# Executive Summary

|  |
| --- |
| The current paper analyzed a total of 138 ESG indicators in order to recommend actions for the improvement of the availability, organization and quality of ESG indicators. A total of 36 indicators have been found to be discontinued, mainly due to not being updated by the producers. There is no significant variation between collection, processing and dissemination of indicators and between production frequency and update frequency at the WBG. Country coverage and data point availability decrease generally as we get the most recent year. Most recent values are thinner for indicators produced by other organizations, followed by indicators sourced from UN system and WBG.  62 indicators out of 76 have a CV higher than 100% which translates into a high and very high variance of the indicators. A total of 34 indicators out of 76 have a QCD higher than 50%. 21 of these are social indicators and 13 of them are environmental indicators.  In the short run, increasing the production timeline for WBG sourced indicators appears to yield the most improvements of ESG indicators since a significant number of WBG sourced indicators have no data for 2018.  Data sourced from other organizations and from the UN system have a significant number of indicators with no data for 2018 and could benefit from an increased pace of production.  Increasing the timeline or production, update and dissemination and coverage of indicators identified as having high CV and QCD can improve significantly the quality of the ESG datasets.  Finally, obtaining better publishing rights from organizations such as IEA might increase the availability of ESG data. An ESG data portal should focus in the first instance on indicators that have wide country coverage and significant data coverage for most recent years. |

# Introduction

The current paper is the continuation of the work conducted and resumed in a discussion paper titled “Options for Improving Use of ESG data for Sovereign Bond Analysis”. Following interaction with investors in sovereign bonds that used Environmental, Social and Governance indicators in their country analyses and risk/return profiles of sovereign securities, the World Bank set out to present a set of options for improving accessibility, quality (e.g. timeliness and regularity of publication, geographic coverage) and transparency of World Bank Group (WBG) Emerging Markets data. More specifically of interest were Environmental, Social and Governance data. This paper aims to update the analysis and provide a series of recommendations for improving the accessibility, quality and transparency of ESG indicators.

# 2018 Paper Summary

The 2018 paper assessed a series of three options for improving the accessibility and quality of ESG data and selected a total number of 138 indicators for subsequent analysis in this regard. The previous paper found that financial institutions use ESG data either coming from data providers which gather information and sell it to interested parties (e.g. Sustainalytics, Bloomberg, Verisk Maplecroft, MSCI) and rating agencies, index providers and asset management firms that include ESG data into their proprietary analytical frameworks or offer ESG investment products (e.g. RobecoSAM, JPMorgan, Moody’s, Thomson Reuters, FTSE Russell, Arabesque and Neuberger Berman). A quick analysis of ESG tools provided by Bloomberg, MSCI, Robescam and Maplecroft revealed that at least 137 indicators used in these tools are sourced from various WBG databases, such as World Development Indicators (WDI), Doing Business, World Governance Indicators (WGI), Women Business & Law and Global Financial Development among others. Bloomberg used 46 WBG indicators out of a total of 54 ESG indicators. MSCI used 76 WBG indicators out of the total of 131 ESG indicators while Robescam used a total of 8 WBG indicators of a total of 24 ESG indicators. In 2019, the same number of indicators stemming from WBG databases was used by Bloomberg, MSCI and Robescam.

**Table 2: Data Sources for Select ESG Tools - 2018**

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **INDICATORS USED** | | | | | | | |
|  | **BLOOMBERG** | | **MSCI** | | **ROBECOSAM** | | **MAPLECROFT** | |
| **SECTOR** | **WBG** | **TOTAL** | **WBG** | **TOTAL** | **WBG** | **TOTAL** | **WBG** | **TOTAL** |
| Environment | 11 | 13 | 17 | 39 | 0 | 5 | 3 | n/a |
| Social | 21 | 22 | 46 | 60 | 1 | 6 | 15 | n/a |
| Governance | 14 | 19 | 13 | 32 | 7 | 13 | 25 | n/a |
| **TOTAL** | **46** | **54** | **76** | **131** | **8** | **24** | **43** |  |

The 2018 paper revealed that a lack of dedicated framework on measuring and understanding ESG approaches and how indicators are used for ESG purposes is a significant barrier to the development of their use by the broader finance community. Additionally, it was revealed that data coverage represented a significant issue in WBG data used for ESG because there are a significant number of cases where data was either out of date by several years or where the most recent value (MRV) varied substantially across countries. The previous analysis showed that the number of ESG indicators for which recent values are available (2015, 2016, 2017 or later) decreases. 45% of indicators have recent values (2016 or later) for at least 75% of countries in the WBG system. One extreme example is India for where the most recent available poverty and inequality estimate is more than six years old, based on a survey from 2011-2012. Often, older values are carried forward to more recent years by the ESG community in their efforts to bridge the information gaps, but problems occur when new data comes in and differs from carried forward data, an occurrence that is hard to explain, more so in dynamic economies where this is frequently the case.

Having identified these issues in the previous paper, a series of three options have been proposed to improve the use of WBG indicators for ESG activities.

1. the first option was to improve the availability and organization of ESG data currently in use mainly by compiling a dataset of indicators used for ESG, ideally in an online ESG portal following the similar examples of for health, gender, jobs, SDG and other.
2. option two proposed the exploration of means to fill gaps in existing ESG indicators
3. option three explored leveraging WBG expertise to provide direction for ESG data and standards.

The current paper explores the first two options proposed in the 2018 paper. It aims at providing an overview of the indicators curated and disseminated by World Bank and relevant for ESG (Environmental, Social, Governance) investment. The work has been conducted assembling a metadata database on indicators, doing a series of interviews with data managers and analyzing and researching the set of indicators and their metadata. We selected138 indicators of relevance to ESG investments, selection informed by the research and consultations conducted for the 2018 paper and we provide in the following parts of the paper an analytical overview of these indicators. We start first with a detailed overview of the methodology used to analyze the dataset and we continue with providing a set of insights on data availability, sources, processing entities and dissemination of these indicators and the frequency of their updates and release. The first part of the paper aims at providing an overview of the methodology used for the analysis and a description of the database assembled for the analysis. The second part of the paper aims at providing analytical insights on the ESG indicators with a focus on measuring data quality and availability, to improve availability, fill in gaps and organize ESG data under a common framework.

