristics are collected.

18. UNSD, ‘Principles Governing International Statistical Activities’, Items for Discussion, Committee for the Coordination of Statistical Activities, 4-6 September 2013, <<https://unstats.un.org/unsd/accsub/2013docs-22nd/SA-2013-8-FP-UNSD.pdf>>.

19. United Nations Human Rights, ‘A Human Rights-Based Approach to Data: Leaving no-one behind in the 2030 agenda for sustainable development’, 2018, accessed at [www.ohchr.org/Documents/Issues/HRIndicators/GuidanceNoteonApproachtoData.pdf](http://www.ohchr.org/Documents/Issues/HRIndicators/GuidanceNoteonApproachtoData.pdf)

3. **Self-identification:** For the purposes of data collection, populations of interest should be self-defining. Individuals should have the option to disclose or withhold information about their personal characteristics.
4. **Transparency:** Data collectors should provide clear, openly accessible information about their operations, including research design and data collection methodology. Data collected by state agencies should be openly accessible to the public.
5. **Privacy:** Data disclosed to data collectors should be protected and kept private, and confidentiality of individuals' responses and personal information should be maintained.
6. **Accountability:** Data collectors are accountable for upholding human rights in their operations, and data should be used to hold states and other actors to account on human rights issues.

Considering the specific interests of children, these broad principles have been reflected in the Responsible Data for Children Principles<sup>20</sup> developed jointly by UNICEF and New York GovLAB, as presented below. These are intended to provide a practical foundation for working with data in the field and should be upheld across all our data work related to children.

For UNICEF, this data quality framework and the integration of key quality dimensions across our data work are an important mechanism for upholding these international principles and continuing to ensure the data we produce is trusted to support better decision-making and therefore, better outcomes for children.

PRINCIPLE	DESCRIPTION
<b>Protective of children's rights</b>	Recognizing the distinct rights and requirements for helping children develop to their full potential.
<b>Purpose-driven</b>	Identifying and specifying why the data is needed and how the intended or potential benefits relate to improving children's lives.
<b>Proportional</b>	Aligning the breadth of data collection and duration of data retention with the intended purpose.
<b>Professionally accountable</b>	Operationalizing responsible data practices and principles by establishing institutional processes, roles, and responsibilities.
<b>Prevention of harms across the life cycle</b>	Establishing end-to-end data responsibility by assessing risks during the collecting, storing, preparing, sharing, analyzing, and using stages of the data life cycle.
<b>Participatory</b>	Engaging and informing individuals and groups affected by the use of data for and about children.
<b>People-centric</b>	Ensuring the needs and expectations of children, their caregivers, and their communities are prioritized by actors handling data for and about them.

20. Responsible Data for Children, 'The RD4C Principles', accessed at [www.rd4c.org](http://www.rd4c.org)

# 3. QUALITY DIMENSIONS

While statistical product quality can be summarized in line with the definition in the ISO 9000 Series<sup>21</sup> for any product as fitness for use, there is a need to elaborate this definition in terms of the various quality aspects or dimensions. The data quality dimensions presented here have been adapted from the UN Statistics template for UN agencies<sup>22</sup> to address UNICEF's specific data uses and engagement and draws heavily from this text.

Quality dimensions, as they are used at UNICEF, can be considered in four groups:

- **Access to the data;**
- **Output quality;**
- **Process quality; and**
- **Institutional quality.**

These last two groups underpin output quality through well designed and executed processes within an appropriate institutional setting. Attention to each of these categories is essential in supporting the generation of data trusted in the decision-making process, and which can effectively inform decisions around data that is appropriate for use.



21. ISO, 'Quality Management Systems: Fundamentals and vocabulary', accessed at [www.iso.org/obp/ui/#iso:std:iso:9000:ed-3:v1:en](http://www.iso.org/obp/ui/#iso:std:iso:9000:ed-3:v1:en)

22. UNSD, 'Generic Data Quality Assurance Framework for a UN Agency', 2015, accessed at <https://unstats.un.org/unsd/unsystem/Documents-Sept2015/GSQAF-GenericData-Sept2015.pdf>

FIGURE 1: UNICEF dimensions of data quality

AIM	CATEGORIES	WHAT DO WE ACTUALLY NEED TO KNOW?	DIMENSIONS OF QUALITY
DATA THAT ARE TRUSTED AND USEFUL	ACCESS	Can I find the data I need easily?	Accessibility Target audience Publication platforms Access policy and procedures
	OUTPUT QUALITY	Do the data measure what I need/ what they say they do?	Relevance Concepts and content Coverage Inclusion
		How accurate are the data?	Accuracy Completeness Precision Reliability Consistency
		Are the data timely?	Timeliness Time since reference Punctuality Frequency
		Do the data make sense?	Interpretability Metadata and documentation Comparability
	PROCESS QUALITY	Can I see how the data were obtained?	Appropriate sources Metadata and documentation
		Does the data handling process minimise/ check for data handling errors? Do the methods reflect best practice?	Sound methods and systems Ethical data practice Sustainability/cost-efficiency
		Are there processes to keep up to date with best practice?	Review procedures
	INSTITUTIONAL QUALITY	Is the organization impartial and objective?	Impartiality and objectivity
		Is the organization transparent and credible?	Review procedures Data quality framework Statistical coordination

## 3.1. Access to data (Accessibility)

### Can I easily find the data I need?

Often presented as simply a dimension of output quality, access to the data (or accessibility) is worth considering as a category in its own right as a primary driver of whether data are useful. The accessibility of a statistical output reflects how readily the data can be discovered, located, and accessed from within the UNICEF data holdings. It includes the suitability of formats in which the data are available; the media of dissemination; the availability of metadata and user support services; and, if there is a charge, the affordability of the data to users.

