

IATI Data Quality Index

Background Paper for Consultation

[Introduction](#)

[What does good quality data look like?](#)

[Existing data quality measures with proposed changes](#)

- [3.1. Timeliness](#)
- [3.2. Comprehensiveness - change to Data Availability](#)
- [3.3. Data Availability - Mandatory](#)
- [3.4. Data Availability - Location](#)
- [3.5. Data Availability - Classification](#)
- [3.6. Data Availability - SDGs](#)
- [3.7. Data Availability - Identifiers and traceability](#)
- [3.8. Data Availability - Financial](#)
- [3.9. Data Availability - Humanitarian data](#)
- [3.10. Data Availability- Results](#)
- [3.11. Data Availability - Documents](#)
- [3.13. Data Availability - Coverage](#)

[Additional new measures](#)

- [4.1. Basic validation measure](#)
- [4.2. Data Complementarity](#)
- [4.3. Trust in data](#)

[Discussion and Feedback](#)

1. Introduction

[IATI Strategic Plan \(2020-2025\)](#) sets out three clear objectives of equal importance. Objective one is about driving a significant improvement in the quality of IATI data. It states that by 2025, we will support current and new publishers to meet the highest standards of data quality through improved tools and guidance, **establish our own metrics** and **use the IATI Dashboard to measure progress more effectively** and assist publishers to publish better data.

In order to measure progress more effectively towards our Strategic Plan, in consultation with our members, an [IATI Results Framework](#) was set in place. A specific commitment for 2021 is to develop a Data Quality Index through a consultative process with IATI community members (output indicator 1.1). The Working Group contributing to the development of the Results Framework suggested that the Index should include revised measures for timeliness, comprehensiveness and forward-looking data quality (currently assessed via the [Publisher Statistics](#)). The Index could also include measures for coverage and traceability, among others, and, to the extent possible, will quantify data users' level of trust in published data and utilisation of the gender marker, as feasible. As the next step in establishing the new Data Quality Index, we are launching a broad consultation for the Data Quality Index beginning on September 1st. See consultation timeline on [IATI Connect](#) for further details and next phases.

The **aim** of the IATI Data Quality Index consultation is to reach agreement within the IATI community on **what good quality IATI data looks** like and how this can be promoted through a set of **revised data quality measures for all IATI publishers**, to be implemented in a new Data Quality Index. The consultation will follow a two part process. Discussion 1 focuses on what we want to measure, while Discussion 2 will look at the methodology for each agreed measure. Detailed information on the substance and timing of each phase of the consultation can be found on [IATI Connect](#).

This **background paper forms the basis for Discussion 1**. It sets out changes to existing measures and proposes additional new measures for the creation of a Data Quality Index. The proposals are based on feedback received from the community to date. This includes discussions on our community platform IATI Connect, feedback received during different community events, specific feedback on the current Publishing Statistics measures as well as through our guidance consultations. We have also referenced research commissioned by the IATI Data Use Working Group on Feedback Processes written by [Catalpa in June 2020](#), focusing on the mentioned drivers for data quality improvements. The paper outlines the rationale for each measure and asks our community members for feedback on [IATI Connect](#).

A summary of all proposed measures to be included in the Data Quality Index (DQI) has been provided below. Detailed information about each measure is provided in sections 3 and 4.

IATI Data Quality Index

Name of measure	New or changes	Summary of change
Timeliness	Existing measure with changes	Including two new measures on meaningful updates, consistency and active publishers
Data Availability - Mandatory	Existing measure with changes	Including mandatory and recommended elements
Data Availability - Location	Existing measure with changes	Creating separate measures and including sub-national location details
Data Availability - Classification	Existing measure with changes	Including finance, flow, collaboration type, disbursement channel and also specific gender policy marker
Data Availability - SDG	New	N/A
Data Availability - Identifiers and traceability	Existing measure with changes	Including transaction provider and receiver organisations, participating organisation type and role, and check for correct identifiers
Data Availability - Financial	Existing measure with changes	Including budget measure from both activity and organisation file (previously forward looking measure), as well as specific spend measure and assessment of dates
Data Availability - Humanitarian	New	N/A
Data Availability - Results	New	N/A
Data Availability - Document Links	Existing measure with changes	Including all document link references in both activity and organisation file
Data Availability - Coverage	Existing measure with changes	Including total expenditure from the organisation file
Validator Measure	New	N/A
Data Complementarity	New	N/A
Trust	New	N/A

2. What does good quality data look like?

The number of organisations publishing data to IATI has grown over time with over 1,300 organisations publishing at present to IATI. The IATI Secretariat has strongly encouraged IATI publishers to increase the quality of their data over time. This includes ensuring there are no technical errors in their data, using the [IATI Validator](#), as well as improving its key qualities, making it more timely, comprehensive and forward-looking. We know that good quality data is essential for data use and that good quality data will build trust for our data users.