# Methodology

We assembled a database for analysis consisting of metadata with a total of 34 variables. These variables were meant to provide the most comprehensive overview of each indicator separately. In addition to an individual tracking code assigned to each indicator, we categorized them by characteristics such as the database of which the indicators is part of, its topic, its statistical methodology used for compilation, its aggregation methods, its relevance to development, its periodicity, among others. Out of these 34 variables, we identified a set of 17 variables that contain most of the information that is relevant for the analysis of these indicators in the context of their use in ESG investments. This subset of variables has been set out to capture the collecting, processing and dissemination entity and the granularity of data (monthly, yearly etc) Finally, we tried to capture the statistical methods through which the indicators are created and to capture imputations performed on the dataset. The annex of this paper provides an overview of the initial database of the indicators metadata and the short version with the variables of most interest to this paper. In order to fill in knowledge gaps in metadata, we conducted eight interviews with internal Point of Consultation (PoC), which were selected due to their extensive knowledge of each indicator of interest. The interviews took place over a three week period, in person and over phone and focused mostly on timeliness of updates.

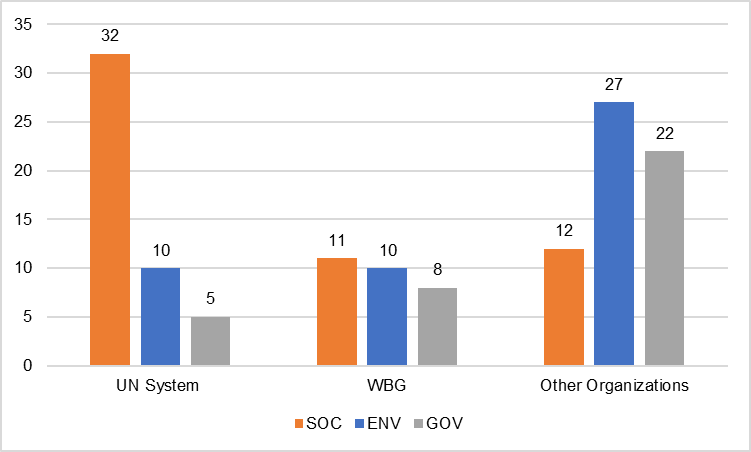
Environmental indicators are related to climate change, carbon emissions, pollution, resource efficiency and biodiversity. Social indicators pertain to human rights, labor standards, health and safety, diversity policies, development of human capital (health & education). Finally, governance indicators refer to corporate governance issues, corruption, rule of law, institutional strength and transparency.

Finally, using the database assembled on the abovementioned guidelines we were able to analyze the metadata of indicators and provide insights on key issues such as update frequency and dissemination, processing methods, data handling techniques, and quality of indicators. Additionally, an analysis of data variability and of data points availability was conducted in order to assess first the impact of changes in indicator values and how old data was.

# Database description

The final database assembled for analysis had a total of 35 variables for a total of 138 indicators. 47 indicators are environmental indicators, 25 governance indicators and 66 social indicators. The World Bank is a source for a total of 29 indicators, of which 11 are social indicators, 10 are environmental indicators and 8 are governance indicators. Organizations pertaining to the UN system are a source for 49 indicators of which 32 social indicators, 10 environmental indicators and 5 governance indicators. 51 indicators are sourced from different organization grouped under the term of other organization. 27 of these indicators are environmental, 12 are social and 22 are governance indicators.

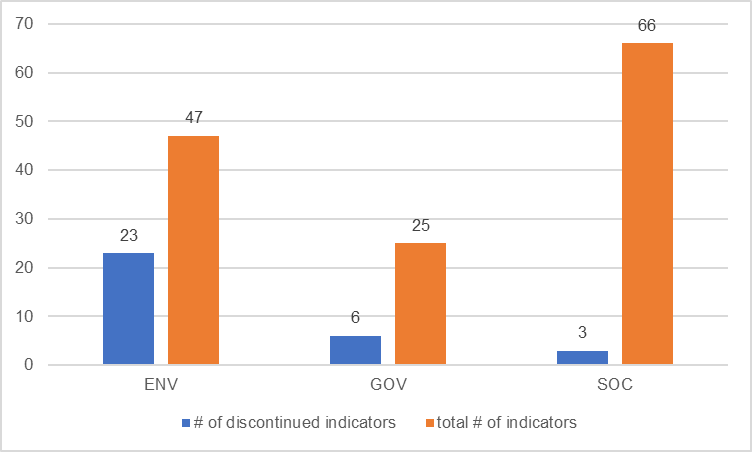
Figure 1: Indicators disaggregated by type of source and within sectors (ENV, SOC, GOV)



# Discontinued Indicators

Following the interviews with the data managers, we identified a total of 32 out of 138 indicators that are discontinued. 23 of these indicators are environmental indicators, 6 are governance indicators and 3 are social indicators. Given the number of discontinued indicators, there are a total of 106 ESG indicators that are continuously updated. A total of 24 environmental indicators from an initial batch of 47 indicators are updated on a continuous basis. This is by far the highest rate of attrition observed in the indicators as 51% of the initial set of indicators appears to be no longer updated. 19 governance indicators are continuously updated from an initial set of 25, representing a 24% attrition rate. Finally, social indicators appear to be the most robust class of indicators with 63 out of 66 original ones being continuously updated equivalent of a 5% attrition rate.

Figure 2: Distribution of discontinued indicators



The preliminary findings related to discontinued indicators pertain to some of these indicators not being updated by the source organization, such as indicators sourced from the Carbon Dioxide Information Analysis Center. Other indicators have been used once only for a publication or have been replace by other indicators such as in the case of air quality, PM10 to PM2.5.

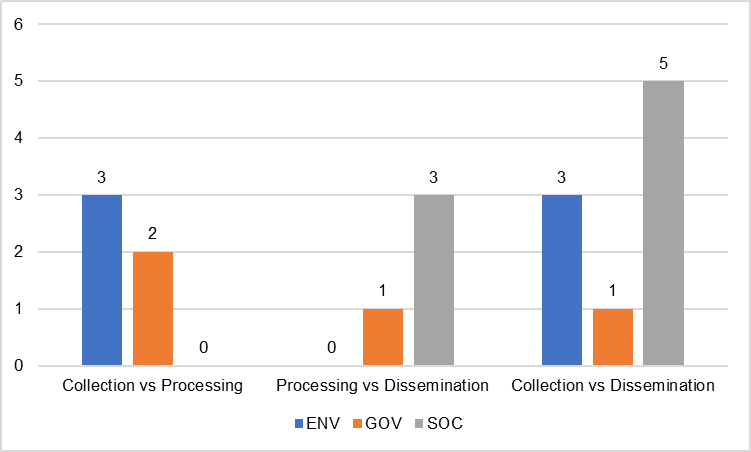
# Data Collection, Processing and Dissemination

We designed three separate source types to capture variation between entities that collect the data, entities that process and entities that disseminate the datasets. We defined a collecting entity an entity that gathers the raw data through different statistical instruments and methods. A processing entity is defined as an entity that takes raw data and transforms it into indicators and numbers that are meaningful and can be released to a wider audience. Finally, a dissemination entity is defined as an entity that takes data ready for dissemination with wider audiences and disseminates it through their own channels such as online databases, brochures, booklets and other means.