From the perspective of data availability, UNICEF users are divided into three distinct groups: internal users; partner governments, agencies, and field partners; and external users (the general public). Internal users may be able to access data earlier, in more detail or in draft form, than other groups and may therefore have a different experience of accessibility of data than the other two groups. For example, unit record data may be available to internal users but may only be released to partners or the public as aggregate data or through a formal release process.

### What does this mean for UNICEF?

Target audience: For each major data collection, statistical product or system, UNICEF should have clearly documented the target audiences, the type of data that should be made available to each of these three groups, and the reasoning behind these decisions. Decisions on what data should be available to each group should be driven by the principles outlined in Section 2 – with a clear focus on privacy and protection of individual or sensitive data,<sup>23</sup> while supporting transparent and open data publication where possible. In identifying the target audience, specific consideration should be given to the population from which the data itself was collected.

Publication platform(s): Data products should be readily accessible to the target audience via platforms that make sense for the context in which users are operating. This may mean searchable websites, data exchange platforms, or published print reports (or a combination of these and other media) that are readily available to the target audience at the time the data is needed. This should also include removing barriers to access where possible. This may mean, for example, making published reports available without cost, limiting log-in requirements to public data sets (unless specifically required for access to more sensitive data), and adopting international data standards to maximize the ability to share data across reporting and access platforms. Navigation and search functions should also be structured to simplify the ability to query and retrieve data.

Key data collections and publications from UNICEF are routinely published on the data.unicef.org site, supported by the UNICEF data warehouse,<sup>24</sup> which is searchable by data type, topic, country, and other key fields.

Access policy and procedures: An essential part of making data accessible is ensuring visibility for key distribution platforms or publications, and a publication policy should be articulated and made publicly known. There should also be a clearly defined policy and/or procedure for requesting access to more sensitive data where underlying data sets have not been made available to all user groups. These procedures, and the criteria on which a decision to release data will be based (in accordance with the principles noted in Section 2), should be readily available to all user groups. It should be noted that in the context in which UNICEF works, some of the data are held or used with the permission of national governments or partner organizations and requests to release more detailed data may need to be re-directed to the appropriate party.

23. In line with UNICEF's 'Policy on Personal Data Protection', accessed at [www.unicef.org/supply/documents/unicef-policy-personal-data-protection](http://www.unicef.org/supply/documents/unicef-policy-personal-data-protection)

24. UNICEF, Data Warehouse, accessed at [https://data.unicef.org/dv\\_index/](https://data.unicef.org/dv_index/)

Requests to access data should be considered in line with the Responsible Data for Children Principles (in particular, purpose and participation), and comply with internal procedures such as required non-disclosure agreements for facilitating access to government data<sup>25</sup> and UNICEF's global data protection policy.<sup>26</sup>

Whenever possible, requests for data by national governments should be coordinated with other UN agencies, funds, and programmes to avoid a large burden on national counterparts.

## 3.2. Output quality dimensions

### Relevance: Do the data measure what I need?

The relevance of a statistical output is the degree to which the data serve to address the purposes for which they are sought by users, or which they are presented as addressing. While relevance is highly subjective, this dimension asks researchers, system managers, and data collectors to consider whether the item being measured accurately represents the question that is being asked. Where possible, this should be supported by evidence to validate the selection of the measurement used.

Typically, a data product has multiple users and intended uses. Thus, measuring relevance requires the identification of user groups and their needs, while recognizing that these may change over time.

Users of UNICEF data may be divided into three main groups:

- **Internal users** – programme managers or colleagues within the UNICEF structure who require data to inform programming decisions and monitor impacts;
- **Partner governments, agencies, and field partners** – including other international organizations, national governments, national statistical offices and other national organizations, non-government organizations, private sector firms, or consultants working collaboratively with UNICEF<sup>27</sup> to develop policies, or plan programmes or interventions;
- **External users** – including academic institutions, media, donors, and members of the general public, etc.

Among these groups, individual key users should be identified and given whatever special consideration is appropriate according to their status. This should specifically include the community from which the data is purported to represent. The content and format of published outputs should be adapted to the full range of potential users.

### What does this mean for UNICEF?

Relevant data is therefore based on a plausible or appropriate premise or measure (content and concept); collected from the population group it claims to represent; and reflects the diversity of this population group adequately – including sub-populations or specific groups that may be hidden or under-counted in routine data collection.

Relevant concepts and content: This is essentially about definition of the data items and how these are applied, with the aim of measuring the factor, characteristic, or outcome that we are seeking to report on as closely as possible. One example would be the time period covered (for example, for a calendar year, financial year, or some other period), and how this relates to the time period the data is intended to measure. The plausibility and specificity of association between the actual measure and what we are trying to represent should also be considered.<sup>28</sup>

25. UNICEF, Internal Document, accessed at <[https://unicef.sharepoint.com/sites/SD/SitePages/Launch-of-Non-Disclosure-Agreement-\(NDA\)-in-VISION-for-Procurement-Documents.aspx](https://unicef.sharepoint.com/sites/SD/SitePages/Launch-of-Non-Disclosure-Agreement-(NDA)-in-VISION-for-Procurement-Documents.aspx)>

26. UNICEF, 'Policy on Personal Data Protection', accessed at [www.unicef.org/supply/documents/unicef-policy-personal-data-protection](http://www.unicef.org/supply/documents/unicef-policy-personal-data-protection)

27. Usually through a formalized agreement such as a contract or data sharing arrangement, in the case of private firms, contractors, or NGOs.

