

Estimating Baseline Aid to the Sustainable Development Goals

Matthew DiLorenzo*, AidData

Sid Ghose, AidData

Jennifer Turner, AidData

Introduction

In 2015, the international community, spearheaded by the United Nations, adopted an ambitious set of 17 Sustainable Development Goals (SDGs). These goals will help shape the next 15 years of development financing priorities. By some estimates, achieving the SDGs will require the international community to mobilize an additional \$2.5 trillion USD per year.¹ Tracking and analyzing this funding will be central to measuring progress, crowding in resources to priority areas, and helping decision-makers make more informed choices.

Unfortunately, current data available on SDG financing are not fit for purpose. Aid reporting systems do not capture sufficient information on the distribution of financing for the SDGs. The AidData Sustainable Development Goals (SDGs) Estimates attempt to fill this gap by providing project-level estimates of contributions to the SDGs (and their associated targets) using development project descriptions. This methodology lets us see where development financing is targeted, allowing comparisons among SDG goals and individual SDG targets.

In prioritizing future financing for the SDGs, it is important to develop a baseline understanding of recent historical trends in financing to the goals. This research note first describes an iteration of a methodology for estimating aid contributions to the SDGs throughout the period of 2000 to 2013 that AidData employed for its flagship report *At the Crossroads*. We then reflect on some methodological and conceptual issues involved in estimating funding for the SDGs. We conclude by discussing our revised, in-progress methodology that we will use in estimating financing for the SDGs in the post-2015 era.

Method

Our methodology is based on an analysis of development project descriptions and builds on an existing activity coding schema developed at AidData, through which student researchers assigned codes based on a project's activities. Students had previously assigned activities and purposes to over 800,000 project descriptions in AidData's core research release (v3.0).

* Corresponding author. E-mail: mdilorenzo@aiddata.wm.edu. Phone: (978) 204-4309. AidData at the Institute for the Theory and Practice of International Relations. College of William and Mary. 427 Scotland Street, Williamsburg VA 23185, USA.

¹ United Nations Conference on Trade and Development 2014

This methodology involves three critical steps: (1) creating a mapping between activity codes and SDG targets, (2) splitting an aid project across designated activities, and (3) splitting activity amounts across SDG targets, as an activity may be linked to multiple targets. From these calculations, we can sum target-level estimates up to the goal level.

We incorporate as much information about an aid project as is possible in generating SDG estimates. Where activity codes are available, we use those as an intermediary to link SDG targets to projects. Where activity codes are unavailable, we use purpose codes, which are not as granular as activity codes. When only purpose codes are available, we generate estimates based on a naïve diffuse assumption about what activities were involved in a project with a given purpose code.² Both methods are based on an initial mapping between the AidData activity codes and the SDG targets, as described below.

Preliminary Method

This section describes the previous iteration of our SDG coding methodology. While we are in the process of implementing a revised version of this methodology based on direct-to-SDG coding, the process of developing the method described here was instructive in terms of thinking about how to track financing to the SDGs, as well as for thinking about how to interpret potential gaps in funding for certain targets.

Mapping Activity Codes to SDG Targets

As a first step, and core to our methodology for linking development projects to the SDGs, we mapped SDG targets to AidData activity codes. Activity codes are based on the OECD's Creditor Reporting System (CRS) sector and purpose codes, but go one step deeper, providing a more disaggregated breakdown of development activities that are relevant to each CRS code.

To link AidData activities to SDG targets, student coders went through the 544 AidData activity codes and assigned SDG targets to each activity.³ Multiple coders contributed to the initial round of coding, with different students coding different sections of activities, resulting in a single mapping of activity codes to SDG targets. Once this initial round of coding was complete, two student research assistants from AidData's Research and Evaluation Unit reviewed the coding and made suggested changes. Three members of the AidData Policy Analysis Unit then reviewed the coding and arbitrated cases of disagreement in consultation with the Policy Analysis Team to resolve unclear cases.