5 out of 103 ESG indicators have a different processing entity as compared to the collecting entity. 3 of these indicators are environmental indicators and 2 are governance indicators. For example, energy intensity level is calculated per unit of GDP at Power Purchase Parity (PPP) by taking data from IEA on energy indicators and data on GDP from World Bank. There are 4 indicators only that have a different dissemination entity as compared with the processing entity. 3 of these indicators are social indicators and 1 is a governance indicator. All these indicators are collected and processed by OECD, but disseminated by UNESCO. Finally, a total of 9 indicators have a different dissemination entity as compared to the collection entity. 5 of these are social indicators, 3 are environmental indicators and one is a governance indicators. 5 of these indicators are disseminated by World Bank and 3 by UNESCO, while three of them are collected by IEA, another 3 by OECD and one each by FAO and UN.

In conclusion, there is little variation between collecting, processing and disseminating entities as organizations usually disseminate through their own channels the datasets that they collect and process. Usually, differences appear when initial data is divided by population to express a certain indicator at the per capita level.

Figure 3: Differences between collection, processing and dissemination

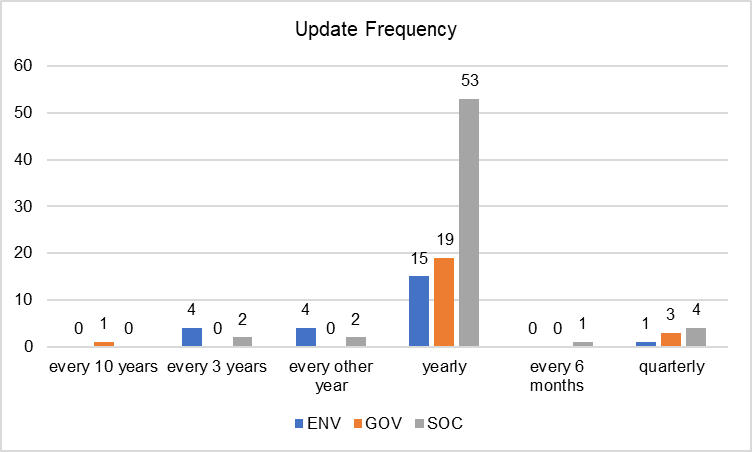
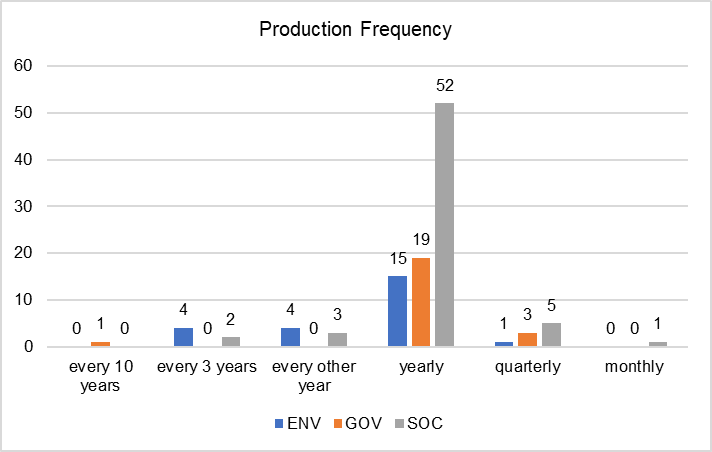


# Highest Frequency Production and Updates

There is insignificant variation between the production update by the organizations producing the indicators and the timeline of update at the World Bank. Two indicators only, both social, come with different update timeline at the World Bank as compared to the production update. One of the indicators, Agriculture, forestry, and fishing, value added (% of GDP), is updated quarterly by the collecting entity while the World Bank updates it every six months and another one, Proportion of seats held by women in national parliaments (%), is updated monthly by the collecting entity while the World Bank updates it annually.

12 out of 20 environmental indicators are produced yearly by the originator and updated yearly by the World Bank. Four environmental indicators are produced and updated every other year while three are produced and updated every three years. 50 out of 61 social indicators are produced and updated on a yearly basis while 13 out of 16 governance indicators are produced yearly. In total, 75 out of 97 indicators are produced and updated, by the World Bank, on a yearly basis making this both the highest production and update frequency.

Figure 4: Production frequency vs. update frequency



# Country Coverage Analysis

## Environmental indicators

Country coverage of indicators point to the comprehensiveness and availability of data of the analyzed indicators. As of 2019, there are only two environmental indicators that do not have a single country data point in 2010, while there are 6 indicators that have no data point in 2014. As we come to more recent years, lack of data at country level, is more prevalent. There are 22 indicators with not a single data point at country level for 2016 while 3 environmental indicators have data for a single country in 2016. Another 10 indicators have data points for around 196 countries each, on average. Country coverage decreases dramatically in 2018. 28 indicators have zero country coverage for 2018 while three only have one country data point. A number of four indicators only have on average 211 country datapoints for 2018. The same analysis was performed in 2018 and the year before there were a total of 5 environmental indicators with no data data since 2013 increasing to 13 indicators with no data from 2015 and finally to 28 indicators with no data for 2017.

## Social Indicators

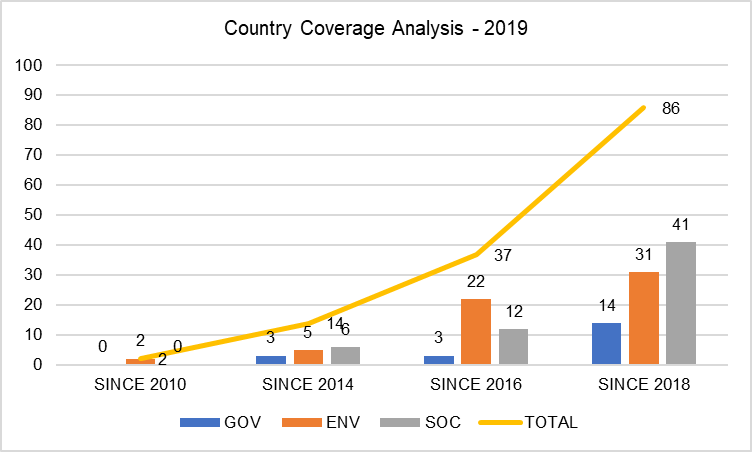
We observed a total of 5 social indicators with no data since 2014 increasing to 22 indicators with no data since 2016 and finally to 31 indicators with no data since 2018. In the previous iteration of the analysis there were 8 indicators with no data for both since 2013 and since 2015 and finally increasing to 36 since 2017.