28. Concepts taken from the Bradford-Hill Criteria for assessing causality: Hill, A.B., 'The Environment and Disease: Association or causation?' Proceedings of the Royal Society of Medicine, No. 58, 1965, pp. 295-300, accessed at <[https://yosemite.epa.gov/oa/rhc/epaadmin.nsf/Filings/4FA30B929E8DE57B85257A5A005A17EC/\\$File/ATTULNMQ.pdf](https://yosemite.epa.gov/oa/rhc/epaadmin.nsf/Filings/4FA30B929E8DE57B85257A5A005A17EC/$File/ATTULNMQ.pdf)>

This concept of relevant content is particularly significant when using proxy indicators (where the item of interest cannot be measured directly), and compound indexes (such as those used in early childhood development, or the human development index, which weights disparate measures to create a single indicator). Special consideration is also necessary when using modelling, big data, and/or machine learning to draw inferences from indirect sources.

As relevance is directly linked to how the data are intended to be used, there should also be processes in place to periodically determine the views of users and the uses they make of the data.

**Coverage:** This refers to the target population for which the data system or collection is actually designed to capture data. For example, this may be national or subnational, it may be defined by specific age groups, or by any other specific defined characteristics (usually demographic).

Relevance is therefore a function of how closely the coverage of the data set matches the population for whom we are trying to answer a specific question. For example, if we want to know about literacy in adolescents, we may consider a data set that captures literacy for children in high school (the coverage of the system) as closely relevant, depending on attendance rates. However, there may be discrepancy between the true literacy rates of all adolescents and those at school. Documentation should include discussion of who the data represent, who is excluded, and whether there are any impacts or biases caused by exclusion of particular people, areas, or groups.<sup>29</sup> Coverage should not be confused with completeness, which measures how well the data set actually captures the population it seeks to cover.

**Inclusion (equity, or representativeness):** This dimension is essentially a sub-set of coverage but is of particular interest to UNICEF which has a strong mandate to 'leave no child behind'. While coverage is an indication of who the data system or collection aims to include, inclusion asks data managers to specifically consider who may be 'invisible' in the data collection and what measures can or should be put in place to ensure that data sets reflect the whole of our target audience. In statistical terms, this is called 'selection bias', and the potential for certain sub-groups to be under-represented in the data set should be noted in the accompanying metadata.

## Accuracy: How accurate are the data?

The accuracy of a statistical output is the degree to which the data correctly estimate or describe the quantities or characteristics they are designed to measure. Accuracy refers to the closeness between the values provided in the product and the (unknown) true values (or error). Error in statistical estimates is traditionally decomposed into bias (systematic error) and variance (random error) components.

In the case of data from sample surveys, the major sources of error are coverage, sampling, non-response, response, processing, and seasonal adjustment. For data from censuses there are no sampling errors. For data from administrative sources there are also no sampling errors, but there are additional problems due to mismatching of administrative concepts and/or classifications to statistical requirements, while biases due to incomplete registration or barriers to services are common.

### What does this mean for UNICEF?

The accuracy of the data produced by UNICEF is largely determined by the volume and accuracy of the data received from the contributing organizations within countries. However, UNICEF activities can also improve accuracy. For example, quality checks may result in detection and correction of errors in data provided by countries, thus leading to improvements in the data. Conversely, UNICEF activities may also have an adverse effect, for example by introducing errors during the processing stages. UNICEF may estimate data for countries for which data are not provided. Such estimates typically depend upon statistical models and thus are subject to model error, i.e., the difference between the model-based values and the true (but unmeasured) values.

As it is often not possible to know the 'true' value of most statistics, how 'accurate' our data is, or how close we are to that truth, is assessed in terms of the following sub-dimensions: the completeness of the data set; the precision of the result; the reliability (or repeatability) of the result; and the coherence of the data.

29. Australian Bureau of Statistics, 'The ABS Data Quality Framework', 2010, accessed at <[www.abs.gov.au/websitedbs/D3310114.nsf/home/Quality:+The+ABS+Data+Quality+Framework](http://www.abs.gov.au/websitedbs/D3310114.nsf/home/Quality:+The+ABS+Data+Quality+Framework)>

Completeness: This is a measure of how many of the events or individuals (the 'records') in our target population were actually documented. For example, in a national birth registration system, completeness is the proportion of all births (that should be recorded in the system) actually registered. In a sample survey, completeness would be the equivalent of the response rate.

Completeness may be measured for a collection tools as a whole (whether a registration form was completed or survey instrument filled out) or for individual data elements within the collection.

The completeness of individual data items may be influenced by many factors including, for example, question design, sensitivity of the information being requested, and the actual layout of the data collection tool.

There are no absolute rules around how complete data needs to be in order to be considered accurate, however care should be taken to avoid reporting statistics, or indeed creating data visualizations derived from data where the completeness is low. This is likely to result in systematic biases that will make the data less representative of the population and therefore, less likely to accurately reflect the 'truth'. It is good practice to ensure that completeness of the data set (and key data items, if appropriate) is assessed and clearly reported in the metadata, and to ensure that highly incomplete data is not misrepresented by being reported without this potential bias being clearly noted.

Precision denotes how certain we are of the final values; that is, how sure we are that we have reported the 'true' value. It is generally indicated by how wide the confidence interval is around the data estimate or statistic (a function of variance), or how wide the range of values is that we have determined. The method of data collection and analysis (including measurement instruments), and the sampling size can both have a major impact on data precision.

It is good practice to present 95 per cent confidence intervals or variance with all key statistics where relevant, or to ensure these can be readily sourced if a data user requires it.