However, IATI has never had an agreed definition of what “good quality” data looks like from the different perspectives of all our stakeholders. During this consultation we will go in depth into the measures that are important for both publishers and data users. Following the consultation we want to have a **clear agreed definition of what good quality data looks like** and what all our IATI publishers should be striving for to make their data usable.

3. Existing data quality measures with proposed changes

The [IATI Publishing Statistics](#) has been the only existing set of measures looking at all IATI publishers and focusing on four assessment areas: **timeliness**, **comprehensiveness**, **forward-looking** and **coverage**. These measures were initially developed based on the Global Partnership for effective development co-operation established in 2011 in Busan, South Korea. Based on existing feedback from our community over the years as well as consultations as part of the IATI Strategic Plan, there is support to **keep the existing four measures** with a number of proposed changes and additional measures that will capture what good quality data looks like. Support for supplementing existing measures and not completely removing existing measures, can also be seen in recent [Catalpa Report](#) (June 2020, p.35):

*“The IATI Secretariat is already publishing statistics on data quality for each publisher on both the IATI Dashboard and the new IATI Validator. Data quality statistics should be extended to cover additional data quality issues that can be assessed in an automated way. For example, the data quality assessment could include whether organisations report valid organisational identifiers for funding and implementing agencies, if data is made available in the language of the recipient country, and if titles and descriptions have a minimum length. This would **supplement existing indicators measuring timeliness, comprehensiveness, and forward-looking data**. Work under this recommendation could be covered through the Strategic Plan commitment to “establish metrics and use the IATI Dashboard to measure progress more effectively and assist publishers to publish better data”.*

3.1. Timeliness

Rationale

Having up-to-date information is one of the key benefits of IATI data. It is of critical importance for any data users and decision-makers. The recent COVID-19 pandemic highlighted even more the demand and importance of timely data. Currently organisations publish according to their own timeliness, or without consistent updates, and it's not always apparent to users how up to date the data is.

Current methodology

The current timeliness measures seek to assess how often a publisher updates their data. As transactions are the most numerous element reported in IATI, the adopted methodology assumes that a publisher has updated their data if a transaction that has a more recent transaction date than previously published is detected in the publisher's dataset(s). This information is then used to assess two different measures:

- **Frequency:** Assesses the number of days in each of the last 12 months on which the most recently recorded transaction date changed; provides frequency assessment of monthly, quarterly, six-monthly, annual, and less than annual
- **Timelag:** Assesses the number of transaction dates reported in each of the last 12 months; provides timelag assessment of one month in arrears, a quarter in arrears, six months in arrears, one year in arrears, more than one year in arrears

Additional timeliness measures

While frequency and timelag are useful measures, they look at a single change in a transaction. They won't tell us if a thorough and consistent update to important data such as spending has happened. In order to address this challenge, we propose to add two new measures under timeliness:

- **Timely substantive updates to spending data:** this measure will be assessing whether an organization has published spending data (disbursement or expenditure) by checking whether an update has taken place within the quarter. The check will not be looking at the actual spend values for this assessment
- **Consistency of updates:** this measure will be assessing whether the updates to transactions are made at a consistent interval. It will check if transactions (disbursement and expenditure) are updated consistently for all activities and whether the transaction dates align

Additionally we propose disaggregating the timeliness measures to assess whether a publisher is **active**. An active publisher is a publisher that has made any update (not just transaction, but

a change in any of the elements) over the last 12 months in their IATI activity file. If no updates over the last 12 months they will be counted as inactive. If publishers are inactive they can be flagged in red so that the users are aware that the publishers have not made any updates to their data in the last 12 months.

What data quality improvements do we want to drive with timeliness:

- Increased frequency in updates to an organisation's dataset
- Increased consistency in publishers' updates
- Increased updates overall

Expected outcome

We would expect that adding these measures would help to drive publishers to more frequently update their data in a meaningful and consistent way to make their data more usable.

