Methodology of the ICT Development Index 2023: Version 3

Prepared by the ITU Secretariat. 2 August 2023.

Contents

1 Introduction 1

2 Conceptual framework (step 1) 3

3 Indicator selection and quality assessment (steps 2 and 3) 7

4 Identifying and treating outliers and missing data (step 4) 20

5 Normalization, weighting, and aggregation (step 5) 24

6 Conclusion and next steps 26

References 28

Annex 1: Indicative timeline for the development of the ICT Development Index (IDI) 2023 29

Annex 2: Data availability by economy and indicator 30

Annex 3: Correlation analysis 36

1 Introduction

Created to measure the level of development of the information and communication technology sector (ICT), the ICT Development Index (IDI) was a composite indicator published by ITU from 2009 until 2017. It was discontinued in 2018, owing to issues of data availability and quality (see Box 1).

In October 2022, ITU's Plenipotentiary Conference 2022 in Bucharest adopted a revised text of Resolution 131. This new text (PP Resolution 131 Rev. Bucharest, 2022), referred to as Resolution 131 hereafter) defines, *inter alia*, the main features of the process for developing and adopting a new IDI methodology and of the IDI itself (see Box 2). Consistent with the urgency imposed by Resolution 131, the objective is to launch the IDI in 2023 (see process and timeline in Annex 1).[1]

In this context, and in line with *instructs 8 to the BDT Director*,[2] the Secretariat prepared a Zero draft document, which described a possible framework and structure for the IDI, to inform, facilitate and expedite the process. This document was posted on a discussion forum dedicated to the new IDI (IDI Forum), where members of the Expert Group on ICT Household Indicators (EGH) and the Expert Group on Telecommunication/ICT Indicators (EGTI) were invited to share feedback on the draft methodology, comments and suggestions. This approach is in line with the longstanding working methods of EGTI /EGH, which have their own online forums.

More than 200 members signed up for the IDI Forum and almost 100 comments were posted. A document with a compilation of all the comments received on the content and the respective responses from the ITU

[1] Resolution 131 instructs the BDT Director to "urgently perform the tasks set out in *resolves* above".
[2] "to facilitate the work of EGTI/EGH in fulfilling the tasks set out under resolves above, including through correspondence";

Secretariat was produced and posted on the IDI Forum. The Zero draft document was revised in light of the comments received to produce a Version 1 document, which also contained additional proposals related to the treatment of outliers, aggregation and weightings. A compilation of all the comments received and the Secretariat's responses to each of them were appended as Annex 4 to the Version 1 document.

In Circular BDT/DKH/IDA/009 of 21 April 2023, Member States were invited to comment on the Version 1 document, as per *instructs the Director of the Telecommunication Development Bureau* 7 of Resolution 131.[3] The period of consultation for Member States ran from 21 April 2023 to 19 May 2023.[4] Fourteen Member States submitted a total of 71 comments.[5] Based on the comments received, the Secretariat updated the Version 1 document and produced a Version 2 document. Annex 4 of that document features a compilation of the comments submitted by Member States on the Version 1 document and the Secretariat's responses to each of them.

Following these consultations, the ITU Secretariat identified a number of outstanding issues, which were presented in the Version 2 document (and summarized in text boxes labelled 'Issues for discussions at the IDI meeting'). These outstanding issues were discussed at the virtual joint meeting of EGTI and EGH from 13 to 15 June 2023 (Circular BDT/DKH/IDA/007 of 21 March 2023). The report of the meeting was posted on 27 June. The Secretariat has updated the methodology to reflect the outcomes of the meeting and produced this Version 3 of the methodology. In addition, the Secretariat has maintained a webpage with all details about, and documents related to the process.

In June 2023, during the 30th meeting of the Telecommunication Development Advisory Group, the ITU Secretariat provided an update about the ongoing process (see Addendum 2 Annex 3 of Document 2). In July, during its 2023 session, ITU's Council noted the developments and requested that an additional meeting be organized ahead of the final consultation of Member States. Accordingly, a joint EGTI/EGH session on the IDI will be held on 18-19 September during the annual meetings of EGTI and EGH, with the aim of concluding the elaboration of the methodology.

Following this meeting, the IDI methodology will be finalised and submitted for approval by Member States in October. The methodology will be adopted if 70 per cent of responding Member States approve it, as per Resolution 131. During this consultation, Member States will also have the possibility to opt out from the 2023 edition if the methodology is approved.

The rest of the document is organized as follows: Section 2 presents the conceptual framework (step 1 of the process of developing an index – see Table 1); Section 3 presents a set of selection criteria which combined with the conceptual framework help identify candidate indicators for inclusion (step 2) and the statistical analysis (step 3) used to narrow down and confirm the choice of indicators. Section 4 describes the approach to identify and treat outliers and estimating missing data (step 4). Section 5 describes the approach to normalize indicators and aggregate them (step 5). Section 6 concludes.

Box 1: A brief history of the IDI

The IDI was published from 2009 to 2017. In the last published edition in 2017, 11 indicators were combined into a composite score.

In March 2017, an extraordinary meeting of the Expert Group on ICT Household Indicators (EGH) and Expert Group on Telecommunication/ICT Indicators (EGTI) adopted a revised set of 14 indicators to be included in the IDI. However, following the shift from 11 to 14 indicators, countries were facing challenges in collecting and

[3] to invite Member States to contribute and comment on the IDI methodology and structure;
[4] The consultation period was initially to end on 19 May but was extended to 22 May to allow for some ITU focal points (or the person to whom they had delegated the task) who experienced technical difficulties to post their comments.
[5] The 14 countries were: Algeria, Bahrain, Brazil, Egypt, India, Kenya, Korea (Rep. of), Oman, Pakistan, Qatar, Russian Federation, Singapore, and the United Arab Emirates. Another country, Japan, posted comments that did not relate to the methodology, but to data availability in Japan of the proposed indicators.

submitting quality data. For the calculation of the 2018 IDI for example, 58 per cent of the data points would have to be estimated. Furthermore, there were issues with the harmonization and quality of the data used, and the methodology applied to derive some of the newly adopted indicators. Because of these flaws it was not possible to compute a methodologically sound index that reflected the true state of ICT development.

Since 2018, attempts either to publish the IDI in line with the Plenipotentiary Conference Resolution 131 “Measuring information and communication technologies to build an integrating and inclusive information society” (Rev. Dubai, 2018) or to develop an entirely new index have been unsuccessful, as no consensus could be reached.

To address these implementation challenges, Resolution 131 was revised at the 2022 Plenipotentiary Conference 2022 in Bucharest. Refer to the ITU website for more on the history of the IDI.

Box 2: Main implications of Resolution 131 for the development of the IDI

Resolution 131 (Rev. Bucharest, 2022) describes the main features of the process for developing the IDI methodology and of the IDI itself (relevant paragraphs of the resolution appear in brackets):

- ITU must publish a new IDI “urgently” (*instructs to BDT Director 1*);
- The new IDI will be published without ranking (*resolves 3*);
- ITU should establish a valid structure and methodology for the IDI, working through EGTI/EGH, and through formal consultations (*resolves 3*);
- ITU should establish the criteria on the minimum data availability for Member States to feature in the IDI, working through EGTI/EGH (*resolves 6*);
- The BDT Director should facilitate the work of EGTI/EGH (*instructs to BDT Director 8*);
- Methodology will be submitted to Member States for approval and adopted if 70 percent of respondents approve it (*resolves 3*);
- If adopted, the methodology will be valid for four editions, namely 2023-2026 (*resolves 4*);
- Member States will have with the option to decline to participate in the IDI during the given period of validity, though with the choice to re-join the exercise on an annual basis (*resolves 5*);
- A meeting of EGTI/EGH will be convened following a formal consultation of Member States with a view to resolving any contentious issues and seeking consensus (*instructs to BDT Director 9*);
- Integrity of all ITU's statistical work must be preserved, in strict adherence to UN principles on good statistics (*instructs to BDT Director 12*).

In addition to the IDI, Resolution 131 covers other topics not discussed here.

2 Conceptual framework (step 1)

ICT development is an inherently multidimensional concept. An evidence-based assessment of country performance therefore requires multiple indicators. An aggregate measure, or composite indicator, serves the purpose of summarizing a range of metrics into a single number. There are both advantages and disadvantages to using composite indicators, summarised in Table 1.

Table 1: Pros and cons of a composite indicator

Pros	Cons
• Can summarise complex, multi-dimensional realities with a view to supporting decision-makers. • Are easier to interpret than a battery of many separate indicators. • Can assess progress of countries over time. • Reduce the visible size of a set of indicators without dropping the underlying information base, making it possible to include more information within the existing size limit. • Uses the power of numbers to advocate an issue of concern and introduce it in the policy arena. • Facilitate communication with the public (i.e., citizens, media, etc.) and promote accountability. • Help to construct/underpin narratives for lay and literate audiences. • Enable users to compare complex dimensions effectively. • Bring public attention to the need to develop and refine statistical data collection.	• May send misleading policy messages if poorly constructed or misinterpreted. • May invite simplistic policy conclusions. • May be misused, e.g., to support a desired policy, if the construction process is not transparent and/or lacks sound statistical or conceptual principles. • The selection of indicators and weights could be the subject of political dispute and may be biased by data availability. • May disguise serious failings in some dimensions and increase the difficulty of identifying proper remedial action if the construction process is not transparent. • May lead to inappropriate policies if dimensions of performance that are difficult to measure are ignored, or if measurement lags are not taken into consideration. • May hide, inequalities within territorial units and trade-offs between alternatives, by presenting the average of averages. • May give the false impression that units are independent competitors, while hiding interdependencies and common underlying processes transcending borders.

Source: Based on OECD (2008).

Aggregation necessarily involves simplification. To guarantee a conceptually and statistically sound index, its construction must follow an iterative process, as formalised in the *OECD-JRC Handbook on Constructing Composite Indicators* (2008) and *Your 10-Step Pocket Guide to Composite Indicators & Scoreboards* from the European Commission (2019) and presented in Table 2.

Table 2: Steps for developing a composite indicator

	Step
1	Develop the conceptual framework based on the stated objective.
2	Identify potential indicators that capture those concepts.
3	For each considered indicator, assess coverage, methodological soundness, quality of data.
	Based on this assessment, revisit the framework, concepts, and/or indicators (steps 1-3) if necessary.
4	Identify and treat any outliers and missing data.
5	Define the suitable normalization, weighting, and aggregation methods.
6	Calculate the index.
7	Assess the statistical and conceptual coherence of the index.
8	Conduct sensitivity analyses and assess the impact of uncertainties on resulting scores.
	Based on the results of the sensitivity analysis, revisit steps 1-8 if necessary.
9	Make sense of the data and validate the results.
10	Communicate the results and underlying information.

Source: OECD (2008) and European Commission (2019).

Step 1 consists in developing a conceptual framework based on the objective of the composite indicator. When the IDI was developed in 2009, the objective was to assess the development of the ICT sector. Such development was seen as a simple progression from *access* to *use* to *impacts*, a sequence that provided the framework for the old IDI. However, the framework focused on the quantity of ICTs and less on the qualitative aspect. This shortcoming is addressed by the concept of *universal and meaningful connectivity* (UMC). UMC is the possibility for everyone to enjoy a safe, satisfying, enriching, productive online experience at an affordable cost. Digital connectivity must be universal *and* meaningful to maximize its impact on society and the economy. UMC reflects the need for a holistic strategy for closing all aspects of the digital divide, across and within countries.

UMC has gained significant traction over the past two years. The concept of UMC was formalised in 2021, in the context of the implementation of the UN Secretary-General's Roadmap for Digital Cooperation. The ITU and the Office of the UN Secretary-General's Envoy on Technology convened a multi-stakeholder sub-working group (SWG) to work on a baseline and aspirational targets for UMC. The baseline and targets were launched in April 2022 along with a background document detailing the concept of UMC.

At the World Telecommunication Development Conference (WTDC) 2022 and ITU's Plenipotentiary Conference (PP) 2022, universal and meaningful connectivity was front and centre. The concept is mentioned multiple times in the Final Report of WTDC 2022: notably in Resolution 2 (Study Groups), Resolution 87 (Connecting every school to the Internet), Resolution 88 (Partner2Connect), Regional initiatives (Europe, Arab States). UMC is also captured in the first Strategic Goal ("Universal Connectivity: Enable and foster universal access to affordable, high-quality and secure telecommunications/ICTs") of the Strategic Plan 2024-2027, adopted at PP 2022.

For these reasons – its relevance and its recognition by ITU constituency – the concept of UMC was adopted as the conceptual framework to guide the development of a new IDI. The choice of the conceptual framework also defines the *objective* of the index: to assess the extent to which a country's connectivity is universal and meaningful.

During the consultations, the conceptual framework based on UMC received extremely broad support from EGTI and EGH members and Member States (see Version 1 and Version 2 documents) . The remainder of this section describes the concept of UMC. For more information about the concept of UMC, refer to ITU and OSET (2022).

Figure 1 illustrates the two dimensions of UMC: use – ranging from *none* to *universal*; and quality – ranging from *no connectivity* to *meaningful connectivity*. "Universal connectivity" means connectivity for all. The two dimensions are complementary: neither universal connectivity with poor quality nor meaningful connectivity for the few will yield significant, society-wide benefits. At the same time, the two dimensions reinforce each other: more use can lead to more meaningful connectivity, and vice versa. Based on the definition of universal and meaningful connectivity, the SWG developed a conceptual framework (Figure 2).

Achieving *universal connectivity* (top half of Figure 2) calls for dedicating attention to the connectivity of people, households, communities, and businesses, rather than merely that of the average population.

- Focusing on people helps achieve universality by ensuring that anyone can connect regardless of their urban or rural location, gender, level of education, etc.
- Focusing on households, communities and businesses helps ensure that the main places where people can connect are represented: at home, in schools and community centres, and at work.

