

# SDG data

---

The printable version is no longer supported and may have rendering errors. Please update your browser bookmarks and please use the default browser print function instead.

The International Futures (IFs) model has a unique dashboard for evaluating and forecasting countries' progress toward achieving the United Nations' (UN) 17 Sustainable Development Goals (SDGs). The SDGs comprise 169 targets with 239 associated indicators that are used to track country-level progress toward meeting each goal by 2030. Many of these indicators are further disaggregated by location, sex, and/or age group. The IFs SDG dashboard has two components, a data table and a graphical feature. It currently displays 280 historical and forecasted data series covering 79 targets and 95 indicators for 16 of the 17 SDGs.

Populating the SDG Form with historical series requires performing a data pull from the UN Statistics Division's (UNSD) SDG Indicators Global Database. The Global Database is the official source for SDG indicator data. It includes all available series for each indicator along with a metadata repository that provides detailed documentation on the definition, concepts, sources, and computations behind each variable.

☐

<b>Contents</b>
Identifying Series
IFs Series
UN Series
Naming New Variables
Data Acquisition
Series Exclusions
Documentation
DataDict
Country Concordance
Vetting
References

## Identifying Series

### IFs Series

---

The SDG Form includes 74 variables for which there are existing data series in IFs. Many, though not all, of these “IFs series” include both historical data and a forecast. These series

are incorporated into the form using data and forecasts from the model, and are regularly updated as part of model maintenance. Therefore, they were not updated during the initial SDG pull. Examples of these variables include percentage of the population living below the poverty line, stunting prevalence, and percentage of people with access to electricity, safe water, and sanitation.

## UN Series

---

The SDG Form includes 204 historical series from the UNSD's SDG database. The initial summer 2017 pull populated the SDG Form with "UN series" for Tier I indicators only. A Tier I indicator is one that is, "conceptually clear, has an internationally established methodology and standards are available, and data are regularly produced by countries for at least 50 per cent of countries and of the population in every region where the indicator is relevant<sup>[1]</sup>." The UN Development Program (UNDP) provided these classifications directly to the Pardee Center for the initial data pull.

## Naming New Variables

The addition of new SDG data series into IFs required assigning each a name that conform as closely as possible to the standard IFs naming convention. Broadly speaking, variable names include three to six shortened terms identifying the series' contents, beginning with the major category to which the series belongs and moving downward by conceptual significance for the variable. For example, the series "PopFoodInsecSevFemaleFIES" begins with "Pop" for population, and describes the UN series "estimated number of population in severe food insecurity," which is disaggregated by gender and measured using the Food Insecurity Experience Scale (FIES).

## Data Acquisition

The UNSD's SDG database can be sorted geographically or by SDG target. For the initial pull, the data were downloaded by target, not geographically. The team reviewed each target for which Tier I indicators are currently available, and, except for in a rare number of series (See Series Exclusions), pulled all series in the database that were provided for these targets at the country level. To pull a target's associated series, the target was selected and downloaded as an Excel using the "Excel" button that appears at the top left once the data has fully loaded. The series were then organized and cleaned for import using standard best practices.

## Series Exclusions

Data must be available at the country-level for a series to be used in IFs, resulting in the exclusion of a small number of series that are only available by region or other grouping. In addition, when a large number of series was associated with a target due to disaggregation, (more than ~25) some were dropped for simplification or combined using an index. In these cases, the data team aimed to maintain those series that most directly captured a given target. These cases are noted in the full pull documentation and/or the DataDict. Examples of targets for which not all series were pulled are "8.7.1 Proportion and number of children aged 5-17 years engaged in child labour, by sex and age" and "6.b.1 Proportion of local

administrative units with established and operational policies and procedures for participation of local communities in water and sanitation management."

## Documentation

The team documented the initial pull in a master tracking sheet that was routinely reviewed for accuracy and used to track questions, challenges, and judgement calls at the series- and/or target-level. The master sheet also captured the information used to complete the DataDict entry for each series, and included an additional tab with guidance for team members on how to uniformly complete it. The addition of new series to IFs requires creating new DataDict entries that include the relevant metadata on each series. Ensuring completeness and accuracy of documentation was therefore a major feature of the project.

## DataDict

As noted, the initial SDG pull required developing original DataDict entries for the new SDG variables. Table 1 details the guidelines that the data team used to develop accurate and consistent metadata for each new SDG variable that was pulled into IFs from the UNSD database.