Reliability (or reproducibility): Reliability refers to how stable the result would be if the study or collection were repeated in the same population around the same time using the same method. It can also indicate the closeness of the initially released values of a statistical output to the values that are subsequently released for the same reference period (through a subsequent revision or update due to additional data collection or correction of erroneous values). A reliable value is one that is not subsequently likely to change when re-estimated, whether it is accurate or not.

Consistency: Often referred to as coherence, consistency in its loosest sense implies the data are 'at least reconcilable'. For example, if two output purporting to cover the same phenomenon differ, the differences in time of recording, valuation, and coverage should be identified so that the outputs can be reconciled. Different methods will always yield somewhat different results due to the different timeframes, definitions, and collection mechanisms (as well as random variation). However, similar results for the same data item or statistic from different methods (such as calculation of infant mortality from a household survey like MICS and from civil registration data) can be used to triangulate data and provide additional evidence to support that your data or estimates are accurate.

Coherence within a dataset implies that the elementary data items are based on compatible concepts, definitions, and classifications. Incoherency within a dataset occurs, for example, when component values that should add up to a total do not.

## Timeliness: Are the data timely?

Data are only useful if they are available to decision makers at the time that they are making decisions, and if those data refer to a time period that can be used to make judgements about progress or the best course of action in the present. Timeliness is assessed in terms of a time scale dependent upon the period for which the data are of value for making decisions. The concept applies equally to short-term (sub-annual) or structural (annual or longer period) indicators. In effect, this means that the same data set or source can be timely for some uses (that is, have value to the decisions being made) but not for others. The timeliness of the data is therefore interwoven with the intended data use. For example, if delivery of a nutrition programme in a development context is reviewed only quarterly – this is the fastest period in which services and budgets would be realistically amended – for data to be timely, it would need to be available quarterly and should refer to a reference period not more than a few months old. Data would not need to be ‘daily’ in order to be timely in this circumstance. However, in a humanitarian setting with an acute malnutrition crisis, this may not be considered timely and more frequent data reporting may be necessary.

Although UNICEF processes have an impact on timeliness, for the most part the timeliness of UNICEF outputs is determined by the timeliness of the data it receives from contributing organizations within countries.

There is generally a trade-off between timeliness and accuracy of the data – and the importance of time versus accuracy will depend on what the data is being used for. For example, official statistics for international monitoring and reporting may require a greater level of accuracy (and therefore require more time to prepare) than data used to inform a humanitarian response, where there is greater need to balance the speed at which data can be produced against accuracy that is ‘good enough’ to support action in the field (such as providing estimated affected populations, etc.).

## What does this mean for UNICEF?

Two related concepts are used to understand the timeliness of data at UNICEF:

**Timeliness** (time since the reference period): The timeliness of a data product or a statistical output is determined by the length of time between its availability and the event or phenomenon it describes (the ‘reference period’), and whether the data is received soon enough after the reference period to be useful.

The reference period should be clearly indicated in all final data sets and statistics and efforts made to align data availability with key decision-making processes (such as the development of a strategic plan, country plan cycles, or budget reviews) and uses (i.e., daily reporting during cluster activation in an emergency setting).

**Punctuality**: As punctuality and timeliness are inter-related, punctuality is often grouped with timeliness as a single quality dimension in the QAFs of many organizations. Here they are kept separate as they are achieved by quite distinct means. The punctuality of a statistical output implies the existence of and adherence to an output dissemination schedule. An output is punctual if it is disseminated in accordance with the schedule. In the case of an output published externally, the schedule may comprise a target release date or may involve a commitment to release data within a prescribed time period.

## A dissemination schedule assists:

- internal users, by enhancing their capacity to plan their work based on target internal dissemination dates for data they require;
- partners and external users, by improving their capacity to make timely use of the UNICEF statistics;

There may be occasions when UNICEF simply cannot adhere to the dissemination schedule due to the late acquisition of data from input sources. In such circumstances, a warning regarding the delay in dissemination should be communicated to users.

**Frequency:** Punctuality is closely related to the concept of frequency. While some data is needed only periodically or as a 'one off' to provide a 'snapshot' of a situation to guide policy or response, data for monitoring and programming is required regularly – whether or not this is formally documented in a publication schedule. Frequency of data collection, collation, and publication should be tied as closely as possible with planning mechanisms and cycles at the country level, national office, and organizational level.

## Interpretability: Do the data make sense?

The interpretability (sometimes called clarity) of a statistical output reflects the ease with which users can understand and properly use the data.

### What does this mean for UNICEF?

Metadata and documentation: The degree of interpretability is largely determined by the adequacy of the metadata that accompany the data, including definitions of concepts, target populations, indicators, and other terminology describing the output and its limitations. That allows users to draw appropriate judgements on how data can and should be used while data visualization and query tools interrogate and unpack the data for specific uses.

Returning to the earlier concept of different user target groups, the interpretability of data is largely driven by ensuring an appropriate match between data users and the level of detail provided, in both the data itself and the accompanying metadata. This supports data users to draw appropriate inferences from the data and statistics provided. This may mean adapting the content and format of published outputs for a specific audience and providing increasing layers of detail. Major collections and publications should have, at minimum, explanatory output suitable for use with the communities from which the data was derived and with national government (and non-sectoral specialists), as well as specialist or technical papers on methodology, where appropriate. For example, UNICEF situation indicators are not only publicly available using standard machine-readable open data interfaces that eliminate ambiguity (e.g., [CL UNICEF INDICATORS](#)), they are also fully described in [indicator profiles](#) to maximize their transparency and lineage.