Figure 1: The two dimensions of connectivity

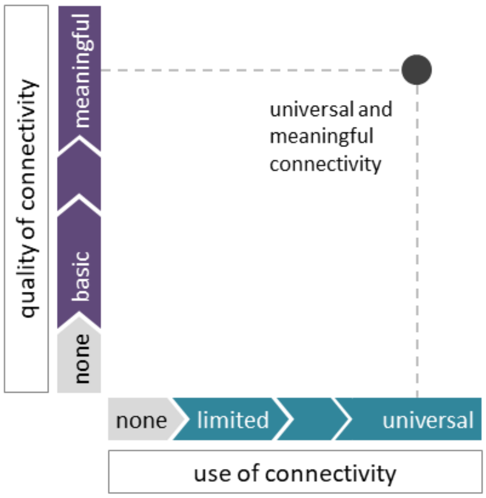
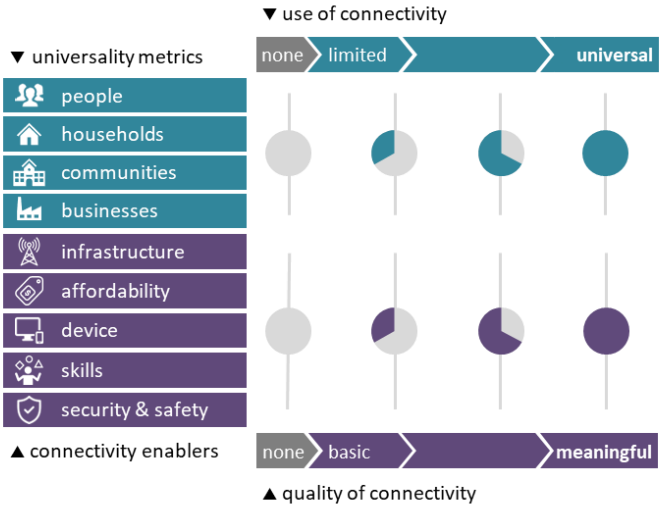


Figure 2: Conceptual framework of universal and meaningful connectivity



Source: ITU and UN OSET (2022).

Meaningful connectivity depends on several factors, called "connectivity enablers": infrastructure, affordability, device, skills, and safety and security (bottom half of Figure 2).

- Meaningful connectivity requires high-quality infrastructure that is not only in place and functioning but allows for a fast and reliable connection. The framework adopts a technology-neutral approach. Satellite connectivity, and fixed and mobile terrestrial networks, all can contribute to connecting people to the Internet.
- Affordable devices and ICT services are essential for enabling people to go online. Affordability is a relative concept that depends on people's social and economic conditions.
- Access to an Internet-enabled device is required to go online. These can be either mobile phones or desktop computers, considering that the most basic models of the former are cheaper, while the latter allow for a richer experience. For mobile phones, it is important to distinguish use from ownership, recognizing that mere access without full possession of a device imposes constraints, including when and for how long one can be online.
- An important barrier keeping people from going online or fully benefiting when they are online is a lack of skills. Meaningful use of the Internet requires that people are digitally literate.
- A safe and sec*ure Internet* is important for people to have the trust to go online.

A country with a highly developed digital eco-system is a country where there is a high Internet usage among the population, empowered by high quality enablers. This means that everyone that wants to can connect to high-quality, affordable, and safe Internet and benefit fully from its services.

The analytical framework defines the scope, but also sets the boundaries of the exercise. The following aspects of connectivity are out of scope:

- **Levers.** Enablers of connectivity representing areas where policymakers and other stakeholders can intervene using tools such as investment, policies, and regulation. They are not included in the framework as it is deliberately agnostic about the means to improve on the various factors, as there is no single pathway and no one-size-fits-all policy mix that can be prescribed to all countries.
- **Catalysts.** Broader factors and trends, such as economic development and technological innovation, that contribute to improving the quality enablers.
- **Content and services.** These are treated as a lever: the more content and services are available, accessible, and relevant, the more likely people are to connect. Content and services are an enabler of connectivity, but they do not directly influence the quality of connectivity, which is what the frameworks aims to assess.

- **Applications.** The framework is deliberately agnostic about what people do with connectivity. The exercise is about measuring the use and quality of connectivity, rather than assessing what people do online.
- **Impacts.** By extension, the societal, environmental, and economic impacts of connectivity and its applications are well beyond the scope of the exercise.

3 Indicator selection and quality assessment (steps 2 and 3)

The next step in the process is to identify potential indicators that capture the concepts of the conceptual framework. Table 3 summarises the criteria for selecting an indicator as candidate for inclusion in the index. These criteria include the instructions from Resolution 131.

Table 3: Indicator selection criteria

	Criterion	Rationale
1	Relevance to the concept	An indicator should measure one aspect of the concept retained for the index, in this case universal and meaningful connectivity and have policy relevance.
2	Clarity/interpretability	Indicators should be easy to interpret and the impact on universal and meaningful connectivity clear.
3	Source	Indicators should rely primarily on official data provided by Member States, based on internationally recognized and transparent methodologies (as per *Instructs to BDT Director 4* of Resolution 131).
4	Reliability	The indicator should be coherently collected by countries according to the harmonized methodology developed by ITU's expert groups EGTI/EGH or by another international organisation.
5	Applicability to measure country performance	The indicator should have a sufficiently high variation to distinguish country performance and signal progress over time. Quantitative indicators are preferred over qualitative indicators.
6	Availability and timeliness	Recent data should be available for as many of the 196 considered economies as possible[6], to ensure the broadest coverage possible and reduce the number of estimates, as per *resolves 3* of Resolution 131.[7]

The first two criteria are self-explanatory steps for any kind of index construction. The third, fourth and sixth criteria stem directly from Resolution 131. The fifth criterion is a best practice in index construction.

The most problematic aspect is data availability. In the context of a composite indicator, maximizing data availability for the countries included is crucial for enabling meaningful comparison. Comparing the performance of a country with 100 per cent data availability against that of a country with only 50 per cent availability is obviously misguided and problematic if the index is meant to help decision making. In addition, limiting the coverage of an index to the sole countries with full or nearly full data coverage would mean excluding most LDCs, and many low- and middle-income economies from the index.

[6] For the purpose of the index, 196 economies are considered: the 193 ITU Member States plus Hong Kong (China), Macao (China), and Palestine.

[7] "[...] that ITU should take further necessary measures to establish a valid structure and methodology for the IDI [...] that allows the publication of the IDI on an annual basis, without ranking, provided that there are sufficient valid data to cover a majority of Member States;"

We follow a two-step approach to indicator selection: 1) indicators that fit the conceptual framework and meet the criteria 2-5 from Table 3 are considered; 2) data availability and timeliness are assessed (criterion 6), using the percentage of economies for which timely official data exists.

Data availability and reference year

To assess data availability, we first identify a *reference year*, which is the year of the majority of the data points. ITU's data collection cycle plays an important role in determining the reference period.

In the ITU questionnaires, countries usually submit data for the previous year. Furthermore, the results of the long questionnaire (LQ), which is conducted in September-October of each year and provides additional and more final data than the short questionnaire (SQ) conducted in March-April, are available at the very end of the year. In addition, all estimates are computed and validated by countries by the end of year, too. The end of the year is therefore a natural cut-off date and defines the reference year. Consequently, the results of the questionnaires of 2022, which contain mostly 2021 data, are used to assess data availability in 2023.

To maximize data availability and reduce the number of estimates, we extend the reference period to the year *preceding* the reference year, that is, 2020 in the present case. The reason is that not all data are collected annually, especially those derived from ICT household surveys.[8] Therefore, the reference period to assess data availability for candidate indicators in 2023 is 2020-2021. When computing the percentage of economies for which data exists, only official data for 2020 and 2021 are considered. Estimated values are *not* considered available data. Annex 2 reports data availability for all indicators for the reference period 2020-2021. Some EGTI/EGH members suggested to extend the reference period to 2019. Considering the trade-off between official data for older years and estimates for more recent years, we believe this would be going too far back, considering the rapid pace of ICT diffusion.

In addition to assessing data availability to guide the *development of the index*, the same principles are to be used to identify the reference period for the *computation of the index* in 2023 and in subsequent years. That is, for an edition of the IDI released in year *t*, the reference period will always be *t-3* and *t-2*.[9] Within this reference period, if data is available for both *t-2* and *t-3*, the most recent (i.e., *t-2*) will be used. Concretely, this means that for the 2023 edition, the reference period will be 2020-2021. For the 2024 edition, the reference period will be 2021-2022, and so on. The only difference with the 2023 edition is that subsequent editions will be launched much earlier in the year.[10]

Indicators for which official data for the reference period 2020-2021 are available for less than 50 per cent of economies (i.e., fewer than 98 economies), are in principle excluded, except if there are compelling reasons to keep them. Estimating more than 50 per cent of data points for an indicator would diminish the reliability of the index. This threshold is already very lenient: a threshold of 65 per cent is more in line with good statistical practices (see for example EC (2019)). But in the case of ICT indicators, this would cause too many indicators to be discarded. Additionally, Resolution 131 limits the use of estimates and other data sources to the strict minimum.[11] Finally, estimating many data points is extremely time-consuming and would delay the release and impacts the timeliness of the assessment.

---

[8] In a handful of cases (typically fewer than 10), countries manage to submit data for household indicators for the current year, if they administered their household survey early in the year and managed to compute and submit the results in the household long questionnaire.
[9] This is different from what was initially proposed in the Zero draft document, where the proposal was that for subsequent edition of the index, we would use *t-1* as reference year for an edition *t* of the index. Upon further investigation, this option proved not feasible, as explained in the text.
[10] If the methodology is approved, the 2023 edition of the IDI will be launched in early or mid December (see Annex 1). The 2024 edition will then be launched in mid-2024 (to allow for six months between two editions). Subsequent editions of the index would be launched at the end of the first quarter of each year.
[11] Resolution 131:"to rely primarily on official data provided by Member States based on internationally recognized and transparent methodologies, while also taking into account their level of ICT and statistical

Throughout the development process, potential indicators were assessed against the six criteria to determine their eligibility for inclusion in the IDI. In the rest of this section, we list the indicators that have been retained for this Version 3 and those that were considered but ultimately not retained as they did not meet one or more selection criteria. Among those, some met all the criteria except data availability. Those are prime candidate for inclusion in a future iteration of the index.

The fact that an indicator was not retained does not mean that it is not important. All 150 or so indicators and subindicators collected by countries and submitted to ITU through the various questionnaires are important. Countries should strive to collect as many of them as possible on a regular basis and in accordance with international standards to ensure reliability. Collectively, these indicators draw a holistic picture of the state of connectivity and digital development. The small set of indicators retained for the IDI can only provide a partial assessment.

Some EGTI/EGH members (see Annex 4 of the Version 1 document) and Member States (see Annex 4 of the Version 2 document) suggested establishing a list of relevant indicators for potential inclusion in a future iteration of the IDI. When data is available, featuring these indicators on the IDI country profiles may contribute to draw a more complete picture of a country's connectivity and encourage countries to collect data.

Indicator selection: Universal connectivity

The notion of universality encompasses four categories: people, households, communities, and businesses (Figure 2). The last three categories represent the main places where people can connect: at home, in schools and community centres, and at work. The following indicators are therefore natural candidates for inclusion: *individuals using the Internet*, *households with Internet access*, *business using the Internet* and *schools using the Internet*.[12] In addition, using the Internet requires a subscription to a service, so *mobile broadband subscriptions* and *fixed broadband subscriptions* feature on the list of candidate indicators.

Retained indicators

Percentage of individuals using the Internet

The percentage of individuals using the Internet is the main indicator for universal connectivity. The indicator is an indicator of the Sustainable Development Goals framework (SDG), defined in *ITU Manual for measuring ICT access and use by households and individuals l* (ITU, 2020a, the 'ITU Household Manual' hereafter). It is also one of the core indicators of the Partnership on Measuring ICT for Development. The source is usually ICT household surveys conducted in countries, often by the national statistical office, but sometimes by other entities as well, such as the regulator. At the international level, data are collected from countries by ITU. Data are also collected by Eurostat for their member countries, as well as by the OECD. For 2021, data were available for 96 economies. Although availability is just below the threshold, the indicator was retained because of its critical relevance in measuring connectivity.

There were comments that noted that countries submit data for different age ranges, proposing to align all countries based on the same population range (e.g., based on 16 to 74 years). This is a very important and relevant point. Although ITU Household Manual (ITU, 2020a, the Chapter 7, page 171) recommends collecting data for all individuals aged 5 and above, many countries do not survey children and/or older persons. This creates comparability issues, particularly where older persons are not surveyed. Countries with available data

database development; only in the absence of such information may other sources be used, after consulting with the focal points of the Member States concerned in advance on other sources used to obtain the information by means of which ITU fulfils the role referred to in considering a) above;"

[12] Internationally comparable data on community centres with Internet access unfortunately do not exist.

consistently report that they are less likely to use the Internet. One option as suggested is to use only the 16-74 age bracket. Though some differences in survey scope would remain, this option has the clear advantage of increasing the comparability between countries. However, there are costs to this approach. First, many countries that provide overall Internet use data do not provide breakdowns by age. Availability of official data for 2020 or later drops from 96 countries to 64 when requiring data for the 16-74 age range – below the threshold set for inclusion in the index. If this indicator was included despite the lack of data, more estimation would be required. In addition, using Internet use for only the 16-74 age range for the purposes of the index diminishes the importance of children and older persons when assessing ICT development in countries, which would be in contradiction with the concept of universality. Therefore, the costs outweigh the benefits.

Percentage of households with Internet access

This indicator covers the most common place where people connect to the Internet: at home. The indicator is defined in the ITU Household Manual (ITU, 2020a). It is one of the core indicators of the Partnership on Measuring ICT for Development. The source is usually ICT household surveys conducted in countries, often by the national statistical office, but sometimes by other entities as well, such as the regulator. At the international level, data are collected from countries by ITU. Data are also collected by Eurostat for their member countries, as well as by the OECD. For 2020-2021, data were available for 94 economies. Although availability is just below the threshold, the indicator was retained because of its importance in the conceptual framework.

Active mobile-broadband subscriptions per 100 inhabitants

A subscription is necessary to use the Internet, and a mobile phone is the most common way for people to go online. To allow for a meaningful connection, the subscription needs to be to a broadband network, which is a 3G or more advanced technology. The indicator is defined in the ITU's *Handbook for the Collection of Administrative Data on Telecommunications/ICT* (ITU, 2020b, referred to as the 'ITU Handbook' hereafter), (ITU, 2020b). It is one of the core indicators of the Partnership on Measuring ICT for Development. The data are usually collected by the ICT regulator, which collects the data from the various operators in the country. At the international level, data are collected from countries by ITU. Data are also collected by Eurostat for their member countries, as well as by the OECD. For 2020-2021, data were available for 170 economies.

Indicators considered but not retained

Fixed-broadband subscriptions penetration rate

An indicator on fixed-broadband subscriptions is necessary to complement the indicator on mobile broadband subscriptions, to avoid an imbalance with and a bias towards mobile infrastructure. Mobile broadband technology is not yet a perfect substitute for wired connections, particularly fibre optic, which remains critical for businesses. The inclusion of fixed broadband penetration increases the likelihood that the index reflects the infrastructure needed to generate positive economic outcomes.

The number of fixed broadband subscriptions is an indicator that is defined in the ITU Handbook (ITU, 2020b). It is one of the core indicators of the Partnership on Measuring ICT for Development. The data are usually collected by the national regulator, which collects the data from the operators. At the international level, data are collected from countries by ITU. For 2020-2021, data were available for 170 economies.