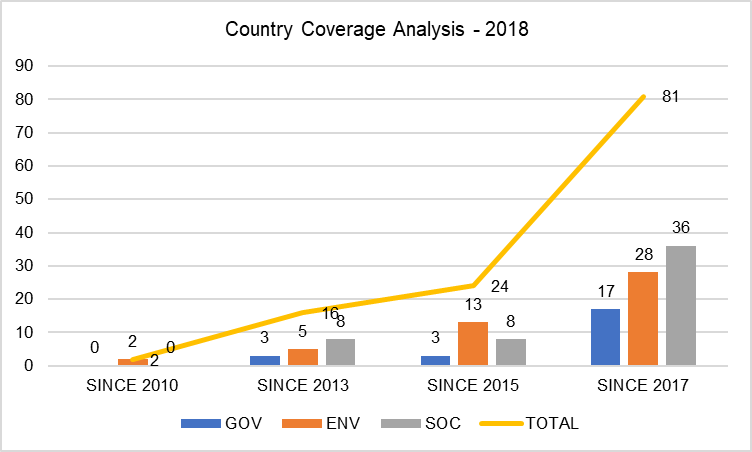
## Governance Indicators

We observed a total of 3 indicators with no data since both 2014 and 2016. This number has increased to 14 indicators with no data since 2018. In the analysis performed for the previous paper, we observed the same number of 3 indicators with no data since both 2013 and 2015 respectively and 17 indicators with no data since 2017.

The first insight that results from this analysis is that the number of indicators with no country coverage since the year before the analysis has increased from 81 to 86 in the current analysis. The number of environmental indicators with no country coverage for the previous year has increased from 27 in the first analysis to 31 indicators currently. The highest increase is observed for governance indicators, from 36 in the first analysis to 41 in the current analysis. Finally, the number of governance indicators without country coverage for the year before the analysis has declined from 17 to 14.

Figure 5: Country Coverage of Indicators

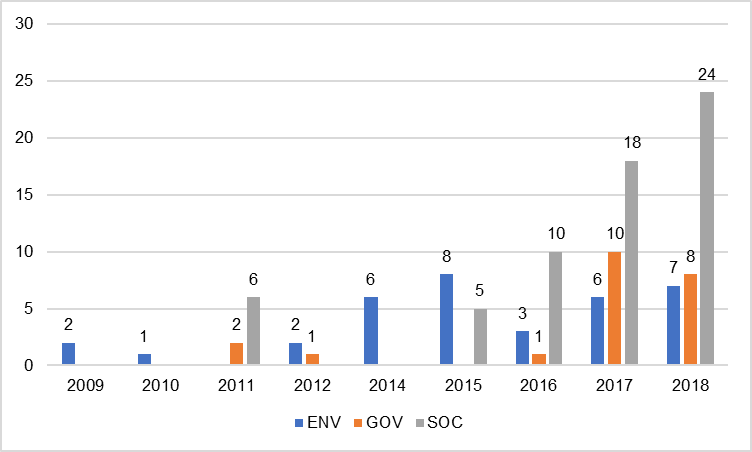




## Most Recent Values

Figure 6 gives an overview of the availability of data of the ESG indicators. Each value represents the number of environment, governance, social indicators that have the most recent value the respective year. The graph helps in identifying the number of indicators with missing data for recent years. The graph progresses linearly as the older the year the less the number of indicators that have a most recent value in that year given that they are updated regularly. Thus, currently there are 2 environmental indicators that have their most recent value in 2009. One indicator is sourced from a research institute and refers to droughts and the other one from UNFCCC and pertains to GHG net emissions. On the other end, the number of indicators that have 2018 as most recent value increases significantly. There are 24 indicators social indicators out of a total of 63 social indicators that have 2018 data. This translates into a rate of coverage with 2018 data for social indicators of 40%. Furthermore, this means that 60%, or 39 of social indicators do not have 2018 data.

Figure 6: Most recent values of indicators



There are only 8 governance indicators of a total of 19 governance indicators that have data for 2018. Thus, more than 58% of indicators, 11, do not have 2018 data.

Finally, only 7 environmental indicators have data for 2018, representing 29% of all environmental indicators.

Globally, only 39 indicators of a total of 106 have data for 2018, meaning that 63% of indicators in this analysis do not have data for 2018.

Figure 7: Number of Indicators – 2018 Most Recent Value

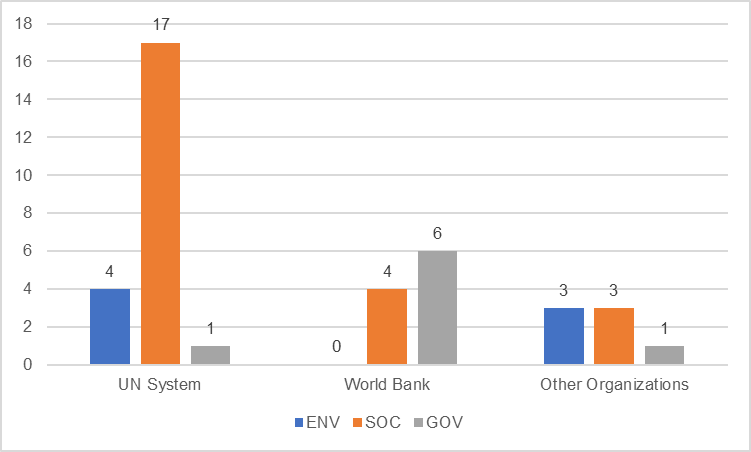


Figure 7 presents the source and number of indicators that have 2018 data. A total of 17 social indicators out of 32 sourced from the United Nations system have values for 2018, a rate of 53%. Furthermore, 4 out of 10 environment indicators sourced from the UN system have data for 2018 while only 1 social indicators out of 5 UN system sourced indicators has 2018 data.

Data sourced from within the World Bank Group indicated that 6 governance indicators out of a total of 8 indicators have 2018 data. Furthermore, 4 out of 11 social indicators have 2018 data and none of 10 environmental indicators have 2018 data. In terms of 2018 data availability, this is the largest gap in data observed so far.

Finally, the most significant occurrence is observed in data sourced from other organizations. The availability of data for 2018 is very low. Only 3 out of 27 environmental indicators have 2018 data. Only 3 social indicators out of 12 have 2018 data and 1 only our of 22 governance indicators have data for 2018.

Thus, UN and WBG sourced indicators have one of the largest availability of 2018 data, while data sourced from other organizations contains a significant less number of 2018 data points.

## Variability Analysis

## Coefficient of Variation and Quartile Coefficient of Dispersion – Indicator Level

We have analyzed two measures of variability. Firstly, we analyzed the coefficient of variation (CV), also known as relative standard variation which in its essence is a measure of dispersion, expressed as a percentage and calculated as the ratio of the standard deviation to the mean. A result of zero means that the dataset has constant values only. As the percentage result increases, the variation is higher.