Students linked AidData activity codes to SDG targets using a few guidelines. First, we required that students take into account both the text of the activity code and the purpose category name. Second, cases in which the coders were unable to find a link between an activity and a specific target but felt that an activity was relevant to an overall goal were coded to the number of the SDG and appended with ".0" to indicate that they were linked to the goal but not to any specific target.⁴ Third, we advised students that coding from specific activities to general targets was appropriate, while coding from general or vague activities to specific targets was not

² In particular, we assume that any activity that appears under that purpose code is equally likely to have been part of a project.

³ There are a total of 17 SDGs with 169 associated targets across the 17 goals.

⁴ For example, a project on health policy that could not be linked to any specific goal 3 target would be coded as 3.0, indicating that it was relevant to SDG3 but not any specific target under SDG3.

appropriate. Fourth, we directed the students to link activities to as many SDG targets as appropriate. Finally, we instructed students not to make inferences about the likely effects of aid projects with given activities in terms of how they would achieve the SDG targets. In other words, we wanted students to avoid imagining a chain of events that might potentially link an activity to outcomes that would not be directly related to that activity. This likely provides a conservative estimate of funding that would contribute to the various SDGs and explains some of the coding difficulties and features of the data we describe later.

Within groups of SDG targets, there is a distinction between level 1 and level 2 targets. The level 1 targets pertain to more specific, concrete aspirations with associated deadlines, while the level 2 targets summarize broader, open-ended aspirations that apply mainly to developed countries. We include both sets of targets in our coding scheme. Some of the level 2 SDG targets are focused on increasing international support (i.e., foreign aid) for various goals, and in some cases where these targets also allude to broader aspirations, we code specific activities to these targets. Otherwise, our coding is focused on making more direct connections between foreign aid projects and the specific level 1 SDG targets. Researchers interested in studying trends in contributions to these level 2 projects would likely be interested in summing all contributions to any target under a given goal.

Splitting Project Values Across Activities

After mapping activity codes and SDG targets, the next step in estimating historical funding to the SDGs is splitting an aid project's value across activities. Although most development projects in AidData's core research release (v3.0) have been activity coded, project values have not been split across activities. We assume that dollar amounts for a project are distributed equally across activities. Although projects will actually have different distributions of dollar amounts across activities in practice, there is unfortunately no reliable way to infer this given existing data sources. While this requires a strong assumption about the relative prominence of different activities within a project, this approach is similar to previously published research on tracking aid projects for nutrition (Ickes, Trichler, and Parks 2015).

Distributing Activity-dollars Across SDGs

Having split the dollar value of a project across unique activities, the next task is to distribute those activity-dollar amounts across the SDGs. Since we are primarily interested in the *goals* rather than *targets*, we "roll up" to goals from targets and weight an activity's contribution to the SDGs proportional to how often the targets associated with a goal appear in the mapping between that activity and the targets. For example, if activity a_j with $j = \{1, \dots, 544\}$ is linked to targets 1.1, 1.2, 1.3, 2.1, and 3.1, we say that $\frac{3}{5}$ of activity j contributed to SDG 1, $\frac{1}{5}$ to SDG 2, and $\frac{1}{5}$ to SDG 3. For each activity, we have a vector of weights of length 17 (corresponding to the seventeen SDGs), that we can call ω_j . This vector satisfies the condition that $\omega_j \geq 0$ and $\sum \omega_j \in \{0, 1\}$. In words, every element of ω_j is greater than or equal to zero, and the sum of the seventeen entries must be either 0 or 1. This means that if an activity is linked to at least one target, the entire value of that "activity-dollar amount" will be distributed to the SDGs, either to one or multiple goals. In this case, $\sum \omega_j = 1$. If an activity is not linked to any targets, then all of the entries in ω_j are zeroes, and $\sum \omega_j = 0$.