The choice of the denominator has proven problematic.. In previous versions of the IDI, and consistent with the definition adopted by EGTI and codified in the ITU Handbook (ITU, 2020b), the indicator has traditionally been divided by 100 population, which is also an SDG indicator. However, several commenters on the IDI Forum and participants in the IDI meeting considered the number of households to be a denominator that is more aligned with policy objectives than population. This would account for the fact that fixed-broadband subscriptions are often shared within a household and that the average size of households varies across countries. However, up-to-date, internationally comparable data on the number of households are too scarce for using households as denominator.

Participants in the IDI meeting agreed that the indicator *percentage of households with Internet access using fixed broadband*, from ICT household surveys, is the most appropriate indicator to measure household penetration of fixed broadband. However, it was concluded that it was not a feasible alternative due to a lack of sufficient countries collecting and reporting the data. The Secretariat proposed an alternative: the *number of fixed broadband subscriptions per 100 inhabitants aged 18+*. By focusing on adults only, this would lessen disparities in household size that creates comparability issues when using the full population. However, when this alternative was presented during the IDI meeting, there were about as many participants in favour as against it.

As participants shared many different views and no acceptable solution emerged, an indicator on fixed broadband subscriptions is not included in this version of the methodology. The participants acknowledged that this is a big loss for the index, because the indicator is very relevant for the ICT development of countries. The conclusion was that the inclusion of an indicator of fixed broadband penetration must be deferred to 2027, when the IDI will be reviewed. It is hoped that – by then – there will be sufficient data on the number of households and/or on households with Internet access using fixed broadband. The ITU Secretariat will encourage and help to the extent it can Member States to provide these household survey data to the ITU and to provide data on household size to the Population Division of the UN Department of Statistics, with whom it will continue the dialogue on the subject. An EGTI subgroup is discussing the issue of measurement of fixed broadband penetration and the strengths and weaknesses of different denominators as part of the EGTI's 2023 work programme. Conclusions of the subgroup may inform decisions on the measurement of fixed broadband penetration for future iterations of the IDI.

For *fixed broadband subscriptions*, the breakdown by speed tier could also be considered. The argument is that faster subscriptions allow for more content, a better user experience and more connected devices. Some of the comments on the IDI forum highlighted this as well. While this is certainly true, there are limitations. First, the indicator reflects *advertised* speed, and not *actual* speed.[13] There are other indicators that provide a direct measure of speed or an indicator on fixed broadband traffic. These are discussed below, in the *Infrastructure* section. A second consideration is conceptual: the definition of meaningful connectivity implies that a user should be able to do whatever they want, without prescribing any online behaviour. While a faster connection is preferrable, it is not possible to set a goal post as this would amount to prescribing an ideal speed, which in turn would prescribe a certain type of usage. Finally, using the indicator for total fixed broadband subscriptions instead of the breakdown by speed tiers increases the availability of data from 74 to 87 per cent of economies.

Percentage of businesses (10+ employees) using the Internet

This indicator covers a common place where people connect to the Internet: at work. The indicator is defined in the UNCTAD Manual (UNCTAD, 2021). It is one of the core indicators of the Partnership on Measuring ICT for Development. The source is usually ICT business surveys conducted in countries, often by the national statistical office, but sometimes by other entities as well. At the international level, data are collected from countries by UNCTAD. Data are also collected by Eurostat for their member countries, as well as by the OECD. However, for 2020-2021, data were only available for 8 economies. Therefore, while this is a very relevant indicator, as highlighted by many, the indicator is excluded.

Percentage of schools using the Internet

This indicator covers a common place where people connect to the Internet: at school. This is an SDG indicator, defined by the UNESCO Institute for Statistics (UIS) in the SDG 4 Data Digest (UIS, 2019). It is also one of the core indicators of the Partnership on Measuring ICT for Development. UIS collects these data from Ministries of Economies from all economies in the world. A secondary source is Giga, the ITU-UNICEF joint For 2020-2021, data were available for 70 economies, far below the threshold to be included in the IDI, despite its relevance to the conceptual framework.

[13] In general, differences between advertised speed and actual speed are due to network overload, user congestion, or more devices being added to the network (connected devices). Other factors that may also affect performance are, for example, interference or environmental factors.

Indicator selection: Meaningful connectivity

The UMC framework features five connectivity enablers: infrastructure, affordability, device, skills, and safety and security. Ideally, the index would feature indicators capturing each of these areas provided they satisfy the criteria of data availability and data quality.

Meaningful connectivity — Infrastructure

Access to a signal is a prerequisite for using the Internet. The minimum requirement for meaningful use of the Internet is access to a 3G mobile network. The *population covered by at least a 3G mobile network* should therefore be included. Since *higher quality networks* are preferred, these would be assessed at the same time. If and how these different indicators are aggregated is to be determined later. In a similar vein, the *number of households passed by a fixed network* could be included in the index, as this is a prerequisite for subscribing to a fixed broadband service. Another indicator considered that may have an impact on the quality of a fixed broadband connection is the *population that lives within physical reach of (fiber) nodes on core terrestrial transmission networks*. It would also be relevant to include other measures of *middle-mile and last-mile connectivity*.

Concerning the quality of Internet connections, indicators considered include *Internet traffic*, the *speed of Internet connections* and *international bandwidth capacity and bandwidth usage*.

Meaningful connectivity — Infrastructure: retained indicators

Percentage of population covered by a mobile network

Access to a signal is a prerequisite for using the Internet. The minimum requirement for meaningful use of the Internet is access to a 3G mobile network. More advanced technologies with increased capacity and faster connection speeds facilitate more meaningful Internet usage. The indicator is defined in the ITU Handbook (ITU, 2020b). The population covered by a 3G mobile network is one of the core indicators of the Partnership on Measuring ICT for Development. The data are usually collected by the ICT regulator, which collects the data from the various operators in the country. At the international level, data are collected from countries by ITU. For “at least 3G” data were available for 158 countries, for “at least LTE/WiMAX” for 156 countries, and for “at least 5G” for 44 countries. The indicator is included for ‘at least 3G’ and ‘at least LTE/WiMAX’. While ‘at least 5G’ is very relevant, it cannot be included yet, because of poor data availability. It will be reconsidered in 2027. A proposal to combine the different technologies is made in the *Weighting and aggregation* section below.

Mobile broadband Internet traffic per mobile broadband subscription

This indicator measures the intensity of Internet usage by mobile broadband subscribers. A range of specific connectivity needs can only be accommodated through the availability of data-intensive connections at the disposal of users who are able to change their physical location. The indicator reflects the quality of the ICT infrastructure from the end-user’s perspective. The indicator is defined in the ITU Handbook (ITU, 2020b). The data are usually collected by the ICT regulator, which collects the data from the various operators in the country. At the international level, data are collected from countries by ITU.

There was one issue for discussion for this enabler: the choice of the denominator to use, where the choice was to divide either by mobile broadband subscriptions or by Internet users. Dividing by Internet users focuses on assessing traffic at the country level, for the average Internet user. However, it would be less precise, as it would combine data from different sources and it would require estimates for the number of Internet users for a fair amount of countries. In addition, no distinction can be made between Internet users using fixed or mobile broadband. Dividing by subscriptions has the advantage that both the numerator and the denominator would come from the same source, being monitored by telecom operators and national regulatory authorities. Traffic per subscription is also a performance indicator used by many regulators. Eventually, the participants agreed to use subscriptions as denominator.

Fixed-broadband Internet traffic per fixed broadband subscription

This indicator measures the intensity of Internet usage by fixed Internet subscribers. Given today's most widely available technologies, certain user needs can only be accommodated by data-intensive, fast fixed broadband connections. The indicator reflects the quality of the ICT infrastructure from the end-user's perspective. The indicator is defined in the ITU Handbook (ITU, 2020b). The data are usually collected by the ICT regulator, which collects the data from the various operators in the country. At the international level, data are collected from countries by ITU. For 2021, data were available for 109 economies.

There was one issue for discussion for this enabler: the choice of the denominator to use, where the choice was to divide either by fixed broadband subscriptions or by Internet users. Dividing by Internet users focuses on assessing traffic at the country level, for the average Internet user. However, it would be less precise, as it would combine data from different sources and it would require estimates for the number of Internet users for a fair amount of countries. In addition, no distinction can be made between Internet users using fixed or mobile broadband. Dividing by subscriptions has the advantage that both the numerator and the denominator would come from the same source, being monitored by telecom operators and national regulatory authorities. Traffic per subscription is also a performance indicator used by many regulators. Eventually, the participants agreed to use subscriptions as denominator.

Meaningful connectivity — Infrastructure: indicators considered but not retained

Percentage of households covered by a fixed network

Being covered by a fixed network at home is a necessary condition to contract a fixed broadband subscription. The indicator is defined in the ITU Handbook (ITU, 2020b). Comparability issues were also highlighted by the EGTI subgroup examining the fixed network coverage indicators in 2022 in the context of the long questionnaire indicators review.[14] The data are usually collected by the ICT regulator from operators. At the international level, data are collected from countries by ITU. For 2021, data were available for 66 economies, for 2020-2021 for 71 economies. The indicator is excluded for data availability and quality reasons.

Percentage of population that lives within physical reach of (fiber) nodes on core terrestrial transmission networks

The indicator is relevant as a proxy for infrastructure density or territorial distribution. It was defined by EGTI and approved at the 10th World Telecommunication/ICT Indicators Meeting in 2012, where it was decided that the data would initially be collected through an ITU pilot project, with external collaborators obtaining the data from operators to create interactive transmission maps. The data thus collected would be shared with national regulators or ministries for verification, ahead of their publication. This practice has evolved into the ITU Broadband Map initiative, run by the Infrastructure Division of ITU-D. Data on the nodes are mostly collected by ITU through desk research, and are subsequently validated by telecom and network operators, with Member States's focal points copied on the correspondence. It is possible that some nodes are missing. The calculation of the percentage of population within a certain distance of the nodes is done by ITU, using a variety of (open) sources. Because of limited resources, the data may not be up to date. The indicator was proposed in the Zero draft document. Based on comments received by EGTI/EGH members on the Zero draft and considering that the data for this indicator is not necessarily updated annually, coverage of nodes may be partial, and that various are used sources in addition to official ones, the indicator was rejected.

International bandwidth usage (bit/s) per Internet user

International bandwidth capacity and bandwidth usage indicators provide information about the availability and utilisation of infrastructure for international data linkages (including submarine or overland cables, satellite linkages, etc.). These statistics can also signal the presence of barriers to international connectivity. The indicator is normalised by dividing by the number of Internet users in the country. However, international bandwidth usage measures suffer from several limitations. First, end-user experience (which is a key concern

[14] https://www.itu.int/itu-d/meetings/statistics/wp-content/uploads/sites/8/2022/09/EGTI2022_LQ_Review_Report.pdf

for meaningful connectivity) is not only determined by international, but also by middle-mile and last-mile connectivity. However, ITU is not collecting statistics on many of the middle mile elements that influence international bandwidth usage (such as local cache, off-peak load, presence of CDN). Second, while low values of the indicator can signal lack of connectivity for users, high values can often be biased if a country is a connectivity transit hub. Third, many countries do not collect this indicator, and many are estimating it based on domestic traffic data, thus limiting international comparability. The problem is made worse by the fact that a non-negligible share of traffic is not carried over the open Internet and by a lack of transparency of international cable operators about pricing and usage. For these reasons, this indicator was not considered a suitable candidate for inclusion and was therefore rejected.

Middle-mile and last-mile connectivity

It would be relevant to include measures of middle-mile and last-mile connectivity. One example is statistics on Internet exchange points, such as the number in a country, their size measured in terms of traffic or peering partners, or their environmental footprint. The 13th meeting of EGTI in 2022 recognized both the relevance of statistics on middle-mile connectivity, as well as the need to investigate the feasibility to develop internationally comparable measures, given the limitations of information readily available at sources such as Packet Clearing House or IXPDB. This was added to the work programme of EGTI for 2023, but at this stage, given the limitations, it is premature to propose middle-mile connectivity indicators for inclusion.

Speed of Internet connections

Meaningful use of the Internet requires a fast connection. High quality data on the speed of Internet connections or user experience metrics would be relevant to include in the index, which was highlighted by a few commenters as well. Various data sources exist, such as crowd sourced speed test data from Ookla, OpenSignal, or M-Lab. These are all non-official sources with some limitations (e.g., means of collection, number of observations, country coverage). Consequently, no indicator on the speed of the Internet connection is proposed. Refining existing quality of service (QoS) metrics and developing new ones, notably speed measurement, will be part of EGTI's work agenda.

Meaningful connectivity — Affordability

One of the main barriers preventing people from going online is the cost of the device and/or of the service. Affordability is a critical enabler of meaningful connectivity. There is no indicator on the affordability of Internet-enabled devices for which there is enough internationally comparable data. For the affordability of Internet services, two indicators collected by ITU were initially proposed: the *price of a data-only mobile-broadband basket as a percentage of GNI per capita* and the *price of a fixed broadband basket as a percentage of GNI per capita*.

Based on comments received, the choice of the mobile broadband basket was discussed during the IDI meeting. The original suggestion was to use the data-only mobile broadband basket, which is the simplest to collect and interpret, is fairly representative of the affordability measures based on the other mobile broadband baskets published by the ITU, and is also used for the UN Broadband Commission's policy target. Many participants argued that since consumers in many markets typically use combined voice and data rather than data-only mobile broadband services, a bundle basket was a better fit for the framework. The Secretariat explained that data availability is not a limitation to use any of the options, and it was clarified that in many countries the price for a mobile broadband data and voice basket may not be the price of a "bundle".[15] Participants still preferred the data and voice basket over the data-only mobile broadband basket. The participants also decided to use the *mobile broadband data and voice high-consumption basket* rather than

[15] Bundle in the sense that the services are marketed in one single package. Depending on countries and operators, combined data, voice and text messaging services are offered in various forms, in bundles, packages or add-ons added to base plans. In the 2020 review, EGTI agreed not to apply a restrictive definition but one that corresponds to the cheapest option reaching the minimum allowance and validity thresholds. See methodological details at: https://www.itu.int/en/ITU-D/Statistics/Documents/publications/prices2021/ITU_ICT_Prices_Methodology.pdf.

the low-consumption one, due to its relevance for meaningful connectivity. It was also seen as a reasonable compromise since the high-consumption data and voice basket applies the same 2 GB data allowance threshold as the data-only mobile broadband basket, and correlation analysis reported in the Version 2 document confirmed the high similarity between these two (as well as the fact that the high- and low-consumption mobile data- and voice baskets are nearly identical statistically).