3.2. Comprehensiveness - change to Data Availability

Rationale

Comprehensiveness is an important measure of data quality because it assesses the availability of information. This is of critical importance to data users because the absence of data in particular fields can make it difficult or impossible to use the data. As data acquires meaning when multiple pieces of information are connected together, we know that the added value of IATI data is in the breadth and richness it provides, including not only financial data, but also location, SDGs, results and others.

Current methodology

Comprehensiveness looks at **what** data is being published by organisations and what the important elements for data users are. Comprehensiveness is an existing measure that was looking at three specific areas:

- **Core:** focusing on all IATI mandatory elements (Version, Reporting-Organisation, IATI-identifier, Participating Organisation, Title, Description, Status, Activity Date, Sector, and Country or Region)
- **Value added elements:** additional IATI elements that increase the richness and usefulness of IATI data (Contacts, Location Details, Geographic Coordinates, DAC Sectors, Capital Spend, Activity Documents, Aid Type, Recipient language, and Result/Indicator)
- **Financial information:** assessing four aspects of financial reporting: the reporting of commitments and spend, the ability to track funds across activities and organisations,

and the existence of activity budgets (which are also given added weight in the forward-looking component)

Proposed change

With the maturing of the Standard, the data, and user needs, we need to be a lot more granular in the assessment of how comprehensive an IATI dataset is, assessing in greater depth specific elements that are critical in making the data usable. As such we propose to retain all elements that have previously been measured and add additional elements, splitting them in the following assessment categories (see below) that will give users an easy way to view/ get an assessment of data availability. We are also proposing to rename this assessment area to Data Availability.

This is a summary of Data Availability assessment criteria:

- Data Availability - Mandatory
- Data Availability - Location
- Data Availability - Classification
- Data Availability - SDGs
- Data Availability - Identifiers and traceability
- Data Availability - Financial
- Data Availability - Humanitarian data
- Data Availability - Results
- Data Availability - Documents and contacts
- Data Availability - Coverage

What data quality improvements do we want to drive with this measure:

- Increased data availability in assessment areas important for data users, including known data gaps
- Increased coverage in the data published to IATI

3.3. Data Availability - Mandatory

Rationale

It is important that all publishers include all mandatory and recommended IATI elements. Previously called “core”, this measure will assess and check if an organisation has included [all mandatory and recommended elements](#). These are Reporting-Organisation, IATI-identifier, Participating Organisation, Title, Description, Activity Status, Activity Date, Sector, Country or Region, Budget and Transaction.

Checking for mandatory elements is one of the schema checks that the IATI Validator follows. As such, we propose to include mandatory elements under Data Availability and have an additional separate Validator measure including all validation checks (see section 4.1).

3.4. Data Availability - Location

Rationale

Knowing **where** activities are taking place is essential for many data users. For instance, knowing if an activity is taking place in a specific country, region, or an admin area in a specific city is key for certain purposes. In order to assess whether organisations are publishing good location data, this measure will look at three areas:

- **Country and regions** - assessing whether an organisation publishes country or region data at activity and/or transaction level. If at activity, the assessment looks at the percentages being assigned
- **Sub-national location** - assessing whether an organisation is publishing **geographic coordinates** and **location details** (location name, location description, administrative area)
- **Recipient language** - assessing whether an organisation contains title and description elements written in at least one of the official languages spoken in the reported recipient country

3.5. Data Availability - Classification

Rationale

How IATI activities are being classified is an important piece of information for data users. For example, we want to know if funding is going to health or education, whether the activities are addressing gender equality, whether they are provided as general budget support, standard grant or loan, ODA or other flows. Information on disbursement channels and type of collaboration are particularly useful for the government to identify whether resources are off/on budget.