Cases Without Activity Codes

Approximately 58% of projects in AidData's core research release have been activity coded. The remaining projects that do not have activity codes have purpose codes, which are less

granular than activity codes. Where only purpose codes are available, we generate estimates based on a naïve diffuse assumption about what activities were involved in a project with a given purpose code by compiling the list of activity codes “under” a given purpose code along with the list of targets associated with those activities. This is a second-best solution, but a reasonable one given the limitations of the data. We strip out the “target-level” information to obtain a list of SDGs for each activity. We estimate the weights to each goal as the proportion of times that the goal appears for any activity under a given purpose code out of the total “goal appearances” for a purpose code. Finally, we scale these weights with a proportional measure that indicates the ratio of activities that are linked to at least one SDG to the total number of SDGs under a purpose code.

For example, for purpose code *A*, we might have activities *A.1*, *A.2*, and *A.3*.

Activity	SDG Targets
A.1	1.1, 2.1
A.2	(none)
A.3	1.1, 3.1

Stripping away the target-level information, we have the following goal “appearances”: 1, 2, 1, 3. Based on this, we assign Goal 1 a weight of $\frac{1}{2}$, Goal 2 a weight of $\frac{1}{4}$, and Goal 3 a weight of $\frac{1}{4}$. However, since activity *A.2* is not linked to any targets, we rescale these weights. Since two thirds of the activities under purpose code *A* contribute to the SDGs, we multiply each weight by $\frac{2}{3}$. This results in the final weights for purpose code *A*: Goal 1 gets a weight of $\frac{1}{3}$, Goal 2 a weight of $\frac{1}{6}$, and Goal 3 a weight of $\frac{1}{6}$. We use this same weighting scheme when generating the target-level estimates, but instead of 17 categories, there are 169, corresponding to the SDG targets.

Coding Decisions and Difficulties

This section details some of the issues we encountered and decisions we made in mapping activity codes to SDG targets that may be relevant for analysts interested in using our data or tracking funding to the SDGs.

Requiring Activities to be “Sustainable.” Many of the SDG targets make specific reference to “sustainable” investments. We generally do not require that activities have a “sustainable” focus even when some of the targets include references to sustainability. For example, Target 2.4 seeks to “ensure sustainable food production systems” by 2030. We link a number of activities to this target that are relevant for food production systems, not all of which have a sustainable component. Similarly, Target 8.9 aims to promote “sustainable” tourism. We link the AidData “Tourism policy and administrative management” activities (codes 33210.01 through 33210.04) to this target.

How Will the Activity Likely Affect the Target? We attempt to account for whether the activity in question should generally advance or work against progress on the target in question, and we link activities to targets only when we think there is a good argument that the activity should advance the goal. In other words, we are not interested in tagging activities that are simply *relevant* for a target, but activities that will plausibly contribute to *progress* on that target. For example, we do not count activity 31140.05 “Ground Water Exploitation” (for agriculture) as contributing to the water SDG or any of its targets. The same goes for activities related to

increasing fish catch. These activities are clearly *relevant* for the SDG targets related to sustainable water resources and fisheries, but, if anything, they would seem to work against those targets. So that our baseline estimates of funding *for* the SDGs are not biased by activities that are nominally related to the SDG targets but inconsistent with their aims, we made an effort to discern their likely impacts on SDG targets.

Indicators as a Clue for Target Intent. In cases where the wording of SDG targets is vague, we looked to the SDG indicators to get a better sense of the priority of the target. For example, Target 9.3 aims to “Increase the access of small-scale industrial and other enterprises, in particular in developing countries, to financial services, including affordable credit, and their integration into value chains and markets.” From the text of the target alone, it is unclear whether the aim is limited to “small-scale” enterprises. The indicators for this target make clear that “small-scale” is, in fact, an essential part of the target.

Budget Support and Humanitarian Aid. Two major categories of aid do not get mapped to the SDG targets in our method: budget support and humanitarian aid. Given research that suggests recipient governments often do not put budget support to use in ways that promote development goals and the fact that it is impossible to know in advance how recipients will use budget support, we do not directly link budget support to any SDG. Although humanitarian aid is relevant in the short-term for many SDGs, the focus of the SDGs is on factors that promote sustainable development in the long-term. As such, we did not link humanitarian aid directly to any of the targets, even though it is easy to imagine ways in which humanitarian aid might indirectly advance certain goals. This is also consistent with OECD reporting standards, which track humanitarian assistance separately from other forms of official development assistance.