In addition, a clarification was provided by the Secretariat on the necessity of expressing the prices of the baskets as a share of gross national income per capita to measure affordability of a connectivity basket. Using prices, including expressed in international dollars (or power purchasing parity dollars) does not provide a measure of affordability: indeed a price in country X may be lower than in country Y, but if income in X is even lower than in Y, X is less affordable, because someone in X must spend a *higher* share of her income than someone in Y for the same basket.

Meaningful connectivity — Affordability: retained indicators

Mobile data and voice high-consumption basket as a percentage of GNI per capita

The high-consumption data and voice basket reflects typical consumer preference for combined data, voice, and text messaging services, is relevant to the concept of meaningful connectivity, and applies a similar data allowance threshold as the data-only mobile broadband basket.

The indicator is defined in the ITU Handbook (ITU, 2020b); the methodology can also be retrieved from the price methodology on the ITU website[16]. The source of retail price data are the non-promotional advertised prices of selected services for residential customers effective at the time of data collection, from operators with the largest market share in an economy, measured by the number of subscriptions. Data are submitted by countries to ITU, complemented by ITU research. GNI per capita levels are from the World Bank World Development Indicators, referring to the preceding year. For 2021, data are available for 182 economies.

Fixed broadband basket as a percentage of GNI per capita

The affordability of fixed broadband basket measures the share of the average income required for connecting to the Internet from a fixed location. It represents one of the main means of meaningful connectivity for individuals and households. The minimum data threshold of 5 GB was established by EGTI to meet entry-level consumption patterns, while in practice, the cheapest available option in most economies was found to contain unlimited data allowance, making it a relevant indicator.

The indicator is defined in the ITU Handbook (ITU, 2020b); the methodology can also be retrieved from the price methodology on the ITU website. It is one of the core indicators of the Partnership on Measuring ICT for Development. The source of retail price data are the non-promotional advertised prices of selected services for residential customers effective at the time of data collection, from operators with the largest market share in an economy, measured by the number of subscriptions. Data are submitted by countries to ITU, complemented by ITU research. GNI per capita levels are from the World Bank World Development Indicators, referring to the preceding year. For 2021, data are available for 171 economies.

Meaningful connectivity — Affordability: indicators considered but not retained

Data-only mobile broadband basket as a percentage of GNI per capita

The indicator was replaced by a combined data and voice high-consumption basket.

Meaningful connectivity — Device

Access to an Internet-enabled device is required to go online. The index could consider both mobile phones and desktop computers, recognizing that the most basic models of the former are cheaper, while the latter allow for a richer experience. For computers, the indicator considered is *households with access to a computer*.

[16] https://www.itu.int/en/ITU-D/Statistics/Documents/datacollection/IPB_Rules_2022.pdf

For *mobile phones*, the indicator considered is *ownership*, recognizing that mere access to a device imposes constraints, including when and for how long one can be online.

Meaningful connectivity — Device: retained indicator

Percentage of individuals owning a mobile phone

A mobile phone is one of the most common devices used to go online. The indicator is an SDG indicator and one of the core indicators of the Partnership on Measuring ICT for Development. The source is usually ICT household surveys conducted in countries, often by the national statistical office, but sometimes by other entities as well, such as the regulator. At the international level, data are collected from countries by ITU. For 2021, data were available for 47 economies, for 2020-2021 for 59 economies.

Initially, the indicator was excluded because of low data availability. Because there was a broad call for inclusion by EGTI and EGH members, the indicator was included in the Version 1 document, supported by the fact that estimates have already been made for this indicator to calculate aggregates for *Facts and Figures*, a practice which will be continued in coming years, using an established methodology. However, during the consultation with Member States, some countries objected based on data availability. Therefore, the indicator was tabled for discussion. During the meeting, there was a clear majority in favour of including the indicator, therefore it is retained.

There were calls for including ownership of a *smartphone*, instead of mobile phone. In the ITU data collection, smartphone is a subcategory of mobile phone, but unfortunately not enough countries submit disaggregated data – only 26 countries have reported data on smartphone ownership since 2019. In addition, ownership of a mobile phone, even a feature phone, is better than no phone.. For these reasons, overall mobile phone ownership remains the best option for the IDI.

Other comments by Member States concerned the age scope and proposed to only consider the 15+ or 18+ population. Children are indeed less likely to own mobile phones. However, the definition of ownership covers individuals who are in sole possession of a mobile phone. That is, another person (e.g., a parent) may have paid for the phone and any ongoing subscriptions, but if the individual in question has full access to the mobile phone, she/he is considered its owner (see Chapter 7 of the ITU Household Manual, ITU, 2020a).

As a result, ages 18 and older may be too high of minimum age for this indicator. In addition, as countries submit data for different age ranges aligning all countries based on the same in-scope population range is not possible. Although the ITU Household Manual recommends collecting data for all individuals aged 5 and above, many countries do not survey children. The median lower bound for in-scope age was 10 years. This corresponds to the age that children might reasonably be expected to begin to own mobile phones where families have resources.[17]Finally, many countries do not provide ownership data disaggregated by age.

Meaningful connectivity — Device: indicator considered but not retained

Percentage of households with a computer

A computer is one of the devices that allows a user to go online. The indicator is defined in the ITU Household Manual (ITU, 2020a). It is one of the core indicators of the Partnership on Measuring ICT for Development. The source is usually ICT household surveys conducted in countries, often by the national statistical office, but sometimes by other entities as well, such as the regulator. At the international level, data are collected from countries by ITU. For 2021, data were available for 53 economies, for 2020-2021 for 67 economies. The indicator is excluded for data availability reasons.

While data availability was roughly the same as for mobile phone ownership there were not similarly compelling reasons to retain this indicator. Notably, there was not a comparable call for inclusion by Member States during consultations. The percentage of households with a computer may be losing some of its value as an indicator of ICT development. Many activities that required a computer in the past (like emailing, web browsing, online shopping, social networking) can now be performed on a mobile device. In addition, because

[17] See https://childmind.org/article/when-should-you-get-your-kid-a-phone.

of these technological and behavioral changes, estimates have not been made for this indicator in recent years to calculate aggregates for *Facts and Figures.* Retaining this indicator would necessitate the establishment of new and untested methodologies to estimate the percentage of households with a computer.

Meaningful connectivity — Skills

Digital literacy is a requirement for fully leveraging connectivity. **The percentage of individuals with ICT skills** is an indicator of digital literacy. In the absence of sufficient harmonised data for this indicator, two education proxies can be proposed, **Expected years of schooling** and **Mean years of schooling.**

Meaningful connectivity — Skills: Indicator considered but not retained

Percentage of individuals with ICT skills

The percentage of individuals with ICT skills is an indicator of digital literacy. The indicator is an SDG indicator, defined in the ITU Household Manual (ITU, 2020a). It is also one of the core indicators of the Partnership on Measuring ICT for Development. The source is usually ICT household surveys conducted in countries, often by the national statistical office, but sometimes by other entities as well, such as the regulator. At the international level, data are collected from countries by ITU. For 2021, data were available for 61 economies, for 2020-2021 for 69 economies.

Because self-reporting of individuals' ICT skills may be subjective, ICT skills are measured based on whether an individual has recently performed certain activities that require different types of skill. The assumption is that performing these activities implies that one has a certain level of the required skills. At first, these activities were grouped in three broad categories: basic, standard and advanced ICT skills. A subgroup of EGH has been at work since 2018 to group the activities in more relevant categories and to propose an overall score based on the reported activities. As a result of the work of the subgroup, activities are now grouped into five categories of digital skills: communication/collaboration; problem solving; safety; content creation; and information/data literacy. Work is still ongoing to aggregate the data into one overall skills score.

Until the work of the subgroup is finalised, and data availability is sufficient, this indicator cannot be included in the IDI. Therefore, the indicator is deferred to 2027, in the hope and expectation that by then the methodological issues will be resolved and data availability will be sufficient.

Expected years of schooling (school life expectancy)

This indicator is one of the proxies for ICT skills in conjunction with mean years of schooling. The methodology is defined by the UNESCO Institute for Statistics (UIS). For a child of a certain age, expected years of schooling is calculated as the sum of the age specific enrolment rates for the levels of education specified. The part of the enrolment that is not distributed by age is divided by the school-age population for the level of education they are enrolled in and multiplied by the duration of that level of education. The result is then added to the sum of the age-specific enrolment rates. Estimates are made by UNDP for use in the HDI.

Mean years of schooling (ISCED 1 or higher), population 25+ years

This indicator is one of the proxies for ICT skills in conjunction with expected years of schooling. The methodology is defined by the UNESCO Institute for Statistics (UIS). It is defined as the average number of completed years of education of a country's population aged 25 years and older, excluding years spent repeating individual grades. Estimates are made by UNDP for use in the HDI. For 2021, data were available for 190 countries, using the data used for the HDI.

In the previous IDI, three indicators were used as proxy for ICT skills: mean years of schooling, gross enrolment ratio in secondary education and gross enrolment ratio in tertiary education. In the Zero draft document, the two enrolment indicators were proposed, but many EGTI/EGH members objected during the first consultation. In the Version 1 and Version 2 documents of the proposed IDI methodology, *Mean years of schooling* and *Expected years of schooling* were proposed as alternate proxy indicators. The two indicators make up the *Knowledge pillar* of UNDP's Human Development Index and it was proposed to use the HDI data set, which includes estimates for the missing values. During the IDI meeting, there was some support for including these indicators as proxies for ICT skills, but a larger part of the audience considered these indicators as poor proxies for ICT skills. Eventually, the participants recommended not to include any indicators on ICT skills in this iteration of the index.

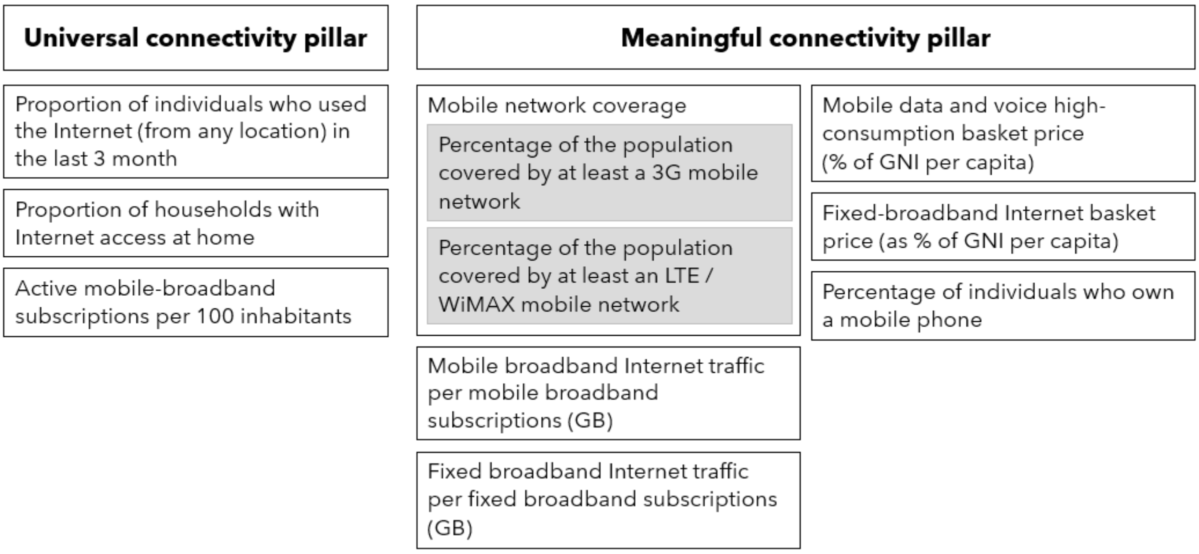
Meaningful connectivity — Safety and security

During the IDI meeting, concerns were expressed that the enabler 'Safety and security' was not captured in the proposed structure and a proposal was made to use ITU's Global Cybersecurity Index (GCI). In response, the Secretariat acknowledged the critical importance of this enabler and explained that the GCI had been considered in the Zero draft but was rejected, and laid out the reasons (also laid out in the Zero draft, Version 1 and Version 2 documents). The GCI assesses countries' commitments to cybersecurity. As such, it does not fit the framework, which focuses on outputs rather than inputs. In addition, the GCI's methodology is still evolving and is not stable yet. Introducing it in the IDI would affect comparability over time, as a change in this indicator may be due to a change in the methodology rather than a change in the performance. Therefore, the inclusion of the GCI was excluded.

Summary of retained indicators

Figure 3 summarizes the indicators retained for inclusion for each UMC dimension.

Figure 3: Structure of the IDI as of 15 June 2023



Country coverage

Based on the list of retained indicators, it is possible to estimate the number of economies the index can cover. Resolution 131 requires that the methodology of the IDI be established so as "to cover a majority of Member States" (*resolves 3*). As explained in the *Data availability and reference year* section, the assessment is based on the criterion of having at least one non-estimated data point available within the period 2020-2021 in the case of the 2023 edition (Table 4).

Table 4: Data availability for the selected indicators

	Category/Code	Indicator	countries with data available ≥2021	≥2020
Universal connectivity				
1	yHH7	Proportion of individuals who used the Internet (from any location) in the last 3 months	81	**94**
2	xHH6	Proportion of households with Internet access at home	81	**94**
3	i911mw	Active mobile-broadband subscriptions per 100 inhabitants	160	**170**
Meaningful connectivity - Infrastructure				
4	i271G	Percentage of the population covered by at least a 3G mobile network	158	**170**
5	I271GA	Percentage of the population covered by at least an LTE/WiMAX mobile network	156	**168**
6	i136mwi_subs	Mobile broadband Internet traffic per mobile broadband subscription (GB)	131	**143**
7	i135tfb_subs	Fixed broadband Internet traffic per fixed broadband subscription (GB)	109	**115**
Meaningful connectivity - Affordability				
8	i271mb_high_ts_GNI	Mobile data and voice high-consumption basket (as a % of GNI per capita)	182	**185**
9	i154_FBB_ts_GNI	Fixed-broadband Internet basket price (as % of GNI per capita)	171	**175**
Meaningful connectivity – Device				
10	xHH18	Percentage of individuals owning a mobile phone	47	**59**

Estimating data points adds uncertainty to the calculation of index scores. By setting a higher threshold for data availability, the number of data points to be estimated decreases (implying that the index would be more robust), but so does the number of economies for which the index can be computed. This requires striking a balance. As Table 5 shows, setting the country inclusion threshold at 70 per cent of indicators available would allow 130 economies to be included. In the extreme case where no estimates would be used, the index could be computed for just 40 economies.