Figure 8 presents an overview of the variation coefficients for the indicators grouped under Social and Environmental sectors. The range of variation for environmental indicators is higher than the range of variation for social indicators. The lowest CV starts at 28% for the environmental indicators (renewable internal freshwater resources) and at 89% for social indicators (internet users). The highest CV ends at 213% for environmental indicators (P.M 2.5 air pollution) and 173% for social indicators (youth unemployment, total). The 20th percentile starts at 134% CV for social indicators and ends at 166%, 80th percentile and for environmental indicators it starts at 85% and ends at 161%. The CV is more clustered around the mean for social indicators, at 147% than it is for environmental indicators with a mean of 127%. In total, 62 indicators out of 76 have a CV higher than 100% which translates into a high and very high variance of the indicators.

A number of 6 environmental indicators have CV above 160%. These are the following ones: PM 2.5 air pollution, fossil fuel energy consumption, energy imports, GHG net emissions, energy intensity level of primary energy, food production index and CO2 emissions. For social indicators, there are a total of 12 indicators for which CV surpasses 160%. These are the following: Unemployment, youth total (% of total labor force ages 15-24) (modeled ILO estimate); Net migration; Unemployment, total (% of total labor force) (modeled ILO estimate); Unemployment, female (% of female labor force) (modeled ILO estimate); School enrollment, primary (gross), gender parity index (GPI); Government expenditure on education, total (% of government expenditure); Tuberculosis prevalence rate, low uncertainty bound (per 1000,000 population, WHO); School enrollment, secondary (% net); Vulnerable employment, total (% of total employment) (modeled ILO estimate); Pupil-teacher ratio, secondary; GINI index (World Bank estimate); Employment to population ratio, ages 15-24, total (%) (modeled ILO estimate). Given the high CV of these 18 indicators, improvements in data collection and timelines of production and update of these indicators will yield significant improvement in the ESG dataset of indicators.

Figure 8: Coefficient of Variance – Indicator Level

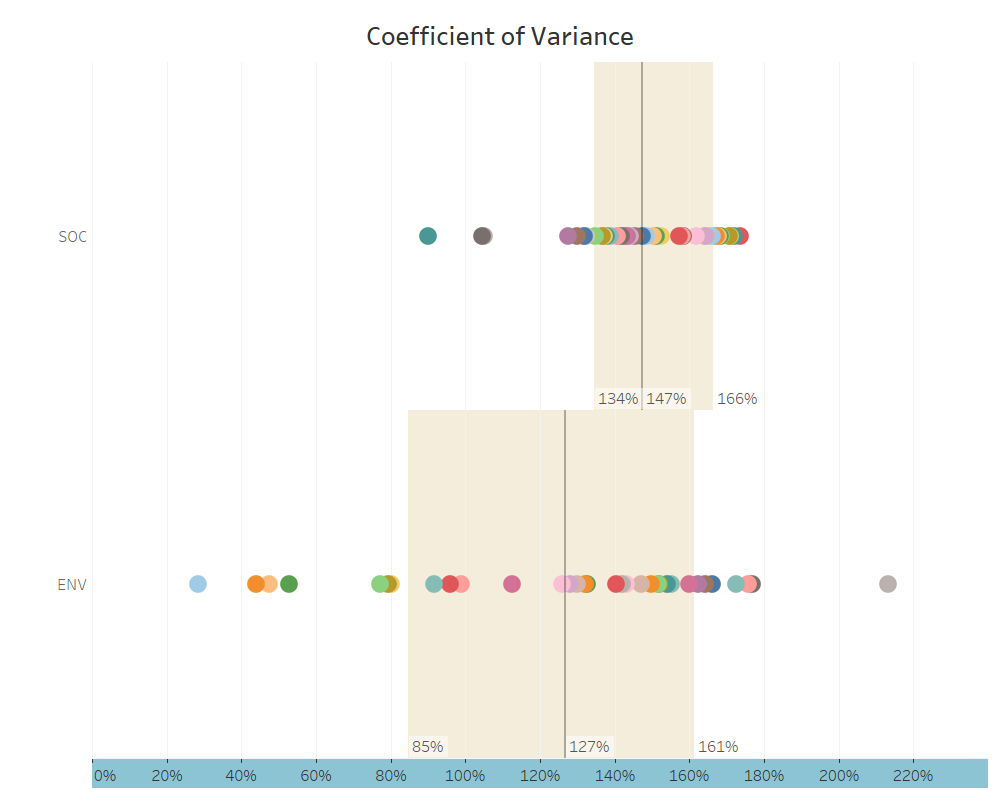
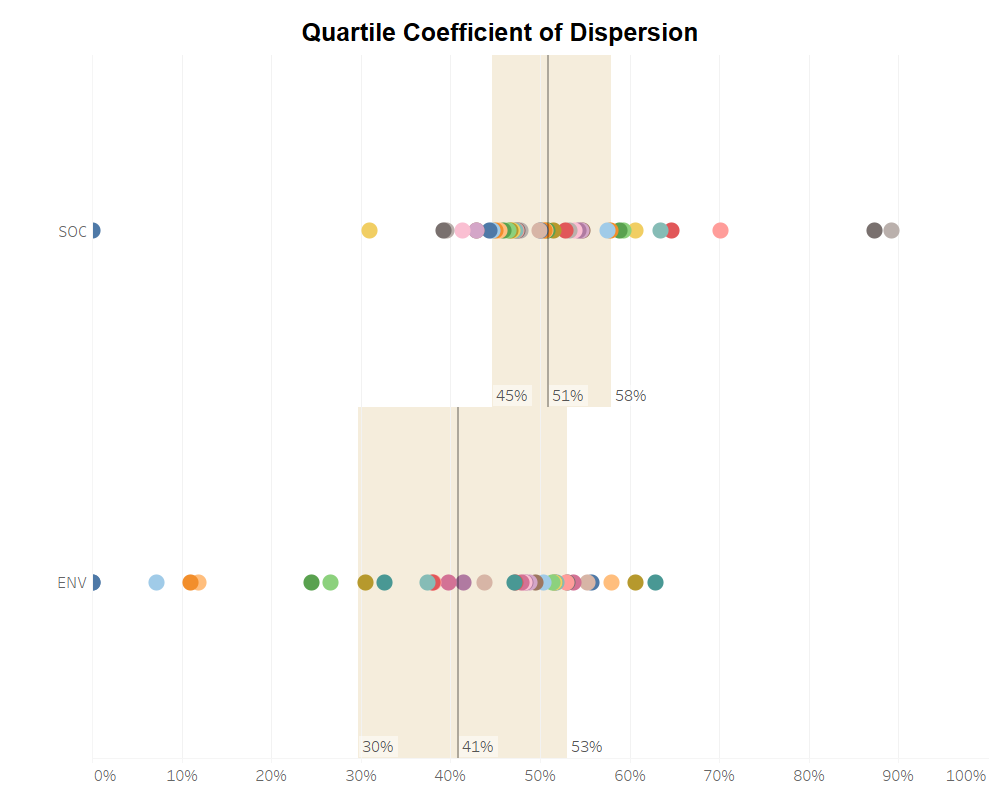


Figure 9 presents an overview of the quantile coefficients of dispersion (QCD) of the indicators grouped under social and environmental categories. QCD is a descriptive statistics that also measures the dispersion and is used to make comparisons within and between indicators. It is calculated using the first and the third quartile. In order to compare indicators between themselves, QCD values of different indicators are divided to obtain a percentage value which translates into how greater the QCD of a given indicator is in comparison to another. The higher the QCD, the highest the dispersion of data points observed in the indicator.