Connection to Recipient Country. In some cases, we do not code when it is not clear that the activity benefits the target/goal *in the recipient country in question*. For example, building fertilizer plants (activity 32120.10) may produce fertilizer that is used in the recipient country or produce fertilizer that the recipient country then exports. For this reason, we linked the activity to Target 9.2 for industrial development but not to any targets pertaining to agricultural productivity. This falls under our general rule for coding targets to activities only when there is a reasonably close and direct link between the activity and the target in question. An additional implication of this coding decision is that there are few activities tagged to targets that emphasize progress on problems that are transnational in nature (e.g., oceans, climate change).

Weaknesses and Challenges

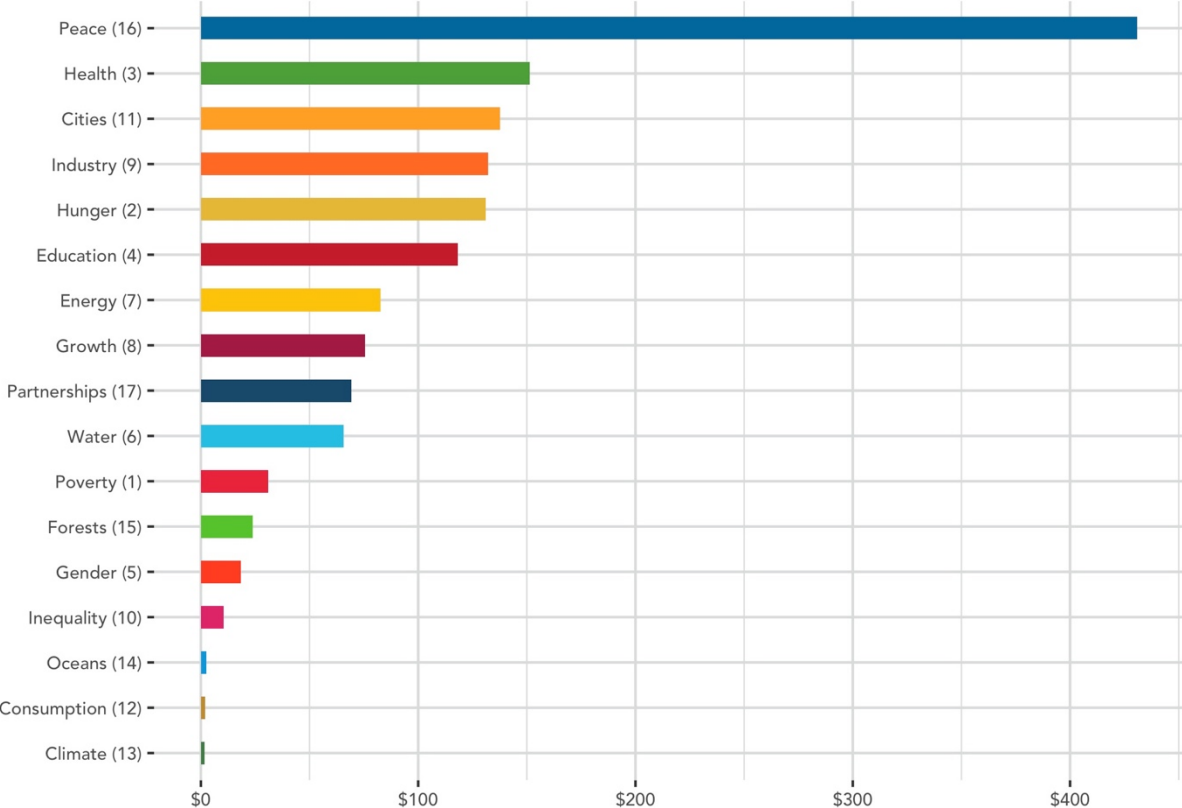
The exercise of developing a map between existing aid categorization schemes (in this case, the AidData activity codes) and SDG targets highlights a number of limitations and challenges in tracking financing to the SDGs. In some cases, this may simply be a reflection of donor practices; donors may not give aid in ways that map well to SDG targets. In other cases, there is a more distinct mismatch between activity codes and SDG targets that likely results in an under-counting of financing to certain goals.