Table 5: Number of economies that can be included in the index with various thresholds

Economy inclusion threshold (% of 10 indicators available in the 2020-2021 reference period)	**50%**	**60%**	**70%**	**80%**	**90%**	**100%**
Nr. of economies meeting the threshold requirement	165	149	130	89	75	40
Nr. of missing data points to be estimated	340	260	184	61	33	0
% of total data points to be estimated	*21%*	*17%*	*14%*	*7%*	*4%*	*0%*

The inclusion threshold is set to 50 per cent. That is, an economy would be included if official data is available for at least 50 per cent of the indicators of the index. With this threshold, and based on data availability as of January 2023, 165 economies could be included in the index.[18]

Statistical assessment of the retained indicators

As discussed above, the retained indicators meet the selection criteria (availability, reliability, relevance, etc.). In addition, an indicator must have certain statistical properties both on its own and vis-à-vis the other indicators of the index in order to add relevant information to the overall index score. This following sections discuss and build on the results of several statistical analyses that aim to determine how each selected indicator fits in the index.

[18] A benefit of an index without ranking is to allow for partial assessment of countries: a country that would normally be excluded for not meeting the overall data availability criterion, could still be assessed on selected components of the index for which sufficient data exists, even though it would not get an overall index score. Without ranking, the inclusion of this country in selected components would be without consequence for other countries. This alternative to outright exclusion would allow to increase the number of countries studied and may incentivise countries to improve data availability.

Specifically, the analyses aim to:

- identify the presence of outliers and recommend treatment methods;
- identify potential constraints in the explanatory power of indicators; and
- explore the statistical association between a set of indicators and the latent structure of the dataset.

The analyses entail an in-depth look at the data, making use of two statistical tools: first, exploring each variable separately and describing them through their descriptive statistics (such as mean, median, min, max, among others), followed by a correlation analysis to explore the statistical relationships between indicator pairs and groups.

The assessments are conducted along the subsequent steps (outlier detection and treatment, normalization, weighting and aggregation) and provide additional information to help better interpret and understand the strengths and weaknesses of the indicators selected on a conceptual basis. The assessments are an integral part of the iterative process of indicator selection and confirmation that ultimately aims at ensuring that the framework is both conceptually and statistically coherent.

4 Identifying and treating outliers and missing data (step 4)

All indicators retained for inclusion contain missing values and, in some cases, outlying values. In this step, we ensure that IDI scores can be computed based on a statistically robust dataset. This involves identifying and treating outliers and setting goalposts when needed, and defining the strategy for treating missing values.

Identifying outliers

An indicator is a useful benchmark if it can meaningfully distinguish performance across units (i.e., economies in the present case) and over time. From a statistical perspective, the range of values (the distance between the minimum and maximum) should not be too narrow, and the distribution not too skewed or peaked (a case when the bulk of the values is concentrated within a small range, with some outlying values further apart). The presence of outliers is particularly problematic in the context of composite indicators. Outlying values are not necessarily errors, but if present in component indicators of a composite indicator, they can significantly bias aggregation results. Outliers would not only become unrealistic or unintended targets, but also imply that a significant portion of the data range will remain empty, while small, marginal differences between countries may be inflated or larger differences underestimated. They can also bias statistical coherence analysis. Identifying and treating outliers is therefore essential in the process of developing an index.[19] Various methods exist to treat outliers, depending on the nature of the data, e.g., applying a log transformation or trimming the distribution by applying caps.

Table 6 reports key descriptive statistic including the number of observations (i.e., economies) for the reference period 2020-2021 and information on range and distribution (minimum and maximum values, mean, standard deviation, median and the 25th and 75th percentile – the range between which half of the observations can be found), as well as skewness and kurtosis coefficients (measures of difference from normal distribution).

[19] There is no single definition for outliers (Aguinis et al, 2013), it depends on the nature of the indicators and the measurement purpose. As a rule of thumb, composite indicator development practitioners typically identify outliers when the absolute skewness (a measure of distribution asymmetry) exceeds 2.0 and kurtosis (a measure of the weight of the tails relative to the centre of the distribution) exceeds 3.5, or if kurtosis alone exceeds 10 (see European Commission, 2019).

Table 6: Descriptive statistics for the retained indicators

	Code	Indicator	N	N/ 196	Min	Max	Mean	St.dev.	25th pctile	Median	75th pctile	Skew.	Kurt.
Universal connectivity													
1	yHH7	Proportion of individuals who used the Internet (from any location) in the last 3 months	94	48%	6.1	100.0	80.3	18.6	75.6	84.8	91.9	-2.1	5.0
2	xHH6	Proportion of households with Internet access at home	94	48%	11.9	100.0	81.3	18.8	79.6	87.3	94.0	-1.7	2.6
3	i911mw	Active mobile-broadband subscriptions per 100 inhabitants	170	87%	2.6	285.1	84.1	43.5	54.5	84.3	107.6	1.0	3.2
Meaningful connectivity – infrastructure													
4	i271G	% of the population covered by at least a 3G mobile network	170	87%	15	100.0	92.2	14.1	92.2	98.4	99.9	-2.9	9.5
5	i271GA	% of the population covered by at least an LTE/WiMAX mobile network.	168	86%	0	100.0	83.6	24.3	80.0	96.0	99.3	-1.7	1.7
6	i136mwi_subs	Mobile broadband Internet traffic per mobile broadband subscriptions (GB)	143	73%	0	1’104.8	93.8	126.0	28.4	62.9	113.5	4.7	31.3
7	i135tfb_subs	Fixed broadband Internet traffic per fixed broadband subscriptions (GB)	115	59%	0	10’484.5	2’273.9	1’892.0	922.3	2’029.7	3’260.7	1.5	3.7
Meaningful connectivity – affordability													
8	i271mb_high_ts_GNI	Data-only mobile-broadband basket price (as % of GNI per capita)	185	94%	0.1	56.9	6.6	9.3	1.1	2.9	7.5	2.6	7.6
9	i154_FBB_ts_GNI	Fixed-broadband Internet basket price (as % of GNI per capita)	175	89%	0.3	164.2	10.0	18.6	1.4	3.5	11.0	4.9	32.3
Meaningful connectivity – device													
10	xHH18	Percentage of individuals owning a mobile phone	59	30%	41.2	100.0	85.4	15.2	75.4	91.3	97.4	-1.2	0.8

Note: Statistics are based on data for 2021 or, where unavailable, 2020.

The descriptive statistics reveal two issues in the dataset: the presence of outliers and the concentration of variation within a very limited range.

- The values for the indicator *Mobile broadband penetration (i911mw)* range from 2.6 to a maximum of 285 subscriptions per 100 inhabitants. Apart from eight countries, values are less than 150 subscriptions per 100 inhabitants. Setting a cap is justified from a statistical as well as a conceptual standpoint to set a more realistically achievable target and allow for a more meaningful cross-country comparison.
- Considering the *mobile broadband coverage* indicators, the *percentage of population covered by at least a 3G mobile network* (i271G) has limited discriminatory power (differences between country performance are often in the decimal digits). Apart from a few outliers in the low range, three-fourth of the observations are found between 92 and 100 per cent. Country performance is somewhat more dispersed for the other indicator, *percentage of population covered by at least an LTE-WiMAX mobile network* (4G, or i271GA). Outlier treatment is not warranted if the two indicators are combined in an aggregate 'coverage' indicator.
- Outliers were detected for both Internet traffic indicators. The distribution of *Mobile broadband traffic per subscription* (i136mwi_subs) values is highly skewed, and while the median is 62.9, around 5 per cent of the countries reported values between 265 to 681 GB per subscription. Such a skewed distribution warrants capping the indicator. A goal post, must be forward looking, considering that Internet traffic is growing by 20 per cent annually.
- *Fixed broadband traffic per subscription* (i135tfb_subs) values are more evenly spread compared to mobile broadband traffic per subscription. However, a few outlying values require treatment before including it in the aggregation for a composite indicator. The median value is 2,030 GB/user, and 95 per cent of the observations are below 5,250 GB/user. Like the previous indicator, setting a cap should take into consideration the fact that traffic is expected to increase for the next four years.
- Both affordability indicators have a very skewed distribution, with a median of 2.9 for mobile and 3.5 per cent of GNI per capita, and 95 per cent of the observations less than 26 and 42 per cent of GNI per capita for mobile and fixed broadband, respectively. However, outliers reach up to 57 and 164 per cent of GNI per capita, respectively. Trimming the distribution is advisable to increase variance across countries, especially because this is an indicator where, contrary to others, the best performer country has the lowest values, thus the direction will have to be reversed at the normalization step.

Treating outliers

Outlier treatment should take into consideration any thresholds and goalposts defined for indicators on a conceptual basis. Such limits may effectively cap the distribution for concerned indicators to acceptable ranges making other adjustments, such as winsorization, unnecessary.[20] Therefore, the data distributions will be re-examined after defining goalposts and thresholds.

Goalposts and thresholds

Depending on the indicator, the goalpost may be a policy target or aspiration, the maximum possible value, or a number derived from statistical analysis of the distribution (e.g., 90th or 95th percentile). Table 7 also shows indicative thresholds and goalposts for the proposed indicators.

A few observations:

- When setting goalposts for the universality indicators, the concept of universality must be interpreted loosely. For individual usage, it is neither expected nor desirable that all children use the Internet. Indeed, approaches to bringing children online varies across geographies. When picking the goalpost, one must also consider that, among the population, some individuals do not want to use the Internet, even if they have access to it and can afford it. For these reasons, the goalpost for Internet users

[20] Earlier version of this document considered two separate steps for outlier treatment and setting goalposts and thresholds. A streamlined approach is followed here, as this requires less intervention.

should be set at a value slightly below the 100% mark. The expert group discussion agreed that the goalpost is set at 95%. This means that a country with a share of 95% or more will get a score (i.e., a normalised value) of 100 on this indicator. The same approach would apply to the indicator "Individuals owning a mobile phone", part of the meaningful connectivity enabler "Device". While universality is the objective, the goalpost should be set at a lower value, because some people may not want to own a device. The same logics applies to the indicator "Households with internet access", reflecting the reality that some households may not want to have access at home and accounting for possible measurement errors.

- For the two traffic indicators, goalposts are defined statistically driven. In order to avoid setting unrealistic targets, the goalposts are set at the 95th percentile of the observed values. According to the 2021 data, these values are 254 and 5'083 GB per subscription per year for mobile and fixed broadband, respectively. Considering the double-digit annual growth of global median traffic, the caps are set respectively at 500 and 10'000 GB.
- In the case of the affordability indicators, where a higher cost corresponds to a worse outcome, the goalpost is lower than the threshold. While initially the goalpost for the affordability indicators corresponded to the 2 per cent policy target of the Broadband Commission, many participants argued for a lower value to better distinguish country performance and motivate further affordability improvements over time. Consequently, the revised goalpost is set at 1 per cent of GNI per capita.

As indicated in the right column of Table 7, applying the goalposts and thresholds adjusts the distribution in a way that no additional outlier treatment is necessary for all but the two traffic indicators. These two indicators display logarithmic distributions, so the appropriate adjustment is applying a logarithmic transformation on the data.[21]

Table 7: Indicative goal posts and thresholds, and outlier treatment

Code	Indicator	Indicative threshold	Indicative goalpost	Additional outlier treatment
yHH7	Proportion of individuals who used the Internet	0%	95%	Not needed
xHH6	Proportion of households with Internet access at home	0%	95%	Not needed
i911mw	Active mobile-broadband subscriptions per 100 inhabitants	0%	95th percentile	Not needed
i271G	% of the population covered by at least a 3G mobile network	0%	100%	Not needed if the two coverage indicators are combined
i271GA	% of the population covered by at least an LTE/WiMAX mobile network.	0%	100%	
i136mwi_subs	Mobile broadband Internet traffic per mobile broadband subscriptions (GB)	Min. value	95th percentile, projected	apply log transformation
i135tfb_subs	Fixed broadband Internet traffic per fixed broadband subscriptions (GB)	Min. value	95th percentile, projected	apply log transformation
i271mb_high_ts_GNI	Data-only mobile-broadband basket price (as % of GNI per capita)*	95th percentile	1%	Not needed
i154_FBB_ts_GNI	Fixed-broadband Internet basket price (as % of GNI per capita)*	95th percentile	1%	Not needed
xHH18	Percentage of individuals owning a mobile phone**	0%	95%	Not needed

* The direction of the affordability indicators is reversed, hence score of 100 will be assigned to values *below* the goal post. Scores of 0 will be assigned to values *above* the threshold.

[21] As standard practice, in order to retain valid 0's, the values are adjusted by a constant of 1 (a negligible value) before calculating the natural logarithm.

Estimating missing data

As explained in the *Country inclusion* section and shown in Table 5 above, a relatively less stringent data availability threshold allows the inclusion of more economies, however, many of them will have missing values for several indicators. This inevitably affects the accuracy of the assessment of the IDI for those countries. Values for ITU indicators that were not submitted by countries in the reference period 2020-2021 will be estimated, when possible, using a model-based approach tailored to the indicator.

The models used to estimate missing values for indicators typically collected in ICT household surveys are based on a diverse range of widely available national indicators on mobile-broadband subscriptions, ICT affordability, GNI per capita and so on, and accounting for their changes over time. In addition to data submitted by Member States, other sources may be used to obtain data and/or cross-check estimates.

In other cases, univariate time series models (such as autoregressive integrated moving average (ARIMA) models may be applied to historical data to predict missing recent values.

It is very difficult to compute reliable estimates for missing traffic data. Instead, it was agreed during the meeting that missing data for these two indicators would not be modelled, but would be imputed using a hot deck imputation method. (The hot deck imputation is a method that involves replacing missing values of one or more variables for a "non-respondent" unit with observed values from another "respondent" unit that is similar ("the nearest neighbor") to the non-respondent with respect to characteristics observed by both cases. The advantages of this method are the use of plausible values (actually observed ones) for imputation, and the fact that it is less sensitive to model misspecification (as opposed to using parametric models). At the same time, it requires good matches between non-respondents and respondents based on the available information.) These estimates will be used to calculate the index, but the underlying estimates will not be published (unlike the estimates for the other indicators that are used for the IDI and for *Facts and Figures*).

Consistent with the iterative nature of the IDI development, the dataset containing estimated values will be subject to outlier detection and treatment as needed. Adding model-based estimates – especially considering that data are not missing at random – will likely change distributions, which will have an impact on thresholds.