In order to assess whether organisations are publishing good quality data under the classification category, this measure will look at the following areas:

- **Sector** - assessing the activities with sector codes published at activity or transaction level, including percentage allocations. We will look at both DAC sectors as well as [other vocabularies](#). For the assessment of other sector vocabularies the check will be done for

the following vocabularies 1, 2, 3, 4, 5,6, 11, 12, 99, 98. SDGs (vocabulary 7-9) and humanitarian clusters (11) will be assessed separately under SDGs and humanitarian

- [Aid-type](#) - assessing the activities that contain details of the type of aid being supplied. This can be done at activity level using default-aid-type, or at transaction level using aid-type
- [Finance-type](#) - assessing the activities that contain details of the finance type being supplied. This can be done at activity level using default-aid-type, or at transaction level using aid-type
- [Flow-type](#) - the percentage of all current activities that contain details of the flow type being supplied. This can be done at activity level using default-aid-type, or at transaction level using aid-type
- [Collaboration-type](#) - assessing the activities that contain details of the collaboration type
- [Disbursement-channel](#) - assessing the activities that contain details of the disbursement channel
- [Policy-marker](#) - assessing the activities that contain policy markers and a specific check/ assessment for using the gender equality marker (1)

3.6. Data Availability - SDGs

Rationale

Monitoring progress towards achieving the Sustainable Development Goals is another important area as a universally recognised and accepted measure of progress in development. In 2019, IATI developed guidance for [publishing data to the SDGs](#) and strongly encouraged all publishers to include SDGs. This measure will be looking specifically at the Sustainable Development Goals and will assess whether organisations have made progress using this guidance by including a reference to either an SDG Goal, Target or Indicator.

3.7. Data Availability - Identifiers and traceability

Rationale

The inclusion of specific data within IATI activities allows users to trace the flow of funds from the original donors through the chain of implementation to the intended beneficiaries. Delivery chains can be long and complex; however, if organisations include the relevant information in the dataset, this would allow users to identify relevant partners and trace the flow of funds. Inclusion of IATI identifiers and organisation names are also critical for data users.

In order to assess whether organisations are publishing data that would allow for traceability this measure will look at the following areas:

- **Correct organisation identifiers:** assessing for both the reporting organisation and also for the participating organisations. For participating organisations we will check by identifying if the organisations match with the list of publishers on the IATI Registry.
- **Transaction provider and receiver information:** assessing for the inclusion of transaction provider and receiver organisation (including the organisation identifier and activity identifier);
- **Participating organisation role and type:** assessing who is taking part in the activity (whether government, national NGO) and what role they are playing, which will tell us where in the delivery chain the organisation is

3.8. Data Availability - [Financial](#)

Rationale

Financial information is what data users mainly look for in IATI data. Financial information in IATI is provided by the use of transactions and budgets. In order to assess whether organisations are providing good quality financial information we will look at the following measures:

- [Transaction type](#): assessing which transaction types organisations are publishing and splitting them by incoming and outgoing transactions
- **Spend:** assessing whether organisations have reported spend. This will be assessed by looking at disbursement and expenditures only
- **Transaction dates:** assessing whether the date of the transaction is the same as the transaction value date
- **Forward-looking budgets:** assessing the number of activities with budgets reported for the next two years. This measure will look at the number of activities with budgets reported for each year compared against the total number of activities currently at the start of each year. Previously this was included as a separate measure under ([Forward-looking](#)). We propose to keep it under financial. Additionally, we will include an assessment of the organisation's **total budget** for the **next three years**. This will give an idea of forward-looking data for both the activity and organisation files

Note: hierarchies will need to be considered when assessing the financial information and will be detailed in discussion 2 on methodology.

3.9. Data Availability - [Humanitarian data](#)

Rationale

This measure highlights the importance of identifying humanitarian data for our humanitarian data users. It aims to assess the extent to which IATI publishers are reporting on humanitarian activities. This is important because IATI data is recognised as one of the key platforms for accessing timely and open data on humanitarian assistance. At the UN World Humanitarian Summit 2016, the [Grand Bargain agreement](#) committed governments and international organisations to timely, transparent, harmonised and open high-quality data on humanitarian funding using the IATI Standard.

In order to assess if organisations are providing good quality financial information we will look at the following measures. This is also in line with the Grand Bargain recommendation on best practice in humanitarian data reporting.

- **Humanitarian flag** - assessing whether the activity is humanitarian or not
- **Humanitarian scope** - assessing whether the activity refer to an [appeal or emergency](#) and whether it references the [Humanitarian response Plans or Global Emergency Identifier \(GLIDE\)](#)
- **Humanitarian clusters** - assessing whether organisations include which humanitarian cluster the activity relates to (i.e. WASH)
- **Localisation** - assess whether organisations are including local actors in their data (using the participating organisation type: 24 - Partner Country based NGO and 72- Private Sector in Aid Recipient Country)

3.10. Data Availability- Results

Rationale

Results data represents a relatively new area of the IATI Standard that was introduced in 2018. We recognise the importance of including results information for our users and propose to include a simple measure of results that will assess whether organisations have reported results. If so, the percentage of indicators with baseline, target and actual values (if period is complete).