AidData’s activity coding scheme, though more granular than alternative aid coding schemes, is not consistently well-aligned with the SDGs. As a result, our methodology is more reliable at tracking financing to certain goals, such as SDG3 (health) and SDG4 (education), for which activity codes map more neatly to SDG targets.

Little funding is reported for certain SDGs, particularly those focused on the environment (SDGs 13, 14, and 15). While this may reflect lower donor priorities given to the environment during the MDG era, it is also likely due to a mismatch between activity codes and the resulting inability to

link these projects to SDG targets. The SDGs lay out a very specific set of environmental targets under distinct goals related to climate change, oceans, and land ecosystems, while AidData’s activity coding scheme groups many of these projects together under categories such as “General environmental protection.” Since this category is too broad to be linked to any specific SDG, projects that were assigned this activity code are not counted as contributing to the SDGs. For example, a project on preserving marine ecosystems is not counted as contributing to SDG14 even though there is an SDG target directly related to that activity (Target 14.2) because the only relevant activity code is “general environmental protection,” which is too vague to be linked to SDG14.

Figure 1: Historical ODA Financing Towards the SDGs, 2000-2013



Source: OECD CRS (2000-13); Note: All figures are in billions of constant 2011 US\$

This same issue exists for a number of different activity coding categories, such as rural development and population. Even though many projects coded to these categories are seemingly relevant to the SDGs, because they couldn’t be linked to specific SDG goals or targets, the projects are not included as contributing to the SDGs.

Next Steps

In light of these limitations, AidData is moving toward a process of directly coding aid projects to the SDGs. That is, rather than the intermediate step of having human coders assign activity codes to projects and then linking projects to SDGs based on activities, human coders will read project descriptions and assign SDG targets directly to the projects. With a direct coding scheme, projects described as “rural development” would still not be able to be coded, but project descriptions describing the types of activities undertaken in a rural development project, like irrigation development or agriculture training, could be coded to a specific SDG target. This “direct coding” methodology will address some of the shortcomings related to the mismatch between SDGs and AidData activity codes and has the potential to more accurately track financing to the SDGs.

In developing the direct coding methodology, we used the original mapping of activity codes to SDG targets as a starting point for developing a codebook to guide student coders. Since the text of SDG targets is often complex and subject to differing interpretations, members of AidData’s Policy Analysis Unit also developed summaries and keywords relevant to each target that were then used as a basis for assigning codes. As with the original methodology, coders were instructed to code projects to the number of the SDG, appended with “.0,” if they are linked to the goal but not to any specific target. They also were instructed to focus on the most direct link with project activities when assigning SDG targets. For example, many project descriptions state that an aim of the project is to reduce poverty or reduce hunger. However, the activities described are more directly related to agricultural productivity or job training. Students were told to focus on this direct activity rather than desired outcomes that could potentially be attributed to an activity. As a result, certain targets, like Target 1.2 (By 2030, reduce at least by half the proportion of men, women and children of all ages living in poverty in all its dimensions according to national definitions) were assigned to relatively few projects.

We continued to refine and update our codebook during an initial testing phase, with members of the AidData Policy Analysis Unit reviewing students’ coding and making recommendations about how to code borderline projects.

Remaining Challenges and Unresolved Issues

In developing the direct coding scheme, several outstanding issues remain, particularly surrounding what projects and activities count as relevant to the SDGs and how to split financing among a project’s different activities. In this section, we detail several of these issues.