5 Normalization, weighting, and aggregation (step 5)

Normalization

The indicators selected are measured on various scales and expressed in different units. Normalization is applied to bring all indicators on a common scale. The most used method is the min-max approach, which rescales indicators onto an identical range of 0 to 100 by subtracting the minimum value for the given indicator across all economies from each value and dividing by the range of the indicator values. It is relatively easy to apply and to interpret. Formally, we have:

$$\mathrm{score}_{i,c} = \frac{\mathrm{value}_{i,c} - \mathrm{threshold}_i}{\mathrm{goalpost}_i - \mathrm{threshold}_i} \times 100$$

where $\mathrm{value}_{i,c}$ is the value of country c on indicator i, $\mathrm{threshold}_i$ is the minimum value for indicator i and $\mathrm{goalpost}_i$ corresponds to the target value for indicator i. If a value is at or below the threshold value, the corresponding score is 0; if a value is at or above the goalpost, the score is 100.

In the case of the affordability indicators, the directionality is reversed using the following formula:

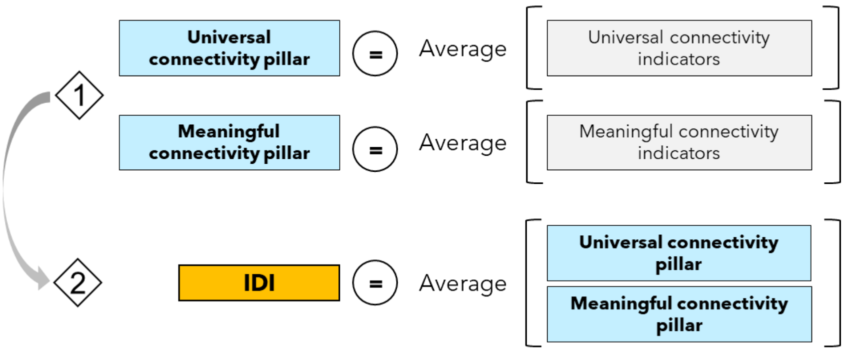
$$\mathrm{score}_{i,c} = \frac{\mathrm{goalpost}_i - \mathrm{value}_{i,c}}{\mathrm{threshold}_i - \mathrm{goalpost}_i} \times 100$$

Weighting and aggregation

Conceptually, there are two groups of indicators: universal connectivity indicators and meaningful connectivity (UMC) indicators. The correlation analysis (presented in detail in Annex 3 of the Version 1 document) revealed that all indicators are positively correlated with one another. This suggests that they measure different aspects of the multidimensional concept of UMC. While no significant trade-offs were identified between the indicators, some compensability cannot be ruled out (i.e., weakness in one indicator may be compensated by strength in another).

For the weighting scheme, one intuitive and neutral approach is to mirror the two dimensions of the UMC concept, by averaging the scores of the Universal connectivity pillar and of the Meaningful connectivity pillar. The scores of the two pillars would be the average of the individual indicators included in each pillar, so that the pillar score provides a balanced summary of the underlying information (Figure 4). In the absence of a clear conceptual and statistical justification, this neutral approach consisting in applying equal weights at each level of aggregation (i.e., pillar level and overall level) should be preferred. With this approach, the assumption is that the main conceptual components are the two pillars and that the individual indicators within each pillar contribute in a similar extent to the performance of the pillar. The subsequent statistical analysis does not reject this neutral and intuitive approach. For these reasons, this approach was proposed in the Version 1 document and again in the Version 2 document. The statistical analysis does not reject this neutral and intuitive approach.

Figure 4: Weighting and aggregation



A possible alternative weighting approach discussed by participants was to consider individual indicators as the main conceptual components (instead of the two pillars). In this case, the overall index score would be the average of the scores of the individual indicators. Each indicator would have the same weight in the overall IDI, unlike the other approach where the implicit weight of individual indicator depends on the number of indicators in the pillar, effectively attributing a stronger weight to the meaningful connectivity indicators. After some debate, the participants agreed to use the first approach.[22]

[22] There is no empirical evidence to justify departing from one of the two approaches above for the weighting scheme.

Universal connectivity pillar

The pillar consists of three indicators. Correlation analysis and the preliminary results of a principal component analysis (PCA)[23] confirm that the three indicators capture a single latent dimension strongly associated with the four indicators, each of which contribute in a fairly similar way to the aggregate measure. This suggests that equal weighting can be applied in this pillar.

Meaningful connectivity pillar

The pillar consists of seven indicators, two of which – *percentage of the population covered by at least a 3G and 4G (LTE/WiMAX) mobile network* – are combined to a mobile broadband coverage score, applying 0.4 and 0.6 as the weights, respectively. This is based on feedback from the IDI forum and expert advice and takes into consideration that having at least 4G technology allows for a more meaningful online experience than having at least 3G technology. It is noted that in practice, the two networks often overlap, in which case often 3G is used for voice and 4G for data communication.

The meaningful connectivity indicators positively correlate with one another, but the structure shows heterogeneity among the indicators. In brief, a moderate compensability was found between the two broadband traffic indicators and the rest of the indicators in the pillar[24] (even after outliers are removed). However, there is no clear statistical justification for departing from the most intuitive approach of applying equal weights to compute the average of the indicator scores in the pillar. One conclusion, in any case, is that it is reasonable to consider the different indicators also by themselves for a comprehensive benchmarking of meaningful connectivity, in addition to using pillar summary scores and the overall aggregate index. This helps understand strengths and weaknesses for each country, delivering more nuanced information for policies.

The IDI scores will be computed by taking the simple average of the meaningful and universal connectivity scores. This *ex-ante* assessment on the structure should, in any case, be revisited in a statistical coherence analysis after the calculation of aggregate scores and after outlier treatment and normalization, as the structure may need some refinements to ensure that the statistical soundness of the IDI. This upcoming step will also take into consideration the results of the statistical audit carried out by the Competence Centre on Composite Indicators of the European Commission's Joint Research Centre.

6 Conclusions and next steps

After the Zero draft, Version 1 and Version 2 documents, this Version 3 document considers the outcomes of the joint EGTI/EGH meeting on the IDI held in June 2023, which allowed to address the remaining issues identified earlier in the process. This document represents the document for the final meeting on the IDI to be held in September 2023.

The document first introduces the approach for developing a conceptually relevant and statistically robust composite indicator – or index. This approach provides the structure for the rest of the document. The first step consists in defining the conceptual framework. The concept of universal and meaningful connectivity (UMC) – the possibility for everyone to enjoy a safe, satisfying, enriching, productive and affordable online experience – is the framework of choice: It is highly relevant as it captures both the quantitative aspects (universal) and qualitative aspect of connectivity (meaningful). As such it reflects the need to go beyond 'connecting' everyone. The concept is also rooted in earlier editions of the IDI and reflects ITU's priorities. In step 2, the conceptual framework of UMC and a set of selection criteria – such as reliability, availability, quality

[23] Principal component analysis is applied to explore the underlying multivariate structure of a set of indicators and helps identify latent dimensions. Only the main conclusions from the analyses are reported in this document, as it is based on a restricted set of economies for which all indicators are available.
[24] Preliminary PCA results on a very restricted number of observations indicate the presence of a second component, associated with the traffic indicators.

– guided the identification of indicators for potential inclusion from a large universe of ICT indicators. In step 3, the data quality of the 10 selected indicators was assessed.

In step 4, statistical analyses of the retained indicators in isolation as well as the analysis of correlation patterns help identify and treat outliers and missing data, and provide a first indication of how well an aggregate index may represent the information shown by the selected indicators. Step 5 consists in transforming the indicators by applying a linear transformation and using thresholds and goalposts, to produce unit-less scores ranging from 0 to 100, with 100 representing the ideal state. The scores for each indicator can then be aggregated to produce the pillar scores, which in turn are combined to produce the overall score.

A statistical analysis reveals that the proposal is statistically sound. Following the aggregation of indicators into a universal connectivity pillar and a meaningful connectivity pillar, together with an overall index will lead to an IDI that will be a fair summary of the information contained in the component indicators part of the baseline proposal. Nevertheless, by its nature, the IDI simplifies the richness of information contained in the individual indicators.

Limited data availability and quality are enormous constraints for the development of the index and impose trade-offs between the depth, completeness, and timeliness of the assessment on the one hand and country coverage on the other. The methodology needs to consider these constraints and trade-offs, while ensuring conceptual relevance and statistical soundness, as per Resolution 131.

The IDI covers certain important aspects of universal and meaningful connectivity, but not all. There are many concepts for which no indicator exists. For other concepts, indicators exist, but data availability or reliability is insufficient or the sources are not official ones. The assessment the IDI is therefore partial. It must be complemented by additional data, information and evidence for a more accurate picture of a country's state of connectivity. In this context, the dozens of ICT indicators maintained by ITU that do not meet the eligibility criteria for inclusion in the IDI are as relevant as ever. In fact, some of the most insightful ITU indicators have the lowest data availability and cannot be included in the IDI. Member States must strive to collect as many of them as possible on a regular basis.[25] The IDI indicators alone – especially when condensed to one single number – will not provide all the necessary information for policymaking.

In its current composition, the index allows to cover *approximately* 165 economies, thus meeting the requirement in Resolution 131 to cover a majority of Member States. A*pproximately* 21% of data points would be estimated, a ratio in line with the requirement in Resolution 131 to rely primarily on official data provided by Member States.

Finally, the complex process of developing an index is also an iterative one. The subsequent steps (6-9) may impose some minor adjustments to the choices made so far to ensure that the final methodology that will be submitted to Member States for their approval is sound.

[25] The technological, policy or market relevance of indicators were recently highlighted in the report of the EGTI subgroup on the review of the indicators collected in the ITU World Telecommunication/ICT Indicators Long Questionnaire, as well as in similar work carried out by the EGH.

References

Aguinis, H., Gottfredson, R. K., & Joo, H. (2013), Best-Practice Recommendations for Defining, Identifying, and Handling Outliers, *Organizational Research Methods*, 16(2), 270–301, https://doi.org/10.1177/1094428112470848

European Commission (EC) (2019), *Your 10-Step Pocket Guide to Composite Indicators & Scoreboards,* https://knowledge4policy.ec.europa.eu/sites/default/files/10-step-pocket-guide-to-composite-indicators-and-scoreboards.pdf

International Telecommunication Union (ITU) (2020a), *Manual for Measuring ICT Access and Use by Households and Individuals*, 2020 Edition, https://www.itu.int/en/ITU-D/Statistics/Pages/publications/manual.aspx

ITU (2020b), *Handbook for the Collection of Administrative Data on Telecommunications/ICT*, 2020 Edition, https://www.itu.int/en/ITU-D/Statistics/Pages/publications/handbook.aspx

ITU and United Nations Office of the UN Secretary-General's Envoy on Technology (UN OSET) (2022), *Achieving Universal and Meaningful Digital Connectivity. Setting a Baseline and Targets for 2030*, https://www.itu.int/itu-d/meetings/statistics/wp-content/uploads/sites/8/2022/04/UniversalMeaningfulDigitalConnectivityTargets2030_BackgroundPaper.pdf

Organisation for Economic Co-Operation and Development (OECD) (2008), *Handbook on Constructing Composite Indicators: Methodology and User Guide*, https://www.oecd.org/sdd/42495745.pdf

UNESCO Institute for Statistics (UIS) (2019), *SDG 4 Data Digest. How to Produce and Use the Global and Thematic Education Indicators*, https://uis.unesco.org/sites/default/files/documents/sdg4-data-digest-2019-en_0.pdf

UNESCO-UIS, OECD and Eurostat (2020), *UOE data collection on formal education. Manual on concepts, definitions and classifications*, https://uis.unesco.org/sites/default/files/documents/uoe-data-collection-manual-2020-en.pdf

United Nations Conference on Trade and Development (UNCTAD) (2021), *Manual for the Production of Statistics on the Digital Economy*, 2020 Revised Edition, https://unctad.org/webflyer/manual-production-statistics-digital-economy-2020

VERSION 3 – NOT FOR CIRCULATION

Annex 1: Indicative timeline for the development of the ICT Development Index (IDI) 2023

As of 31 July 2023.

Mid-February
Circular L describing the process and timeline for developing and publishing the IDI

Notification to EGTI/EGH members inviting to comment on Zero draft on dedicated forum

Early March
Invitation letter to the IDI meeting with registration link

Mid-April
Circular inviting Member States to comment on Version 1; links to compilation of comments from EGTI/EGH members and Secretariat's response

January 2023 | February | March | April

Secretariat prepares Zero draft for Index methodology to facilitate the development process

Consultation of EGTI/EGH members: comments on zero draft are submitted via a new dedicated discussion forum (4 weeks)

Secretariat compiles and responds to comments from EGTI/EGH members and revises methodology based on comments as appropriate (Version 1)

Consultation of MS: comments are submitted via the same discussion forum (4 weeks)

Late May
Circular with links to: compilation of comments from MS and Secretariat's responses; revised IDI methodology (Version 2); and list of outstanding issues to be discussed at joint EGTI/EGH meeting.

13-15 June
Joint EGTI/EGH Meeting on the IDI to discuss and seek to resolve outstanding issues (virtual format)

14 July
Council 2023

Early August
Circular with Version 3

May | June | July | August

Secretariat compiles and responds to comments from MS, and revises the index methodology (Version 2)

Preparation of EGTI/EGH meeting

Based on outcomes of EGTI/EGH meeting, Secretariat updates methodology (Version 3)

Audit of the index methodology by the Composite Indicators unit at the Joint Research Centre of the European Commission

18-19 September
Joint EGTI/EGH session on the IDI during the annual meetings of EGTI and EGH (18-21 September)

Early October
Circular to Member States with final IDI methodology asking MS if they approve the IDI methodology and if they want to opt out from the 2023 edition (assuming the methodology is approved).