3.11. Data Availability - Documents

Rationale

This measure will assess whether organisations have included document links in both the activity and organisation file. We know this is important for data users to be able to read additional key documents, such as annual reports, evaluations and others.

3.13. Data Availability - [Coverage](#)

Rationale

Coverage aims to assess what proportion of an organisation's total operational spend is published to IATI, and is currently not assessed. Presenting how much of an organisation's total spend is published to IATI is a meaningful inclusion and we propose to calculate the coverage statistic for an organisation by using its IATI organisation file. We propose to use the **total expenditure** element in a publisher's organisation file, and compare this to the total spend for a given year in their activity files.

4. Additional new measures

4.1. Basic validation measure

Rationale

IATI launched the [IATI Validator](#) in September 2020. The IATI Validator is an online tool for checking whether data aligns with the rules and guidance of IATI Standard. Alignment to IATI rules and guidance is the bare minimum expectation of compliance by publishers. The IATI Validator allows publishers to check and improve the quality of IATI data and as such it is an important measure that we propose to add to the Data Quality index. Community members have requested (reference below) that the Validator measure be available within the DQI (as opposed to having a separate tool for users to look at).

[Catalpa report](#) (June 2020, p.35):

“An important way to improve communication would be to improve the user experience and navigation among different tools providing information on data quality. Whether or not an overall score is prominently provided, the Publishing Statistics, Dashboard, and Validator should present a consistent and clear user experience in line with the proposals of the 2020 IATI Stocktake. Major and priority issues should be highlighted clearly on a single page for each publisher. This should be simple to understand for non-technical users, while providing the depth that more technical users need to understand issues in more detail.”

Proposed measure

Include statistics from the [IATI Validator](#), showing the percentage change of critical, errors and warnings for each publisher on a weekly basis.

Expected outcome

We hope that all publishers will use the information from the Validator to reduce their errors to the point where they have no validation errors or warnings.

4.2. Data Complementarity

Data becomes more meaningful when multiple pieces of good quality information can be effectively combined to undertake analysis. A Transaction Value alone has little meaning until it is combined with other pieces of information such as a Transaction Type, Date, or Sector. Thus, it is not only important that individual pieces of information are of good quality, but that multiple pieces of information are of sufficient quality to be effectively analyzed together. We would like to propose adding an additional set of measures that assess whether the quality of data enables analysis that data users often need to undertake using IATI data. Below you will find some initial proposals, together with brief explanations of the consequences of implementing / not implementing them; however, we welcome feedback from the community on these proposals as well as other possible measures on what should be included here.

- **Ability to assess predictability:** This would measure whether an organisation is providing data that enables users to compare what they have committed and then what they have actually spent to determine whether they are following through on their commitments.
- **Usefulness of budgets to inform budget planning:** For budgets to inform national budget planning in partner countries, they need to be available for 3 years into the future prior to the start of a country's fiscal year. This measure would attempt to assess whether budgets have been provided for the following 3 years at least one month prior to the start of a new calendar year. It could potentially assess the percentage of the next 3 years that is covered by provided budgets (or less than 3 years if the activity ends prior to this). This would be assessed annually.
- **Ability to assess funds flowing into and out of an organisation based on the data they report:** This measure would leverage existing Availability measures on specific data needed for traceability to assess whether an organisation is providing the necessary information to enable a user to trace the flow of funds into and out of an organisation.

4.3. Trust in data

Trust was highlighted as an important measure that needs to form part of the Data Quality index. Creating an easy measurable assessment of trust is complex, and we really welcome community suggestions in this section.

Some initial proposals:

- Assessing whether an organisation had suddenly made their data unavailable/removed their data
- Assessing the time it takes an organisation to fully resolve data quality issues after first notified
- Assessing whether an organisation is making improvements to their data (based on validator measure) over a specific period of time

5. Discussion and Feedback

Please share your feedback on the background paper and each suggested measure on [IATI Connect](#).