Lack of Alignment Between Donors and the SDG Agenda. In some cases, donors give aid in ways that do not map well to SDG targets. The environmental agenda through the MDG period was more vague, resulting in many broad projects descriptions that are not easily linked to the specific environmental SDGs or targets. Other problems arise from the fact that some donors give few details about their projects, preferring categories like general “development assistance” that cannot be linked to any specific SDG. While many of these projects are likely relevant to the SDGs, we have not found a way to reliably assign them to specific SDG targets or goals.

The Interrelated Nature of the SDGs. A key challenge is that the interrelated nature of the targets themselves— by design of the international community – makes tracking discrete project amounts to individual targets difficult. While progress on one target may reinforce progress on other targets, for the purpose of tracking finance to the targets, it seems unreasonable to assume that project interventions will have their intended effects. For this reason, endeavors focused on tracking financing for the targets should strive to be conservative. Whether progress

spills over to other targets is an empirical question that will be more easily answered further into the SDG era.

Projects with Multiple Activities. Project descriptions often include long lists of ancillary activities that may not be core aspects of the project. A project focused on building a dam may also include small side projects to provide supplies to a local school and health clinic. AidData's activity coding scheme coded every project activity that could be identified in a project description. However, if all three parts of the above project are coded, project financing would be split evenly among the three, significantly overcounting financing to education and health and undercounting financing to the dam project. In the pilot phase of the direct coding methodology, students were instructed to only code to project activities that were considered "significant" to the project.

How to Divide a Project's Financing Among Different SDGs. While we made the decision to split financing evenly among the different activities (and their related SDGs) in our original methodology, questions remain as to whether this is the right decision in all circumstances.

Projects can be coded to multiple SDGs for two reasons. In the first case, a single activity is relevant to multiple SDG targets. Target 3.7 and Target 5.6. both address sexual and reproductive healthcare, so projects relevant to reproductive health are systematically coded to both targets. Similarly, projects to build infrastructure, like roads, hospitals, and electric grids, are coded to both the relevant sectoral targets as well as Target 9.1 for infrastructure. In these cases, splitting financing among the targets undercounts relevant financing for each individual target. If someone is interested in analyzing the total amount of financing going to Target 3.7, it would seem reasonable to assign the full value of the project to that target, rather than only half the value as would happen if financing is split.

In the second case, projects have discrete activities that are coded separately. For a project that provides both job training and water and sanitation services, splitting the financing between the SDGs associated with the two activities seems more reasonable than assigning the full value of the project to each related SDG.

References

Ickes, Scott B, Rachel B Trichler and Bradley C Parks. 2015. "Building a Stronger System for Tracking Nutrition-Sensitive Spending A Methodology and Estimate of Global Spending for Nutrition-Sensitive Foreign Aid." *Food and nutrition bulletin* 36(4):520–533.

"World investment report 2014." United Nations Conference on Trade and Development (UNCTAD) World Investment Report (WIR), 2014.

Appendix A

Documentation

The functions to perform this operation are included in the R package `aidtools` via the `sdg_coder()` and `target_coder()` functions. This package is hosted on GitHub and can be installed and used with the `devtools` package in R as follows:

```
library(devtools)
install_github(mdilorenzo/aidtools)
library(aidtools)
sdg_coded_data <- sdg_coder(data)
```

In using the `sdg_coder()` and `target_coder()` functions, it is important to treat cases without activity codes separately. For example, if a data set has a total of 10 projects, and only 5 of those have activity codes, the user must run the `sdg_coder()` and/or `target_coder()` functions separately on the 5 cases with activity codes and on the 5 cases *without* activity codes, then combine those estimates together into a final data frame. For a data frame object called “data” with a logical variable called `activity_coded` that indicates whether or not a project is activity coded⁵, this would look something like this in R:

```
estimates_1 <- sdg_coder(data[data$activity_coded, ])
estimates_2 <- sdg_coder(data[!data$activity_coded, ], coalesced_purpose = TRUE)
final_data <- rbind(estimates_1, estimates_2)
```

For target-level estimates, the user would simply replace `sdg_coder()` with `target_coder()`.

⁵ One method in R: `data$activity_coded <- nchar(data$aiddata_activity_codes) > 0.`