Late October
Circular announcing results of consultation

Mid-December
If methodology is approved:
Launch of the IDI 2023

September | October | November | December

Finalisation of methodology and preparation of consultation

Member States respond to the two questions

If methodology is approved: Computation of the index and analysis

If methodology is approved: Editing and layout of IDI brief

Annex 2: Data availability by economy and indicator

Economy (ISO code)	Proportion of individuals who used the Internet (from any location) in the last 3 months yHH7	Proportion of households with Internet access at home xHH6	Active mobile-broadband subscriptions per 100 inhabitants i911mw	Percentage of the population covered by at least a 3G mobile network i271G	Percentage of the population covered by at least an LTE/WiMAX mobile network. i271GA	Mobile broadband Internet traffic per mobile broadband subscriptions (GB) i136mwi_subs	Fixed broadband Internet traffic per fixed broadband subscriptions (GB) i135tfb_subs	Mobile data and voice high-consumption basket (as % of GNI per capita) i271mb_high_ts_GNI	Fixed-broadband Internet basket price (as % of GNI per capita) i154_FBB_ts_GNI	Percentage of individuals who own a mobile phone* xHH18	Indicators available for the reference period Number	Share	>50%?
Afghanistan (AFG)			2020	2020	2020	2020	2020	2021	2021		7	70%	Y
Albania (ALB)	2021	2021	2021	2021	2021	2021	2021	2021	2021		9	90%	Y
Algeria (DZA)			2021	2021	2021	2021	2021	2021	2021		7	70%	Y
Andorra (AND)			2021	2021	2021						3	30%	N
Angola (AGO)			2021	2021	2021	2021	2021	2021	2021		7	70%	Y
Antigua and Barbuda (ATG)			2020	2020	2020			2021	2021		5	50%	Y
Argentina (ARG)	2021	2021	2020	2021	2021	2021		2021	2021	2021	9	90%	Y
Armenia (ARM)	2021	2021	2021	2021	2021	2021	2021	2021	2021		9	90%	Y
Australia (AUS)			2021	2021	2021	2021	2021	2021	2021		7	70%	Y
Austria (AUT)	2021	2021	2021	2021	2021	2021		2021	2021		8	80%	Y
Azerbaijan (AZE)	2021	2021	2021	2021	2021	2021	2021	2021	2021	2021	10	100%	Y
Bahamas (BHS)			2020	2020	2020			2021	2021		5	50%	Y
Bahrain (BHR)	2021	2021	2021	2021	2021	2021	2021	2021	2021	2021	10	100%	Y
Bangladesh (BGD)	2021	2021	2021	2021	2021	2021	2021	2021	2021	2021	10	100%	Y
Barbados (BRB)			2021	2021	2021	2021		2021	2021		6	60%	Y
Belarus (BLR)	2021	2021	2021	2021	2021	2021	2021	2021	2021	2021	10	100%	Y
Belgium (BEL)	2021	2021	2021	2021	2021	2021	2021	2021	2021	2021	10	100%	Y
Belize (BLZ)		2021						2021	2021		3	30%	N
Benin (BEN)			2021	2021	2021	2021		2021	2021		6	60%	Y
Bhutan (BTN)	2021	2021	2021	2021	2021	2020		2021	2021	2021	9	90%	Y
Bolivia (Plurinational State of) (BOL)	2021	2021	2021	2021	2020			2021	2021	2020	8	80%	Y
Bosnia and Herzegovina (BIH)	2021	2021	2021	2021	2021	2021	2020	2021	2021		9	90%	Y
Botswana (BWA)			2021	2021	2021	2021	2021	2021	2021		7	70%	Y
Brazil (BRA)	2021	2021	2021	2021	2021	2021	2021	2021	2021	2021	10	100%	Y
Brunei Darussalam (BRN)			2021	2021	2021		2021	2021	2021		6	60%	Y
Bulgaria (BGR)	2021	2021	2021	2021	2021	2021	2021	2021	2021		9	90%	Y
Burkina Faso (BFA)			2021	2021	2021	2021		2021	2021		6	60%	Y
Burundi (BDI)			2021	2021	2021	2021	2021	2021			6	60%	Y
Cabo Verde (CPV)			2021	2021	2021	2021	2021	2021	2021		7	70%	Y
Cambodia (KHM)			2021	2021	2021	2021		2021	2021	2020	7	70%	Y
Cameroon (CMR)			2021	2021	2021	2020	2020	2021	2021		7	70%	Y

Economy (ISO code)	Proportion of individuals who used the Internet (from any location) in the last 3 months yHH7	Proportion of households with Internet access at home xHH6	Active mobile-broadband subscriptions per 100 inhabitants i911mw	Percentage of the population covered by at least a 3G mobile network i271G	Percentage of the population covered by at least an LTE/WiMAX mobile network. i271GA	Mobile broadband Internet traffic per mobile broadband subscriptions (GB) i136mwi_subs	Fixed broadband Internet traffic per fixed broadband subscriptions (GB) i135tfb_subs	Mobile data and voice high-consumption basket (as % of GNI per capita) i271mb_high_ts_GNI	Fixed-broadband Internet basket price (as % of GNI per capita) i154_FBB_ts_GNI	Percentage of individuals who own a mobile phone* xHH18	Indicators available for the reference period Number	Share	>50%?
Canada (CAN)	2020	2020	2021	2021	2021	2021	2021	2021	2021		9	90%	Y
Central African Rep. (CAF)								2021			1	10%	N
Chad (TCD)			2021	2021	2021	2021		2021			5	50%	Y
Chile (CHL)			2021	2021	2021	2021	2021	2021	2021		7	70%	Y
China (CHN)	2021		2021	2021	2021	2021	2021	2021	2021	2020	9	90%	Y
Colombia (COL)	2021	2021	2021	2021	2021	2021	2021	2021	2021	2021	10	100%	Y
Comoros (COM)			2021	2021	2021	2021	2021	2021	2021		7	70%	Y
Congo (Rep. of the) (COG)			2021	2021	2021	2021	2021	2021	2021		7	70%	Y
Costa Rica (CRI)	2021	2021	2021	2021	2021	2021	2021	2021	2021		9	90%	Y
Côte d'Ivoire (CIV)			2021	2021	2021	2021	2021	2021	2021		7	70%	Y
Croatia (HRV)	2021	2021	2021	2021	2021	2021	2021	2021	2021	2020	10	100%	Y
Cuba (CUB)	2021	2021	2021	2021	2021	2021	2021	2020	2020	2021	10	100%	Y
Cyprus (CYP)	2021	2021	2021	2021	2021	2021	2021	2021	2021		9	90%	Y
Czech Republic (CZE)	2021	2021	2021	2021	2021	2021	2021	2021	2021	2021	10	100%	Y
Dem. People's Rep. of Korea (PRK)											0	0%	N
Dem. Rep. of the Congo (COD)			2021	2021	2021	2021		2021			5	50%	Y
Denmark (DNK)	2022	2022	2021	2021	2021	2021	2021	2021	2021		9	90%	Y
Djibouti (DJI)			2021	2021	2021	2021	2021	2021	2021		7	70%	Y
Dominica (DMA)			2021	2021	2021	2020		2021	2021		6	60%	Y
Dominican Rep. (DOM)	2021	2021	2021	2021	2021	2021	2021	2021	2021	2021	10	100%	Y
Ecuador (ECU)	2022	2022	2021	2021	2021	2021	2021	2021	2021	2022	10	100%	Y
Egypt (EGY)	2022	2022	2021	2021	2021	2021	2021	2021	2021	2022	10	100%	Y
El Salvador (SLV)	2020	2020	2021	2021	2021			2021	2021		7	70%	Y
Equatorial Guinea (GNQ)									2021		1	10%	N
Eritrea (ERI)											0	0%	N
Estonia (EST)	2021	2021	2021	2021	2021	2021		2021	2021		8	80%	Y
Eswatini (SWZ)			2021	2021	2021			2021	2021		5	50%	Y
Ethiopia (ETH)	2021		2021	2021	2021			2021	2021		6	60%	Y
Fiji (FJI)			2020	2020	2020			2021	2021		5	50%	Y
Finland (FIN)	2021	2021	2021	2021	2021	2021	2021	2021	2021		9	90%	Y
France (FRA)	2021	2021	2020	2020	2020	2020		2021	2021	2021	9	90%	Y
Gabon (GAB)			2021	2021	2021	2020		2021	2021		6	60%	Y
Gambia (GMB)								2021			1	10%	N
Georgia (GEO)	2021	2021	2021	2021	2021	2021	2021	2021	2021	2021	10	100%	Y
Germany (DEU)	2022	2022	2021	2021	2021	2021	2021	2021	2021		9	90%	Y

Economy (ISO code)	Proportion of individuals who used the Internet (from any location) in the last 3 months yHH7	Proportion of households with Internet access at home xHH6	Active mobile-broadband subscriptions per 100 inhabitants i911mw	Percentage of the population covered by at least a 3G mobile network i271G	Percentage of the population covered by at least an LTE/WiMAX mobile network. i271GA	Mobile broadband Internet traffic per mobile broadband subscriptions (GB) i136mwi_subs	Fixed broadband Internet traffic per fixed broadband subscriptions (GB) i135tfb_subs	Mobile data and voice high-consumption basket (as % of GNI per capita) i271mb_high_ts_GNI	Fixed-broadband Internet basket price (as % of GNI per capita) i154_FBB_ts_GNI	Percentage of individuals who own a mobile phone* xHH18	Indicators available for the reference period Number	 Share	>50%?
Ghana (GHA)	2021		2021	2021	2021	2021	2021	2021	2021	2021	9	90%	Y
Greece (GRC)	2021	2021	2021	2021	2021	2021	2021	2021	2021		9	90%	Y
Grenada (GRD)			2021	2021	2021	2020		2021	2021		6	60%	Y
Guatemala (GTM)	2021	2021	2020	2021	2020			2021	2021	2021	8	80%	Y
Guinea (GIN)								2021	2021		2	20%	N
Guinea-Bissau (GNB)			2021	2021	2021	2021	2021	2021	2021		7	70%	Y
Guyana (GUY)								2021	2021		2	20%	N
Haiti (HTI)								2021	2021		2	20%	N
Honduras (HND)			2021	2021	2021	2021	2021	2021	2021		7	70%	Y
Hong Kong, China (HKG)	2021	2021	2021	2021	2021	2021	2021	2021	2021	2021	10	100%	Y
Hungary (HUN)	2021	2021	2021	2021	2021	2021	2021	2021	2021		9	90%	Y
Iceland (ISL)	2021	2021	2021	2021	2021	2021	2021	2021	2021		9	90%	Y
India (IND)			2021	2021	2021	2021	2021	2021	2021		7	70%	Y
Indonesia (IDN)	2021	2021	2021	2021	2021	2021	2021	2021	2021	2021	10	100%	Y
Iran (Islamic Republic of) (IRN)	2021	2021	2021	2021	2021	2021	2021	2020	2020	2021	10	100%	Y
Iraq (IRQ)			2021	2021	2021	2021	2021	2021			6	60%	Y
Ireland (IRL)	2020	2020	2021	2021	2021	2021	2021	2021	2021		9	90%	Y
Israel (ISR)	2021	2021	2021	2021	2021			2021	2021		7	70%	Y
Italy (ITA)	2021	2021	2021	2021	2021	2021	2021	2021	2021		9	90%	Y
Jamaica (JAM)	2021	2021	2021	2021	2021	2021	2021	2021	2021		9	90%	Y
Japan (JPN)	2021	2021	2021	2021	2021	2021	2021	2021	2021	2021	10	100%	Y
Jordan (JOR)			2021	2020	2020	2021	2021	2021	2021		7	70%	Y
Kazakhstan (KAZ)	2021	2021	2021	2021	2021	2021	2021	2021	2021	2021	10	100%	Y
Kenya (KEN)			2021	2021	2021	2021	2021	2021	2021		7	70%	Y
Kiribati (KIR)			2021	2021	2021	2021	2021	2021			6	60%	Y
Korea (Rep. of) (KOR)	2021	2021	2021	2021	2021	2021	2021	2021	2021	2021	10	100%	Y
Kuwait (KWT)	2021	2021	2021	2021	2021	2021	2021	2021	2021	2021	10	100%	Y
Kyrgyzstan (KGZ)	2020	2020						2021	2021	2020	5	50%	Y
Lao P.D.R. (LAO)	2021	2021	2021	2021	2021	2021	2021	2021	2021		9	90%	Y
Latvia (LVA)	2022	2022	2021	2020	2021	2021	2021	2021	2021		9	90%	Y
Lebanon (LBN)			2020	2020	2020	2020	2020	2021	2021		7	70%	Y
Lesotho (LSO)			2021	2021	2021	2021	2021	2021	2021		7	70%	Y
Liberia (LBR)								2021			1	10%	N
Libya (LBY)								2021	2021		2	20%	N
Liechtenstein (LIE)			2021	2021	2021	2021	2021	2021	2021		7	70%	Y

Economy (ISO code)	Proportion of individuals who used the Internet (from any location) in the last 3 months yHH7	Proportion of households with Internet access at home xHH6	Active mobile-broadband subscriptions per 100 inhabitants i911mw	Percentage of the population covered by at least a 3G mobile network i271G	Percentage of the population covered by at least an LTE/WiMAX mobile network. i271GA	Mobile broadband Internet traffic per mobile broadband subscriptions (GB) i136mwi_subs	Fixed broadband Internet traffic per fixed broadband subscriptions (GB) i135tfb_subs	Mobile data and voice high-consumption basket (as % of GNI per capita) i271mb_high_ts_GNI	Fixed-broadband Internet basket price (as % of GNI per capita) i154_FBB_ts_GNI	Percentage of individuals who own a mobile phone* xHH18	Indicators available for the reference period Number	Share	>50%?
Lithuania (LTU)	2021	2021	2021	2021	2021	2021		2021	2021	2020	9	90%	Y
Luxembourg (LUX)	2021	2021	2021	2021	2021	2021		2021	2021		8	80%	Y
Macao, China (MAC)	2021	2021						2021	2021		4	40%	N
Madagascar (MDG)			2021	2021	2021	2021	2021	2021	2021		7	70%	Y
Malawi (MWI)			2021	2021	2021	2021	2021	2021	2021		7	70%	Y
Malaysia (MYS)	2021	2021	2021	2021	2021	2021	2021	2021	2021	2021	10	100%	Y
Maldives (MDV)			2021	2021	2021	2021	2021	2021	2021		7	70%	Y
Mali (MLI)			2021	2021	2021	2021	2021	2021	2021		7	70%	Y
Malta (MLT)	2021	2021	2021	2021	2021	2021		2021	2021		8	80%	Y
Marshall Islands (MHL)									2021		1	10%	N
Mauritania (MRT)			2021	2021		2021		2021	2021		5	50%	Y
Mauritius (MUS)	2020	2020	2021	2021	2021	2021	2021	2021	2021	2020	10	100%	Y
Mexico (MEX)	2021	2021	2021	2021	2021	2021		2021	2021		8	80%	Y
Micronesia (FSM)								2021	2021		2	20%	N
Moldova (MDA)		2021	2021	2021	2021	2021		2021	2021		7	70%	Y
Monaco (MCO)			2021	2021	2021	2021					4	40%	N
Mongolia (MNG)	2021	2021	2021	2021	2021	2021	2021	2021	2021	2021	10	100%	Y
Montenegro (MNE)	2021	2021	2021	2021	2021	2021	2021	2021	2021		9	90%	Y
Morocco (MAR)	2021	2021	2021	2021	2021	2021	2021	2021	2021	2021	10	100%	Y
Mozambique (MOZ)			2021	2020	2020	2020		2021	2021		6	60%	Y
Myanmar (MMR)			2021	2021	2021	2021	2021	2021	2021		7	70%	Y
Namibia (NAM)			2021	2021	2021	2021		2021	2021		6	60%	Y
Nauru (NRU)								2021			1	10%	N
Nepal (Republic of) (NPL)								2021	2021		2	20%	N
Netherlands (NLD)	2021	2021	2021	2021	2021	2021		2021	2021	2021	9	90%	Y
New Zealand (NZL)			2021	2021	2021	2021	2021	2021	2021		7	70%	Y
Nicaragua (NIC)			2021	2021	2021			2021	2021		5	50%	Y
Niger (NER)								2021	2020		2	20%	N
Nigeria (NGA)			2021	2021	2021	2021	2021	2021	2021		7	70%	Y
North Macedonia (MKD)	2020	2020	2021	2021	2021	2021	2021	2021	2021		9	90%	Y
Norway (NOR)	2021	2021	2020	2020	2020	2020		2021	2021	2021	9	90%	Y
Oman (OMN)	2020	2020	2021	2021	2021	2021	2021	2021	2021	2020	10	100%	Y
Pakistan (PAK)	2020	2020	2021	2021	2021	2021	2021	2021	2021	2020	10	100%	Y
Palestine (WBG)	2022	2022	2021	2021	2021	2021	2021	2021	2021		9	90%	Y
Panama (PAN)			2021	2021	2021			2021	2021		5	50%	Y