**Comparability:** One of the primary influences on the interpretability of data is the comparability of the data over time and between countries or settings. This is essential to enable benchmarking against good practice or to track progress of development indicators over time. Sometimes referred to as 'coherence'<sup>30</sup> or 'harmonization' when applied over time, the ability to compare data relies on agreed data standards and coding, and on metadata that clearly documents changes that may affect the comparability of data over time or between countries.

For UNICEF, ensuring comparability over time means the adoption of international standards or coding where these exist, with SDMX adopted as the international standard data format for aggregate values wherever possible. For some data collections, modelling is also used to 'smooth' data and minimize the impact of different data standards and approaches across countries and time for comparative purposes. For all collections, metadata should indicate any appropriate standards at national or international level and indicate any adjustments that have been made for the sake of comparability. Variations in methodology that might affect data values should not be made without explanation.

30. We have elected not to use the term 'coherence' as this can also relate to the consistency of data within a data set, and between different measures that cannot be directly compared but which related to the same source.

## Balancing output quality dimensions

The data quality dimensions are not interdependent in the sense that there are relationships between the factors contributing to them. Factors leading to improvements with respect to one dimension may result in deterioration with respect to another. Thus, in designing a statistical process and its outputs, it is often necessary to trade-off quality in one dimension with quality in another. The relative importance of each of these dimensions will vary with intended use and audience for the data. The most significant trade-offs to consider are as follows.

- **Accuracy and timeliness:** This is probably the most frequently occurring and important quality trade-off. Improvement in timeliness can be obtained by terminating data acquisition earlier and compiling outputs based on a smaller number of countries and/or reduced editing. However, as this reduces accuracy, there needs to be a trade-off. For major outputs, a compromise is to disseminate a preliminary version of the output based on partial acquisition and then one or two revisions based on successively more acquisition and editing. The difference between preliminary and revised versions is an indicator of the degree of accuracy that is being sacrificed to produce the increased timeliness.
- **Relevance and accuracy:** Relevance can be increased by acquiring more data items, but accuracy may be diminished because the additional data are less accurately reported. Conversely, deciding not to disseminate inaccurate data items increases accuracy but reduces relevance.
- **Relevance and timeliness:** Timeliness may be improved by reducing the number of data items collected or by replacing those that are difficult to collect with ones that are easier to collect. In either case, there is a negative effect on relevance.
- **Relevance and coherence:** Improvements in relevance – for example, by redefining the indicators in the light of a better understanding of user needs or moving to a later version of a classification – reduces comparability over time, perhaps to the point of requiring a series break. Conversely, the desire to retain comparability over time tends to inhibit changes in content that would improve relevance.
- **Accuracy and coherence:** Improved methods may increase accuracy but reduce coherence by introducing changes in data that are attributable to changes in methods not in what is being measured. Conversely, the desire to retain coherence may inhibit changes required to improve accuracy.
- **Timeliness and punctuality:** Having a publication schedule and adhering to it reduces timeliness if the data are available earlier than the due release date.
- **Accuracy and reliability:** If values are not revised in the light of new or better observation data, they are less accurate than they would be if revisions were to take place. On the other hand, if values are revised, reliability is reduced.



## 3.3. Process quality dimensions

Output quality is directly affected by process quality and the measures in place to identify the appropriate data sources, and collate and process this data and how it is presented.

### How were the data obtained?

The selection of appropriate data sources and clear documentation of where the data has come from is the foundation of process quality.

#### What this means for UNICEF:

Appropriate data sources: Appropriate data sources are not absolute and depend greatly on what is available in a given context. Some data sources, however, are better suited for specific uses than others. For example, it is difficult to obtain behavioural data through a routine administrative data system; this information is best collected through a survey process or smaller focus group study, if results do not need to be representative at the population level. On the other hand, monitoring service delivery or national development outcomes where data is needed, either annually or more frequently, is best sourced from a well-functioning administrative data system where this is available due to the infrequent nature of most surveys. For UNICEF this means being able to clearly articulate 'what data when'. That is, what data is needed for what purpose – and matching that back to both the available sources and where we need to be investing in strengthening data collection or systems.

In many circumstances, multiple data sources are preferred in order to triangulate results and address the limitations of using a single data source where the inherent biases are less easily identified or measured.

Metadata – documentation of data sources: In addition to selecting (and investing in) appropriate data sources, it is essential that data sources are clearly documented and presented with core data sets and statistics. This should also include some assessment of the potential biases associated with the data source used and any steps taken to minimize their impact, or guidance as to where to seek this information. This is particularly relevant for data sourced from third parties – such as national governments – where UNICEF does not have direct influence on the design and implementation of the data collection. This, for example, may mean referencing a national administrative data system assessment (such as an evaluation of a national Education Information Management System), or survey design documents (in the case of a MICS survey).<sup>31</sup>

### Does the data handling process minimize/check for data handling errors? Do the methods reflect best practice?

Data quality is very much a function of the life cycle of the data, from collection through to use and potential revision. As such, the way data is handled should reflect deliberate choices designed to ensure output quality (noting the balance between the various output dimensions described earlier, and the need to make choices as to the appropriate balance based on intended use), and to minimize additional errors or uncertainty from the data handling itself. A life cycle approach to data and best practice also explicitly implies a focus on the sustainability of critical data collections and a rights-based ethical approach to data.