**Table 1**

ROW IN DATADICT	DESCRIPTION	EXAMPLE
<b>Variable Name</b>	New variable name you assign; look at similar variables in the DataDict, always start with major category and keep to under 30 characters	AidDonEdScholar
<b>Table</b>	Automatically generated	SeriesAidDonEdScholar
<b>Group</b>	New group(s) you assign; look at similar variables in the DataDict and match to existing groups (e.g. 'Education, Population' for education series)	Economic, SocioPolitical
<b>Subgroup</b>	New subgroup you assign; look at similar variables in the DataDict and match to existing Subgroups (e.g. Education or Attainment for education variables)	Aid
<b>Series</b>	Yes	Yes
<b>CoVaTra</b>	No	No
<b>Cohort</b>	No	No
<b>Definition</b>	New, brief definition you write using metadata for the indicator on the UN website	Gross disbursements of total ODA from all donors for scholarships

<b>Extended Source Definition</b>	Full indicator text including "SDG" and the indicator number at the beginning	SDG 4.b.1 Volume of official development assistance flows for scholarships by sector and type of study
<b>Units</b>	From the Unit column in the UN database; accounts for any transformations	Billion \$ (2011)
<b>CURRENCY</b>	"True" if value is in dollars or another currency (e.g. expenditure on X)	True
<b>Years</b>	Automatically generated, but check to make sure is accurate	2006-2015
<b>Source</b>	"UN SDG Indicators Global Database"	UN SDG Indicators Global Database
<b>Original Source</b>	Original source institution listed in the metadata for the indicator	Organization for Economic Development and Cooperation (OECD)
<b>Notes</b>	Puller and vetter initials, any notes you have, and series type ("official indicator" or "additional series")	AJM; converted to 2011 dollars; official indicator
<b>Last IFs Update</b>	Automatically generated	2017/07/27
<b>Aggregation</b>	Options - POP (population), GDP (GDP), LND (land area), AVG (average), SUM (sum); this tells the model how to aggregate country values to groups such as a region or the world. It is based on the nature of the variable. For examples, anything as a percent of GDP is aggregated by GDP, any relative measure based on population (e.g. teachers per 1,000 students) is aggregated as population, simple counts are aggregated with a sum (e.g. millions of people), geographic variables based on land units (e.g. crop yield per hectare) are aggregated by land area, and indices use an average (e.g. gender parity indices)	SUM

<b>Disaggregation</b>	Automatically generated to be GDP. Options - POP (population), GDP (GDP), LND (land area), AVG (average), SUM (sum); this tells the model how to disaggregate country values to subnational bodies when the model is broken out into its subnational form. Leave as GDP unless you have a good reason to change it	GDP
<b>TreatNullsAs0s</b>	Automatically generated	FALSE
<b>Proprietary</b>	Automatically generated	FALSE
<b>Name in Source</b>	From Series Description column in UN database	Total official flows for scholarships, by recipient
<b>Used in Preprocessor</b>	Automatically generated	FALSE
<b>UsedInPreprocessorFileName</b>	Leave blank	
<b>CompareOtherForecasts</b>	Automatically generated	FALSE
<b>Code in Source</b>	Use Series Code from UN database	DC_TOF_SCHIPSL
<b>Decimal Places</b>	Keep as is or enter 4 or 5 depending on the necessary level of precision (if most values are long decimals use more)	4
<b>Country Concordance</b>	"UN Stats (SDGs)"	UN Stats (SDGs)
<b>Formula</b>	Leave blank	

## Country Concordance

The SDG pull required developing a new country concordance table including the full, original country list used in the UNSD SDG database. Data team members used the same list to import and vet the data, and it has been added permanently to the IFs system for the ease of future updates to these series.

## Vetting

Vetting imported series is a standard and essential practice in all data pulls; however, the process for vetting new series is distinct from that used when vetting existing series. Namely, it requires greater scrutiny of the DataDict and more attention to the accuracy of the data as compared to the original source rather than the data that was previously in IFs. In the SDG pull, vetters examined the following for each series:

- The completeness and accuracy of the DataDict; checking the information provided against the master tracking sheet and against the information in the UNSD database, including the metadata (e.g. that each series has the correct definition and source ID), and that the standard format was used for each column
- The logical consistency of subjective decisions about DataDict entries, such as assigned

variable names and aggregation rules, and group/subgroup designations

- The completeness of the new country concordance list; checking for erroneous country exclusions and whether any countries were inaccurately named or switched (e.g. North and South Korea)
- The accuracy of the data; checking that the data matches the original series in the database it is meant to represent and that none were misnamed or switched during import
- The completeness of the data; checking that no years or values were dropped during import because of issues such as improper formatting of the source Excel sheet
- The completeness and accuracy of any transformations that were made; namely, converting 2015 to 2011 USD (for consistency with other monetary series in the model) and the creation of indices

## References

1. "Tier Classification for Global SDG Indicators." (2017). Retrieved September 1, 2017, from <https://unstats.un.org/sdgs/iaeg-sdgs/tier-classification/>.

---

Retrieved from "https://pardeewiki.du.edu//index.php?title=SDG\_data&oldid=8368"

---

**This page was last edited on 8 September 2017, at 15:30.**