Economy (ISO code)	Proportion of individuals who used the Internet (from any location) in the last 3 months yHH7	Proportion of households with Internet access at home xHH6	Active mobile-broadband subscriptions per 100 inhabitants i911mw	Percentage of the population covered by at least a 3G mobile network i271G	Percentage of the population covered by at least an LTE/WiMAX mobile network. i271GA	Mobile broadband Internet traffic per mobile broadband subscriptions (GB) i136mwi_subs	Fixed broadband Internet traffic per fixed broadband subscriptions (GB) i135tfb_subs	Mobile data and voice high-consumption basket (as % of GNI per capita) i271mb_high_ts_GNI	Fixed-broadband Internet basket price (as % of GNI per capita) i154_FBB_ts_GNI	Percentage of individuals who own a mobile phone* xHH18	Indicators available for the reference period Number	Indicators available for the reference period Share	>50%?
Papua New Guinea (PNG)								2021	2021		2	20%	N
Paraguay (PRY)	2021	2021	2021	2021	2021			2021	2021		7	70%	Y
Peru (PER)	2021	2021	2021	2021	2021			2021	2021	2021	8	80%	Y
Philippines (PHL)			2020	2020	2020	2021		2021	2021		6	60%	Y
Poland (POL)	2021	2021	2021	2021	2021	2021		2021	2021		8	80%	Y
Portugal (PRT)	2021	2021	2021	2020	2021	2021	2021	2021	2021		9	90%	Y
Qatar (QAT)	2020	2020	2021	2021	2021	2021	2021	2021	2021	2020	10	100%	Y
Romania (ROU)	2021	2021	2021	2021	2021	2021	2021	2021	2021	2021	10	100%	Y
Russian Federation (RUS)	2021	2021	2021	2021	2021	2021	2021	2021	2021	2021	10	100%	Y
Rwanda (RWA)	2020		2021	2021	2021	2021	2021	2021	2021	2020	9	90%	Y
Saint Kitts and Nevis (KNA)			2021	2021	2021			2021	2021		5	50%	Y
Saint Lucia (LCA)			2021	2021	2021	2020		2021	2021		6	60%	Y
Saint Vincent and the Grenadines (VCT)			2021	2021	2021	2020		2021	2021		6	60%	Y
Samoa (WSM)								2021	2021		2	20%	N
San Marino (SMR)			2021	2021	2021						3	30%	N
Sao Tome and Principe (STP)			2021	2021		2021	2021	2021	2021		6	60%	Y
Saudi Arabia (SAU)	2021	2021	2021	2021	2021	2021	2021	2021	2021	2021	10	100%	Y
Senegal (SEN)			2021	2021	2021			2021	2021		5	50%	Y
Serbia (SRB)	2021	2021	2021	2021	2021	2021	2021	2021	2021	2021	10	100%	Y
Seychelles (SYC)			2021	2021	2021	2021	2021	2021	2021		7	70%	Y
Sierra Leone (SLE)			2021	2021	2021			2021			4	40%	N
Singapore (SGP)	2022	2022	2021	2021	2021	2021		2021	2021	2021	9	90%	Y
Slovakia (SVK)	2021	2021	2021	2021	2021	2021	2021	2021	2021	2021	10	100%	Y
Slovenia (SVN)	2021	2021	2021	2021	2021	2021		2021	2021	2021	9	90%	Y
Solomon Islands (SLB)								2021	2021		2	20%	N
Somalia (SOM)		2020	2021	2021	2021			2021	2021		6	60%	Y
South Africa (ZAF)		2021	2021	2021	2021	2021	2021	2021	2021		8	80%	Y
South Sudan (SSD)			2021	2021	2021		2021				4	40%	N
Spain (ESP)	2021	2021	2021	2021	2021	2021	2021	2021	2021	2021	10	100%	Y
Sri Lanka (LKA)		2020	2021	2021	2021	2021	2021	2021	2021		8	80%	Y
Sudan (SDN)			2021	2021	2021			2021			4	40%	N
Suriname (SUR)			2021	2021	2021	2021	2020	2021	2021		7	70%	Y
Sweden (SWE)	2022	2021	2021	2021	2021	2021		2021	2021		8	80%	Y
Switzerland (CHE)	2021	2021	2021	2021	2021	2021	2021	2021	2021	2021	10	100%	Y
Syrian Arab Republic (SYR)			2021	2021	2021	2021	2020				5	50%	Y

Economy (ISO code)	Proportion of individuals who used the Internet (from any location) in the last 3 months yHH7	Proportion of households with Internet access at home xHH6	Active mobile-broadband subscriptions per 100 inhabitants i911mw	Percentage of the population covered by at least a 3G mobile network i271G	Percentage of the population covered by at least an LTE/WiMAX mobile network. i271GA	Mobile broadband Internet traffic per mobile broadband subscriptions (GB) i136mwi_subs	Fixed broadband Internet traffic per fixed broadband subscriptions (GB) i135tfb_subs	Mobile data and voice high-consumption basket (as % of GNI per capita) i271mb_high_ts_GNI	Fixed-broadband Internet basket price (as % of GNI per capita) i154_FBB_ts_GNI	Percentage of individuals who own a mobile phone* xHH18	Indicators available for the reference period Number	 Share	>50%?
Tajikistan (TJK)								2021	2021		2	20%	N
Tanzania (TZA)			2021	2021	2021	2021	2021	2021	2021		7	70%	Y
Thailand (THA)	2021	2021	2021	2021	2021	2021	2021	2021	2021	2021	10	100%	Y
Timor-Leste (TLS)			2021	2021	2021			2021	2021		5	50%	Y
Togo (TGO)			2021	2021	2021	2021	2021	2021	2021		7	70%	Y
Tonga (TON)			2021	2021	2021	2021	2021	2021	2021		7	70%	Y
Trinidad and Tobago (TTO)	2021	2021	2021	2021	2021	2021	2021	2021	2021	2021	10	100%	Y
Tunisia (TUN)			2021	2021	2021	2021	2021	2021	2021		7	70%	Y
Türkiye (TUR)	2022	2022	2021	2021	2021	2021	2021	2021	2021	2022	10	100%	Y
Turkmenistan (TKM)								2021	2021		2	20%	N
Tuvalu (TUV)								2021	2021		2	20%	N
Uganda (UGA)	2020		2021	2021	2021	2021	2021	2021			7	70%	Y
Ukraine (UKR)	2021	2021	2021	2021	2021			2021	2021	2021	8	80%	Y
United Arab Emirates (ARE)	2021	2021	2021	2021	2021	2021	2021	2021	2021	2021	10	100%	Y
United Kingdom (GBR)	2020	2020	2021	2021	2021		2021	2021	2021		8	80%	Y
United States (USA)			2021	2021	2021			2021	2021		5	50%	Y
Uruguay (URY)			2021	2021	2021	2021	2021	2021	2021		7	70%	Y
Uzbekistan (UZB)	2021	2021	2021	2021	2021	2021	2021	2021	2021	2021	10	100%	Y
Vanuatu (VUT)			2021	2021	2021	2021	2021	2021	2021		7	70%	Y
Vatican (VAT)											0	0%	N
Venezuela (VEN)			2021	2021	2021	2021	2021				5	50%	Y
Viet Nam (VNM)	2021	2021	2021	2021	2021	2021	2021	2021	2021	2021	10	100%	Y
Yemen (YEM)								2020	2020		2	20%	N
Zambia (ZMB)			2021	2021	2021		2021	2021	2021		6	60%	Y
Zimbabwe (ZWE)	2020	2020	2021	2021	2021	2021	2021	2021	2021	2020	10	100%	Y
Nr. Economies with data available for the reference period (2020-2021)	94	94	170	170	168	143	115	185	175	59			165

Annex 3: Correlation analysis

Correlation analysis is an essential statistical tool for composite indicator development. By helping to understand the statistical relationships among the indicators considered for inclusion, it provides an early indication of the strength of an index and of possible internal consistency problems.

Correlation coefficients indicate overlaps, complementarities, and trade-offs across indicators, which are often not evident when indicators are selected purely for their conceptual relevance. For instance, the stronger the correlation between two indicators, the higher the statistical overlap between them. Near collinearity (i.e., a coefficient close to 1) signals that the two indicators contain the same information with regards to establishing country scores. Conversely, if there is no statistical association between two indicators (correlation coefficients close to 0), the two indicators fully complement one another, each providing very different information about the country performance. Negative correlation would indicate unintended trade-offs (i.e., improving one dimension comes at the detriment of another).

While there is no optimal degree of correlation in the context of an index, it is important to ensure that the selected indicators fit in the aggregation framework based on positive correlation with the other indicators in the same index component (e.g., a pillar) and the overall index. A composite indicator that is the average of uncorrelated component indicators is confusing, because how countries perform according to the index will look very different from how countries perform according to the individual indicators. Yet, component indicators should not be perfectly aligned, as this would not only weaken the case for having multiple indicators instead of using just one, but also imply double counting of the same information. Therefore, components should be positively correlated, but not statistically identical (coefficients close to 1), so that the aggregate index is a summary measure, with the added value that it helps reduce dimensionality in a larger underlying dataset.

Correlation analysis can also inform weighting (e.g., to avoid double counting in case of near collinearity), as well as the structuring of indicators (e.g., if multiple dimensions or pillars are used, ensuring that each indicator is assigned to the dimension with which it shares the highest statistical commonality to ensure coherence of the framework.

Table 8: Correlation table for the selected indicators

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
i99H (1)	1	0.90	0.63	0.59	0.68	0.34	0.34	-0.73	-0.67	0.90
xHH6 (2)	0.90	1	0.69	0.61	0.69	0.31	0.37	-0.67	-0.61	0.81
i911mw (3)	0.63	0.69	1	0.41	0.56	0.28	0.23	-0.52	-0.45	0.64
i271G (4)	0.59	0.61	0.41	1	0.80	0.21	0.26	-0.62	-0.50	0.59
i271GA (5)	0.68	0.69	0.56	0.80	1	0.26	0.26	-0.62	-0.55	0.64
i136mwi_subs (6)	0.34	0.31	0.28	0.21	0.26	1	0.22	-0.22	-0.25	0.30
i135tfb_subs (7)	0.34	0.37	0.23	0.26	0.26	0.22	1	-0.21	-0.11	0.26
i271mb_high_ts_GNI (8)	-0.73	-0.67	-0.52	-0.62	-0.62	-0.22	-0.21	1	0.65	-0.73
i154_FBB_ts_GNI (9)	-0.67	-0.61	-0.45	-0.50	-0.55	-0.25	-0.11	0.65	1	-0.66
xHH18_IDI (10)	0.90	0.81	0.64	0.59	0.64	0.30	0.26	-0.73	-0.66	1

Notes: Pairwise Pearson correlation coefficients shaded by strength and significance.
Indicators (1) to (3) refer to universal connectivity; (4) to (10) refer to meaningful connectivity, among which (4) to (7) refer to infrastructure, (8) to (9) measure affordability, and (10) measures device ownership. See Table 6 for indicator names.

Table 8: shows the correlation coefficients for the selected indicators. This analysis was carried out before any treatment, so some of the patterns are driven by the outliers (see identification in Step 4), and the test should be repeated on the treated dataset. The tests revealed the following information about indicator groups and indicator pairs:

- Overall, the correlation coefficients show the expected signs in the selected indicators set. The negative correlation of the two affordability indicators with the other indicators is also expected, since those indicators are measured in an opposite direction: the lower the prices, the better the situation (this means that the direction should be reversed when normalizing these indicators).
- The indicators in the **universal connectivity group** are positively and moderately to strongly correlated with one another. The two survey-based indicators (share of individuals using the Internet and households accessing the Internet) share the highest degree of similarities, while the somewhat weaker coefficients between the fixed and mobile broadband penetration indicators show that the two technologies are complementary to one another. Similarly, the moderate correlation between the two survey-based measures and the penetration measures based on administrative data shows complementarities between the two approaches. It is possible though that the difference can be explained, to some extent, by the pattern of missing data. Combining indicators of the universal connectivity group into a dimension aggregate appears to make sense from a statistical perspective, as it would not result in a significant loss of information.
- Correlation across indicators in the **meaningful connectivity group** shows greater heterogeneity. Not only does the group stand somewhat apart from the universal connectivity indicators group, but there is also considerable heterogeneity across its different subsets.
- In the **meaningful connectivity – infrastructure group**:
    - The strong positive correlation between the pair of indicators for mobile broadband coverage by at least 3G and 4G technologies suggests that the two indicators can be combined into a single indicator.
    - The two Internet traffic indicators – at least prior to outlier treatment stand apart from the other indicators of the infrastructure group and are also complementary to one another.
    - All this indicates that aggregating these indicators to a single component would involve some degree of compensability among the indicators: countries scoring high on the traffic indicators do not necessarily score high on other indicators in the group. When aggregated, this implies that weaker performance in traffic may be compensated by stronger performance in other indicators, and *vice versa*.
    - The correlation analysis should be revisited after outlier treatment and possible sub-aggregation of the broadband coverage indicators to better understand statistical coherence.
- The **affordability indicators** for the two technologies (mobile data and voice high-consumption basket and fixed broadband basket price, both as a percentage of GNI per capita) are complementary to one another. Interestingly, considering the correlation pattern with the other indicators across the table, while one may expect that all indicators relating to the same technology but measuring different aspects of it (e.g., penetration, traffic, affordability) show greater statistical similarities with one another, correlation patterns show no evidence of that.