31. UNICEF, 'MICS6 Tools', accessed at <https://mics.unicef.org/tools#survey-design>

## What this means for UNICEF:

Sound methods refers to the use of international standards and best practice through all stages of a statistical process, from identification of requirements, through design, data collection, processing, analysis, dissemination, and evaluation, as well as the clear documentation of these standards in readily accessible metadata. Application of standards and best practice not only engenders process and product quality, it also fosters comparability across countries and international organizations. For UNICEF, this means staying abreast of the evidence around key indicators, measures, and methods for both collection and analysis, and addressing gaps in the evidence when these relate to key indicators for which UNICEF is the custodian. UNICEF recognizes SDMX as the preferred UN standard for data and metadata exchange.<sup>32</sup>

Sound systems refers to the use of international standards and best practice in systems development, including liaising with systems developers in other statistical organizations and making optimum use of off-the-shelf or shared statistical products where available. In keeping with the principles outlined in the Data for Children Strategy, sound systems should also build upon existing national infrastructure rather than creating duplicate structures wherever possible.

The data management team responsible for implementing information technology and information management tools and standards in support of UNICEF statistical data organizes its work according to the Generic Statistical Business Process Model (GSBPM).<sup>33</sup> This model was formulated by the UNECE/Eurostat/OECD Steering Group on Statistical Metadata (METIS) in 2008, and formally endorsed by the Conference of European Statisticians (CES) in 2017.<sup>34</sup> The GSBPM provides an end-to-end set of common processes involved in the production of statistics and gives the data management team a robust and widely accepted conceptual model around which to organize its activities.

The data management team implements software components in line with principles found in the Common Statistical Production Architecture (CSPA),<sup>35</sup> and organizes data in line with the Generic Statistical Information Model (GSIM),<sup>36</sup> once again formulated by METIS (and as with the GSBPM, also formally endorsed by CES in 2017). The GSIM finds its most broadly accepted and UN-preferred implementation in Statistical Data and Metadata eXchange (SDMX), an industrial-strength ISO standard (17369:2013)<sup>37</sup> that underpins several components used by the data management team in its 'standards first, tools second' approach.

Such an approach is a natural part of good data governance and works to promote data quality and data quality frameworks that enhance data quality attributes such as discoverability, comparability, exchangeability, validity, auditability, credibility, reusability, localizability, maintainability, granularity, and timeliness.

32. SDMX, 'UN Statistical Commission: SDMX is preferred standard', accessed at [https://sdmx.org/?sdmx\\_news=un-statistical-commission-sdmx-is-preferred-standard](https://sdmx.org/?sdmx_news=un-statistical-commission-sdmx-is-preferred-standard)

33. European Commission, 'GSBPM: Generic Statistical Business Process Model (Theme)', accessed at [https://ec.europa.eu/eurostat/cros/content/gsbpm-generic-statistical-business-process-model-theme\\_en](https://ec.europa.eu/eurostat/cros/content/gsbpm-generic-statistical-business-process-model-theme_en)

34. United Nations Economic and Social Council, 'New Versions of Generic Statistical Business Process Model and Generic Activity Model for Statistical Organizations', Economic Commission for Europe, Conference of European Statisticians, 67th Plenary Session, Paris, June 2019, accessed at <[https://www.unece.org/fileadmin/DAM/stats/documents/ece/ces/2019/ECE\\_CES\\_2019\\_10\\_Add.1.pdf](https://www.unece.org/fileadmin/DAM/stats/documents/ece/ces/2019/ECE_CES_2019_10_Add.1.pdf)>

35. Common Statistical Production Architecture, 2021, accessed at <https://statswiki.unece.org/display/CSPA>

36. Generic Statistical Information Model, 2021, accessed at <https://statswiki.unece.org/display/gsim>

37. SDMX, News, accessed at <https://sdmx.org>

These approaches have led to a nuanced set of cooperating components that together address statistical business processes in a best practices way. Data, metadata, and reference data (e.g., classifications) are held in a [common data warehouse](#) that implements the SDMX standard and pushes to many public channels such as the [Data Explorer](#). The standard fosters many of the quality attributes listed above. Reference metadata are handled separately (e.g., indicator definitions, computation methodologies, rationales, etc.) and are held in the RDM and made public via [indicator profiles](#). Primary data sources are held in a common [Data Catalogue](#) and described using a common set of metadata fields.

Ethical data systems and data responsibility: Best practices are built not only on sound methods and systems, but on explicit recognition of the potential risks associated with handling data, and ensuring that at all times rights are at the center of our approach to data. UNICEF is building out our resources to better address data responsibility for children through policy and practical tools to support good practice in the field. This includes the Responsible Data for Children principles and tools outlined in Section 2,<sup>38</sup> and compliance with global policy including the UNICEF Procedure for Ethical Standards in Evidence Generation (2021).<sup>39</sup> and the [global policy on personal data protection](#).<sup>40</sup> In practice this means reviewing all major data collections, systems, and projects against these principles and standards and wherever possible, ensuring that this is made available publicly.

Cost-efficiency is also often cited as an important dimension of sound processes and systems, although this can perhaps be thought of more appropriately in terms of the sustainability of data systems. It relates to the cost-efficiency with which statistical outputs are produced as a measure of the costs incurred and resources expended relative to the benefits of the products. The aim is to produce a specified set of output of specified quality at minimum cost.

Efficiency can affect all dimensions of output quality in the sense that, if an output can be produced more efficiently with the same quality, then the resources released can be used to improve any dimension of the quality of that output or other outputs, or to create new outputs.

Three types of costs may be incurred, as described below:

1. The direct costs to UNICEF of acquiring, processing, and disseminating the data.
2. The costs incurred by national organizations from which the data are acquired. These costs depend significantly on whether the national organizations in any case collect and use all the data for national purposes. If they do, then the costs imposed by UNICEF are only that of repackaging and transmitting data already collected. Otherwise, the costs are those of the additional data collection by the national organizations.
3. The costs incurred by the households, businesses, or other units providing their data to the national organizations (respondent burden).