Figure 9: QCD – Indicator Level



The range of QCD varies from 0% (droughts, floods, extreme temperatures) to 63% for environmental indicators ( adjusted savings: particulate emissions damage. The range of QCD varies from 0% (diabetes prevalence) to 89% for mobile cellular subscriptions. Most of the environmental indicators vary between 30% to 53% QCD while the variation for social indicators ranges mostly from 45% to 58% which is closer to the mean of 51%. The average QCD of social indicators is 51%, 10 percentage points higher than the average QCD of environmental indicators. As indicated previously, the dispersion is higher in social indicators in comparison with environmental indicators.

A total of 34 indicators out of 76 have a QCD higher than 50%. 21 of these are social indicators and 13 of them are environmental indicators. Table 2 gives an overview of these indicators.

Table:

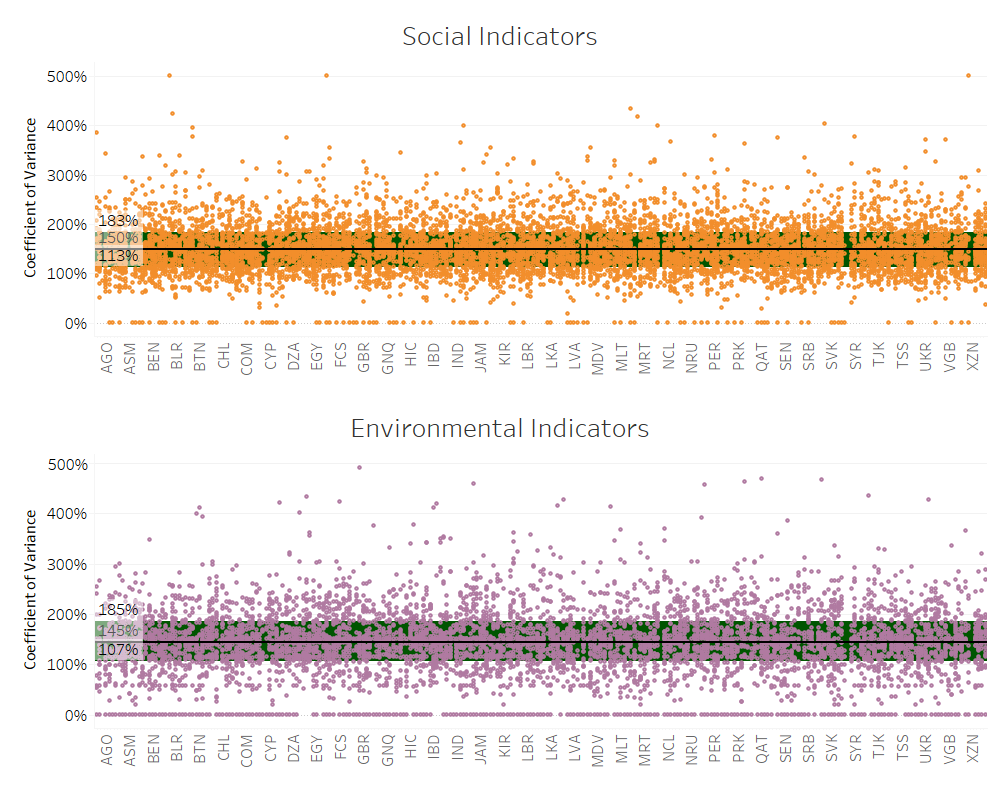
|  |  |  |
| --- | --- | --- |
| Name | Qcd | Sector |
| Adjusted savings: particulate emission damage (% of GNI) | 0.628942 | ENV |
| GDP per unit of energy use (PPP $ per kg of oil equivalent) | 0.606957 | ENV |
| Adjusted savings: mineral depletion (current US$) | 0.57969 | ENV |
| Terrestrial and marine protected areas (% of total territorial area) | 0.557717 | ENV |
| Total greenhouse gas emissions (kt of CO2 equivalent) | 0.552771 | ENV |
| Renewable internal freshwater resources per capita (cubic meters) | 0.537311 | ENV |
| Nitrous oxide emissions (thousand metric tons of CO2 equivalent) | 0.531214 | ENV |
| Adjusted savings: natural resources depletion (% of GNI) | 0.529903 | ENV |
| Agricultural land (% of land area) | 0.518166 | ENV |
| CO2 emissions (kt) | 0.516496 | ENV |
| Population density (people per sq. km of land area) | 0.516423 | ENV |
| Natural capital, subsoil assets: gas (constant 2014 US$) | 0.514481 | ENV |
| Natural capital, subsoil assets: oil (constant 2014 US$) | 0.503617 | ENV |
| Mobile cellular subscriptions (per 100 people) | 0.89252 | SOC |
| Internet users (per 1,000 people) | 0.872793 | SOC |
| GNI per capita, Atlas method (current US$) | 0.70166 | SOC |
| Fixed broadband subscriptions (per 100 people) | 0.646366 | SOC |
| Incidence of malaria (per 1,000 population at risk) | 0.634825 | SOC |
| Mortality rate, infant (per 1,000 live births) | 0.607052 | SOC |
| Maternal mortality ratio (modeled estimate, per 100,000 live births) | 0.593221 | SOC |
| School enrollment, tertiary (% gross) | 0.589285 | SOC |
| Agriculture, forestry, and fishing, value added (% of GDP) | 0.579061 | SOC |
| Prevalence of obesity, male (% of male population ages 18+) | 0.574815 | SOC |
| Poverty headcount ratio at $1.90 a day (2011 PPP) (% of population) | 0.547859 | SOC |
| Life expectancy at birth, total (years) | 0.546228 | SOC |
| Prevalence of obesity, female (% of female population ages 18+) | 0.542571 | SOC |
| Net official development assistance received (current US$) | 0.538606 | SOC |
| Prevalence of undernourishment (% of population) | 0.533133 | SOC |
| Population ages 65 and above (% of total) | 0.527949 | SOC |
| Proportion of seats held by women in national parliaments (%) | 0.514525 | SOC |
| Pupil-teacher ratio, primary | 0.50811 | SOC |
| Vulnerable employment, total (% of total employment) (modeled ILO estimate) | 0.506376 | SOC |
| Military expenditure (% of GDP) | 0.505033 | SOC |
| Tuberculosis prevalence rate, low uncertainty bound (per 1000,000 population, WHO) | 0.500098 | SOC |

As in the previous case with the indicators with the highest CV, improving the timelines of data production and update and coverage of these indicators will yield significant improvements of ESG indicators.

## Coefficient of Variation and Quartile Coefficient of Dispersion – Country Level

The average CV for environmental indicators is 145%, varying between 107% for the 20th percentile to 185% for the 80th percentile. The average CV for social indicators is very close to the one observed for environmental indicators, 150% while the 20th percentile is 113% and the 80th percentile is 183%.

Figure 10: Distribution of CV - Country Level



Both sets of indicators exhibits extreme cases of CV for indicator data at country level, around 500% such as access to electricity - BHS, prevalence of HIV – ESP. A number of 65 countries have social indictors in 77 instances where CV is higher than 300% for a total of 18 unique indicators. These indicators are presented in the table below.

Table 3: Social indicators with CV higher than 300% at country level.

|  |
| --- |
| Name |
| Access to electricity (% of population) |
| Prevalence of HIV, total (% of population ages 15-49) |
| School enrollment, primary (gross), gender parity index (GPI) |
| Net official development assistance received (current US$) |
| Hospital beds (per 1,000 people) |
| Physicians (per 1,000 people) |
| Proportion of seats held by women in national parliaments (%) |
| Agriculture, forestry, and fishing, value added (% of GDP) |
| Government expenditure on education, total (% of government expenditure) |
| Vulnerable employment, total (% of total employment) (modeled ILO estimate) |
| Military expenditure (% of GDP) |
| School enrollment, secondary (% net) |
| GINI index (World Bank estimate) |
| Net migration |
| Unemployment, female (% of female labor force) (modeled ILO estimate) |
| Maternal mortality ratio (modeled estimate, per 100,000 live births) |
| Unemployment, youth total (% of total labor force ages 15-24) (modeled ILO estimate) |

In the case of environmental indicators, there are a total of 83 instances at country level where CV is above 300% for a total of 65 unique countries and 17 unique indicators that are presented in the table below.

Table 4: Environmental indicators with CV higher than 300% at country level.

|  |
| --- |
| Name |
| Agricultural land (% of land area) |
| Fossil fuel energy consumption (% of total) |
| Renewable electricity output (% of total electricity output) |
| Electricity production from nuclear sources (% of total) |
| Land area (sq. km) |
| Renewable energy consumption (% of total final energy consumption) |
| Electricity production from coal sources (% of total) |
| Energy imports, net (% of energy use) |
| Food production index (2004-2006 = 100) |
| Energy intensity level of primary energy (MJ/$2011 PPP GDP) |
| CO2 emissions (metric tons per capita) |
| Energy use (kg of oil equivalent per capita) |
| Population density (people per sq. km of land area) |
| PM2.5 air pollution, mean annual exposure (micrograms per cubic meter) |
| Renewable internal freshwater resources, total (billion cubic meters) |
| Nitrous oxide emissions (thousand metric tons of CO2 equivalent) |
|  |

The distribution of quartile coefficients of dispersion is almost similar between the two types of indicators, social and environemntal. The average QCD of environmental indicators at country level is 49% while the average QCD of social indicators is 52%. Both sets of indicators vary between 28% and 33% respectively to 70% and 72% respectively for their 20th to 80th percentiles.

The more interesting occurrences at country level is the fact that there are 1034 data points of indicators at country level where QCD is above 80% for social indicators. This occurs in 254 individual countries and territories for a total of 38 indicators unique indicators as provided in the table below.