In considering sustainability of the system and recognizing the cost of data collection, it may be necessary to work with key users to prioritize data needs against potential impact for children. For this reason, UNICEF's work to support countries to nationalize the Sustainable Development Goals in line with UN resolutions, and their commitment to developing new approaches to address critical data gaps, is an essential part of the organization's approach to data quality.

38. Responsible Data For Children, 'The RD4C Principles', accessed at [www.rd4c.org](http://www.rd4c.org)

39. UNICEF Internal Document, accessed at <https://unicef.sharepoint.com/sites/OoR-EEG/SitePages/Procedures.aspx?CT=1621223224100&OR=OWA-NT&CID=06dfef80-c391-3f4e-7f47-457517b9468d>

40. UNICEF, 'Policy on Personal Data Protection', accessed at [www.unicef.org/supply/documents/unicef-policy-personal-data-protection](http://www.unicef.org/supply/documents/unicef-policy-personal-data-protection)

## Are there processes to keep 'up to date' with best practice?

Best practices, processes, and systems will become obsolete and undermine data quality over time without a dedicated effort to ensure these are routinely reviewed and updated. While some of this work is undertaken routinely by project teams and data managers, it is also necessary to have a more structured approach across the organization.

### What this means for UNICEF

Evaluation and review: UNICEF's strategic plan is updated every four years, and includes a review of key priorities and data needed as a core part of this process. Similarly, major collections, data policies, and systems should be formally reviewed periodically to adjust for changing priorities, assess new opportunities and methods, and re-visit data privacy and protection concerns. These evaluations should be readily available to internal users across the organization and key funding partners.

### What this means for UNICEF

While impartiality and objectivity is implicit in UNICEF's core mandate, the nature of the organization and our ownership by member nations – as well as potential influence from donors – means that like any international organization, UNICEF must ensure that we remain objective in our use of data and are seen as such. In practical terms this means:

- Statistical methods and outputs are determined by statistical considerations and not by pressure from providers, users, or other stakeholders;
- Commentaries and press releases are objective and non-partisan, with data clearly distinguishable from advocacy or interpretation;
- That source data and metadata (clearly indicating output data quality) are easily accessed from advocacy pieces or other publications;
- The quality of data processes across the organization is managed to ensure the use of international standards and best practice, and a data quality framework is in place for the full extent of data and statistical-related activities across the organization.

## 3.1. Institutional quality dimensions

### Is the organization impartial and objective?

It is well understood that data can be manipulated and presented in ways that support specific opinions or political goals. For users to have confidence in the quality of data produced by UNICEF, and to subsequently use this to inform planning and decision making, there needs to be confidence that data will be presented in an impartial and objective manner.

### Is the organization transparent and credible?

Closely related to the concepts of impartiality and objectivity are those of transparency and credibility. The credibility of UNICEF and its output refers to the confidence that users have in the products, based primarily on their image of UNICEF and its statistical outputs, i.e., the brand image. Credibility is built over time. It depends significantly on accuracy, or more precisely, on users' perceptions of accuracy, which in turn depends upon all the other quality dimensions. Other particularly important contributors to credibility are trust in the objectivity and impartiality of methods used; the perception that outputs are produced professionally in accordance with appropriate statistical standards; that data are not manipulated to suit the countries to which they refer; that estimation methods are appropriate; that the release of outputs is not timed in response to political pressure.



## What this means for UNICEF

As noted above, credibility is built on a holistic approach to data quality and transparency around how data decisions are made at UNICEF. Key dimensions of data quality that are of particular note here (but already discussed previously) include: accessibility; documentation of metadata; sound processes; and ethical use of data and data responsibility.

In addition to dimensions discussed earlier, transparency and credibility also require:

Adoption of a data quality framework that requires that:

- users are informed about sources and methods, and about changes to these that might affect the outputs;
- the limitations of the outputs – and the processes by which they are produced – are acknowledged; and
- errors discovered in published data are corrected as soon as possible and publicized.

Procedures for review, and to respond to complaints or concerns: There should be clear channels open for all data users to engage with the organization to highlight any concerns or complaints regarding the data. A documented procedure should be in place for how these will be referred for consideration and who will be responsible for ensuring this process. There may be different points of contact for different data sets and systems, and these should be readily identifiable. As a first point of contact, internal users may use the data helpdesk contact point: [dataforchildren@unicef.org](mailto:dataforchildren@unicef.org), while external users are able to make contact through the [data.unicef.org website](https://data.unicef.org).

Statistical coordination: Credibility is heavily influenced by the clarity of messaging and consistency of data generated from different parts of the organization. Statistics produced by UNICEF should be harmonized in the sense that they use common concepts, classifications and definitions wherever possible and that the data are mutually compatible. This is achieved through statistical coordination, i.e., coordination of statistical activities within the UNICEF. The Data Task Force, co-chaired by the Associate Director of Data and Analytics and the Associate Director of the Programme Division, as well as annual meetings of the Monitoring, Data and Research Functions (MADR) from divisions across the organization, play an integral role in statistical coordination at UNICEF.



## 4. FURTHER INFORMATION AND NEXT STEPS

This framework and set of definitions are the first step in a broader process to strengthen the availability and transparency of data quality information and tools at UNICEF. Additional information and support on selecting data to work with are available internally through the [Data for Children website](#) and [Data Collection website](#). Externally, additional data on the types of data collected, collated, and used at UNICEF is available through our public site [www.data.unicef.org](http://www.data.unicef.org).