Figure 11: QCD – Country Level

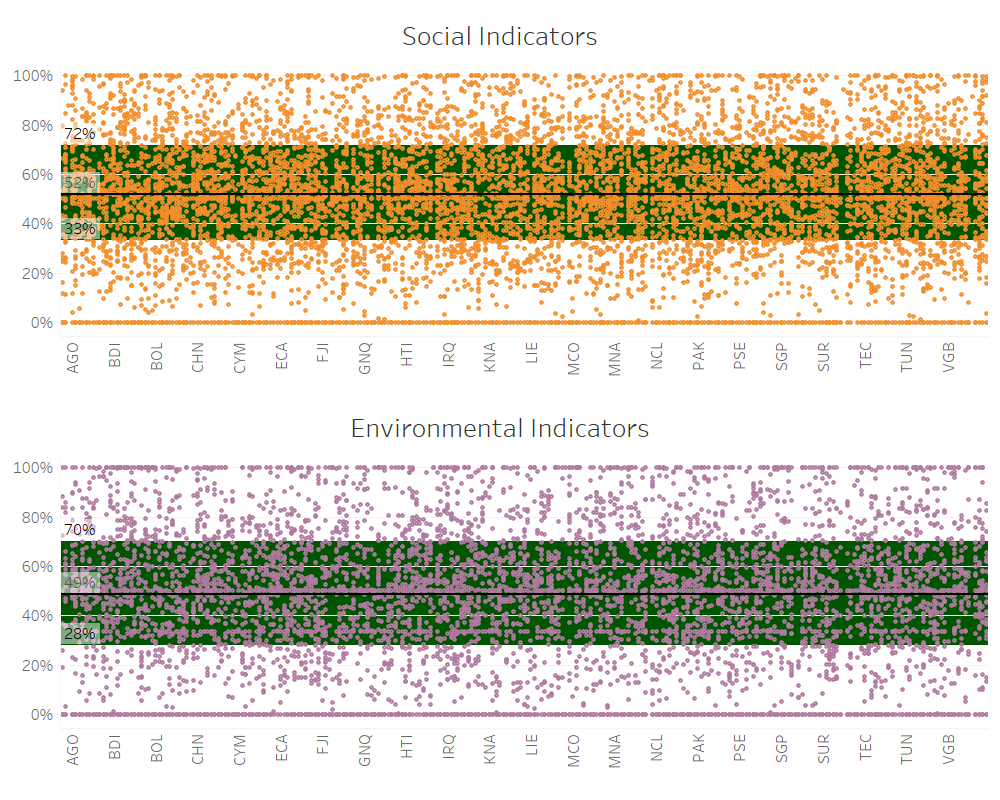


Table 5 : Social indicators with QCD above 80%

|  |
| --- |
| Name |
| Employment to population ratio, ages 15-24, total (%) (modeled ILO estimate) |
| Proportion of seats held by women in national parliaments (%) |
| School enrollment, tertiary (% gross) |
| Mobile cellular subscriptions (per 100 people) |
| GNI per capita, Atlas method (current US$) |
| Net official development assistance received (current US$) |
| Prevalence of HIV, total (% of population ages 15-49) |
| Life expectancy at birth, total (years) |
| Unemployment, youth total (% of total labor force ages 15-24) (modeled ILO estimate) |
| School enrollment, primary (gross), gender parity index (GPI) |
| Incidence of malaria (per 1,000 population at risk) |
| Internet users (per 1,000 people) |
| Pupil-teacher ratio, secondary |
| Fixed broadband subscriptions (per 100 people) |
| Unemployment, female (% of female labor force) (modeled ILO estimate) |
| Military expenditure (% of GDP) |
| Nurses and midwives (per 1,000 people) |
| Poverty headcount ratio at $1.90 a day (2011 PPP) (% of population) |
| Unemployment, total (% of total labor force) (modeled ILO estimate) |
| Technicians in R&D (per million people) |
| Maternal mortality ratio (modeled estimate, per 100,000 live births) |
| Pupil-teacher ratio, primary |
| Tuberculosis prevalence rate, low uncertainty bound (per 1000,000 population, WHO) |
| Vulnerable employment, total (% of total employment) (modeled ILO estimate) |
| Prevalence of undernourishment (% of population) |
| Physicians (per 1,000 people) |
| School enrollment, secondary (% net) |
| Researchers in R&D (per million people) |
| Agriculture, forestry, and fishing, value added (% of GDP) |
| Government expenditure on education, total (% of government expenditure) |
| Population ages 65 and above (% of total) |
| Net migration |
| Hospital beds (per 1,000 people) |
| Literacy rate, adult total (% of people ages 15 and above) |
| Poverty headcount ratio at national poverty lines (% of population) |
| Mortality rate, infant (per 1,000 live births) |
| GINI index (World Bank estimate) |
| Access to electricity (% of population) |

There are a total of 834 data points where QCD is higher than 80% for the environemntal indicators for country level data. This occurence happens in a total of 244 unique countries and territories and for a total of 33 unique indicators as described below.

Table 6: Environmental indicators with QCD higher than 80%

|  |
| --- |
| Name |
| Total greenhouse gas emissions (kt of CO2 equivalent) |
| Forest area (% of land area) |
| Agricultural land (% of land area) |
| GDP per unit of energy use (PPP $ per kg of oil equivalent) |
| Land area (sq. km) |
| Renewable electricity output (% of total electricity output) |
| CO2 emissions (metric tons per capita) |
| Energy imports, net (% of energy use) |
| Fossil fuel energy consumption (% of total) |
| CO2 emissions (kt) |
| Natural capital, subsoil assets: gas (constant 2014 US$) |
| Natural capital, subsoil assets: oil (constant 2014 US$) |
| Adjusted savings: particulate emission damage (% of GNI) |
| Energy use (kg of oil equivalent per capita) |
| Food production index (2004-2006 = 100) |
| Nitrous oxide emissions (thousand metric tons of CO2 equivalent) |
| Energy intensity level of primary energy (MJ/$2011 PPP GDP) |
| Adjusted savings: natural resources depletion (% of GNI) |
| Annual freshwater withdrawals, total (% of internal resources) |
| Renewable energy consumption (% of total final energy consumption) |
| Natural capital, subsoil assets: coal (constant 2014 US$) |
| Electricity production from coal sources (% of total) |
| GHG net emissions/removals by LUCF (Mt of CO2 equivalent) |
| Adjusted savings: energy depletion (% of GNI) |
| PM2.5 air pollution, mean annual exposure (micrograms per cubic meter) |
| Renewable internal freshwater resources, total (billion cubic meters) |
| Adjusted savings: net forest depletion (% of GNI) |
| Population density (people per sq. km of land area) |
| Renewable internal freshwater resources per capita (cubic meters) |
| Adjusted savings: mineral depletion (current US$) |
| \* |
| Terrestrial and marine protected areas (% of total territorial area) |

## Recommendations

A significant attention needs to be paid to data sourced from other organizations than the United Nations system and the World Bank. Data stemming from sources outsides these two bodies tend to have little 2018 data.

One of the issues related that need to be addressed is related to the publishing rights obtained from organizations such as IEA which although it has more recent data, it limits the publishing rights of the World Bank to one or two year below the most recent data in their databases.

Secondly, significant improvement of data availability and quality can be obtained if WBG sourced indicators are more timely updated, especially social and environmental indicators. The same can be obtained if data sourced from UN system is more often updated.

These being said, given that there is little variation between production frequency and update frequency at the World Bank of the ESG indicators, the most fruitful results might be obtained from increasing the rate of production of WBG indicators and from getting better publishing agreements from organizations such as IEA.

Data stemming from the UN system appears to be updated as soon as it is released by the organizations producing it and besides broad recommendations of more timely data updates and making clear the significance of the timeliness of production process for the ESG community, there is little room for improvement of these data streams.

Increasing the timeline or production, update and dissemination and coverage of indicators identified as having high CV and QCD can improve significantly the quality of the ESG datasets.

Finally, an ESG dedicated portal should include the indicators that have the best country and year coverage and these would be the most useful in the first phase for the ESG community. Additional indicators should be kept on an on hold position and as soon as their country coverage and most recent values improve should be introduced into the ESG portal.

## Conclusions

The current paper presented an overview of the 2018 paper and the options to improve the ESG data. It analyzed 138 ESG indicators assembling a database of 35 metadata variables.

The ESG data stems from three broad sources, UN system, WBG and other organizations, mostly international organizations. The vast majority of indicators is updated on a yearly basis and there is little variation between the entities collecting the data, processing and disseminating the data.

A total of 32 indicators out of 138 are discontinued due to indicators having changed to a different one, to indicators having used for a single purpose or indicators no longer being updated by the source organization. Furthermore, a significant number of indicators have low country coverage as they approach the most recent years.

Finally, a significant number of indicators come without 2018 data, their latest year with data available being 2017 or earlier. Most of these indicators are indicators sourced mostly from international organizations other than the UN or WBG. Indicators from UN system come second and those sourced from WBG third.

The first action point to improve data availability would be to increase the production timeline of WBG sourced indicators where only 10 out of 2019 indicators have 2018 data. Secondly, it will be to recommend the same measures for the indicators sourced from UN system. Lastly, obtaining better publishing rights from organizations such as IEA might contribute to having more recent data in the ESG datasets.