This document is being released as a working document and will be reviewed and updated in line with feedback from across the organization. Please contact the [data helpdesk](#) if you would like to provide feedback on this process. Additional information on data quality for key collections at UNICEF, data governance and procedures, and tools including a self-assessment will be developed and released on the Data for Children site as they become available.



## 5. ANNEX

### **Good practices as outlined in the Principles Governing International Statistical Activities**

A series of good practice principles were outlined and endorsed by the chief statisticians/coordinators of statistical activities of UN agencies in 2005 to illustrate how agencies should address the Principles Governing International Statistical Activities.<sup>41</sup> They are as follows:

41. Committee findings.

1. High quality international statistics, accessible to all, are a fundamental element of global information systems.

**Good practices include:**

- Having regular consultations with key users, both inside and outside the relevant organization, to ascertain if their needs are being met;
- Periodic review of statistical programmes to ensure their relevance;
- Compiling and disseminating international statistics based on impartiality;
- Providing equal access to statistics for all users;
- Ensuring free public accessibility to key statistics.

2. To maintain trust in international statistics, their production is to be impartial and strictly based on the highest professional standards.

**Good practices include:**

- Using rigorously professional considerations for decisions on methodology, terminology, and data presentation;
- Developing and using professional codes of conduct;
- Making a clear distinction in statistical publications between statistical and analytical comments on the one hand, and policy prescriptive and advocacy comments on the other.

3. The public has a right to be informed about the mandates for the statistical work of the organizations.

**Good practices include:**

- Making decisions about statistical work programmes publicly available;
- Making documents for, and reports of, statistical meetings publicly available.

4. Concepts, definitions, classifications, sources, methods, and procedures employed in the production of international statistics are chosen to meet professional scientific standards and are made transparent to users.

**Good practices include:**

- Continuously aiming to introduce methodological improvements and systems to manage and improve the quality and transparency of statistics;
- Enhancing the professional level of staff by encouraging them to attend training courses, to do analytical work, to publish scientific papers, and to participate in seminars and conferences.
- Documenting the concepts, definitions and classifications, as well as data collection and processing procedures used and the quality assessments carried out and making this information publicly accessible.
- Documenting how data are collected, processed, and disseminated, including information about editing mechanisms applied to country data.
- Giving credit, in the dissemination of international statistics, to the original source and using agreed quotation standards when reusing statistics originally collected by others.
- Making officially agreed standards publicly available.

5. Sources and methods for data collection are appropriately chosen to ensure timeliness and other aspects of quality, to be cost-efficient, and to minimize the reporting burden for data providers.

**Good practices include:**

- Facilitating the provision of data by countries;
- Working systematically on the improvement of the timeliness of international statistics;
- Periodic reviews of statistical programmes to minimize the burden on data providers;
- Sharing collected data with other organizations and collecting data jointly where appropriate;
- Contributing to an integrated presentation of statistical programmes, including data collection plans, thereby making gaps or overlaps clearly visible;
- Ensuring that national statistical offices and other national organizations for official statistics are duly involved, and advocating that the Fundamental Principles of Official Statistics are applied when data are collected in countries.

6. Individual data collected about natural persons and legal entities, or about small aggregates that are subject to national confidentiality rules, are to be kept strictly confidential and are to be used exclusively for statistical purposes or for purposes mandated by legislation.

**Good practices include:**

- Putting measures in place to prevent the direct or indirect disclosure of data on persons, households, businesses, and other individual respondents;
- Developing a framework describing methods and procedures to provide sets of anonymous micro-data for further analysis by bona fide researchers, maintaining the requirements of confidentiality.

7. Erroneous interpretation and misuse of statistics is to be immediately and appropriately addressed.

**Good practices include:**

- Responding to perceived erroneous interpretation and misuse of statistics;
- Enhancing the use of statistics by developing educational material for important user groups.

**8. Standards for national and international statistics are to be developed on the basis of sound professional criteria, while meeting the test of practical utility and feasibility.**

**Good practices include:**

- Systematically involving national statistical offices, and other national organizations for official statistics, in the development of international statistical programmes, including the development and promulgation of methods, standards, and good practices;
- Ensuring that decisions on such standards are free from conflict of interest, and are perceived to be so;
- Advising countries on implementation issues concerning international standards;
- Monitoring the implementation of agreed standards.

**9. Coordination of international statistical programmes is essential to strengthen the quality, coherence, and governance of international statistics, and avoid duplication of work.**

**Good practices include:**

- Designating one or more statistical units to implement statistical programmes, including one unit to coordinate the statistical work of the organization and represent the organization in international statistical meetings;
- Participating in international statistical meetings and bilateral and multilateral consultations whenever necessary;
- Working systematically towards agreements about common concepts, classifications, standards, and methods;
- Working systematically towards agreement on which series to consider as authoritative for each important set of statistics;

- Coordinating technical cooperation activities with countries, between donors, and between different organizations in the national statistical system to avoid duplication of effort and to encourage complementarities and synergy to the improvement of statistics in the organizations and in countries.

**10. Bilateral and multilateral cooperation in statistics contributes to the professional growth of the statisticians involved and to the improvement of statistics in organizations and in countries.**

**Good practices include:**

- Cooperating and sharing knowledge among international organizations and with countries and regions to further develop national and regional statistical systems;
- Basing cooperation projects on user requirements, promoting full participation of the main stakeholders, taking account of local circumstances and the stage of statistical development;
- Empowering recipient national statistical systems and governments to take the lead
- Advocating the implementation of the Fundamental Principles of Official Statistics in countries
- Setting cooperation projects within a balanced overall strategic framework for national development of official statistics.

