Indian Development Finance Dataset, Version 1.0

The Indian Development Finance Dataset contains all Indian development cooperation projects reported in official documents of India's Ministry of External Affairs (MEA) and the Export-Import (Exim) Bank of India between 2007 and 2014. The project locations are geocoded with various levels of precision. We release the data in four different datasets at different levels of aggregation and detail. The full data cover 124 unique Exim projects in 247 locations, and 1,072 MEA projects in 4,064 locations in 169 countries. At the first administrative unit level the data cover 642 unique projects in 2,174 locations, while there are 615 unique projects in 1,971 locations at the second administrative unit level – both in 94 countries.
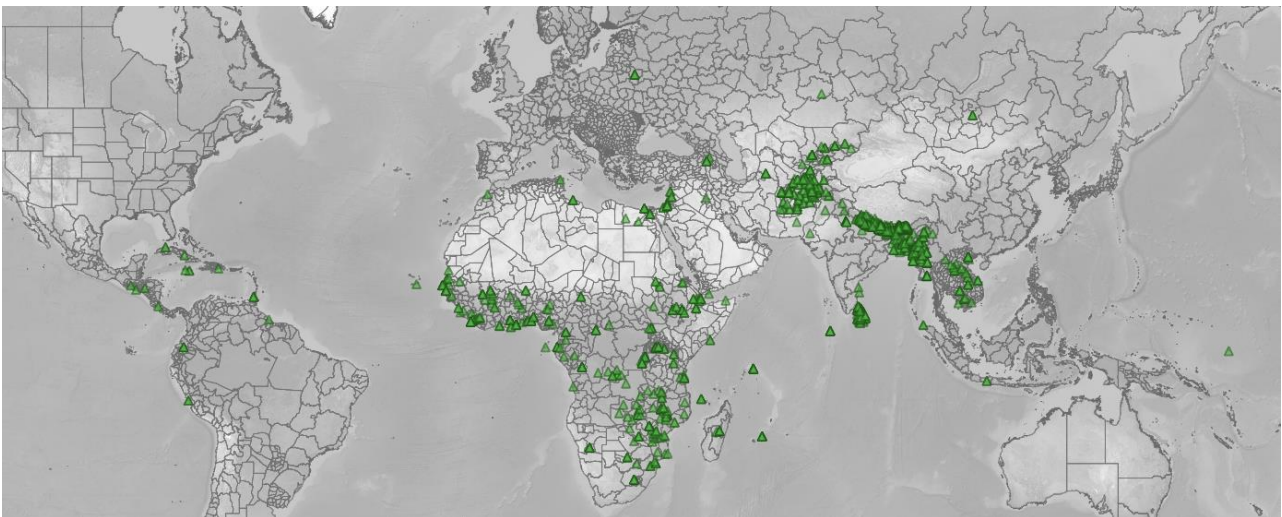

Figure 1: Geographical distribution of projects by the Indian MEA and Exim Bank, 2007-2014.

Citation

When using either dataset, please cite as:

> Asmus, G., Eichenauer, V.Z., Fuchs, A., and Parks, B. (2021). Does India Use Development Finance to Compete with China? A Subnational Analysis. AidData Working Paper #110. Williamsburg, VA: AidData at William & Mary.

About the data collection process

In collaboration with AidData, a research lab at William & Mary, we collected project-level information on the two major Indian agencies that provide official financing to countries with lower or middle incomes: the MEA and the Exim Bank.

The first step of the data collection process was to retrieve project-level information from official government sources. We obtained information about MEA aid projects from the ministry's "outcome budget" documents, and information about Exim Bank credit-financed projects from the bank's press releases announcing new lines of credit. From these documents, we retrieved information about project titles and descriptions, commitment years, recipient countries, financing types, commitment, and disbursement amounts. These steps resulted in a dataset consisting of 1,196 projects in 169 countries, geocoded at various levels of geographic precision, from 2007 to 2014. In total, the monetary value of these projects amounts to US$14.58 billion (in constant 2014 US$), of which US$9.56 billion originated from Exim Bank and US$5.02 billion from MEA.

The second step of the data collection process was to assign all projects to (sub)sectors according to the OECD's Development Assistance Committee (DAC) definitions, such as "Education" and "Humanitarian Assistance" (see Figure 2 for a comparison).
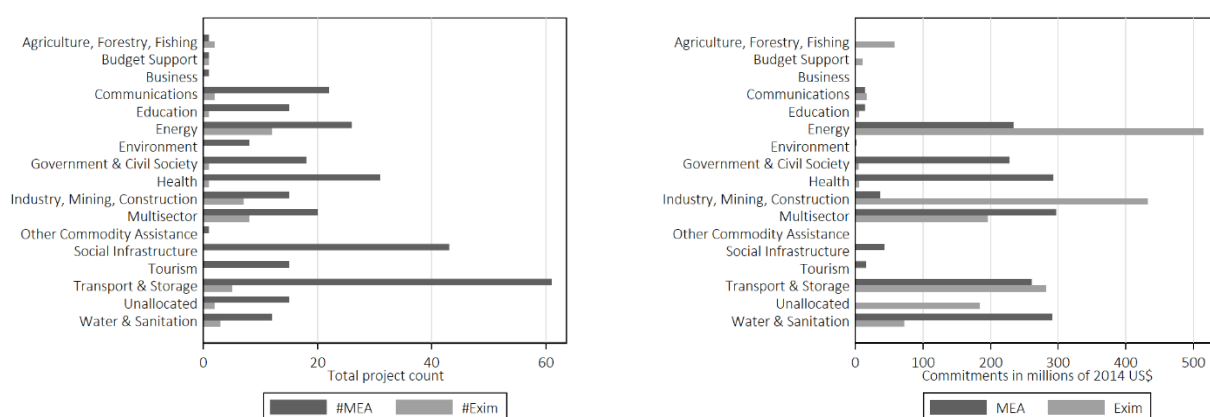

Figure 1: Indian sectoral commitments by MEA and Exim Bank: number of projects and monetary amount, 2007-2014.

About the geocoding process

In a third and final step, we geocoded all of the projects in the dataset. For 50% of the projects, we applied AidData's "double blind" geocoding methodology: Two of our nine research assistants independently extracted relevant geographic information on a project from the documents and searched the web for secondary information on the project locations. Each coder created a separate entry and sets of spatial coordinates for each project. The authors of the research project then reviewed all of these double entries. Whenever the location information provided by the two research assistants differed for a project, the project was re-evaluated in an arbitration process. All research assistants were continually trained based on the insights gained from this arbitration process. Once geocoding quality was assured through these double coding, arbitration, and subsequent training processes, the remaining projects were coded by a single research assistant and quality-assured by the authors. The resulting dataset consists of 1,196 projects across 4,064 intervention sites that were coded with varying levels of geographic precision (see 'Precision Codes' section below). Note that we assigned the projects to the locations where they have been implemented. The data contain 99 cases in which the project has been implemented in India (iso3c = IND).

While it is possible to find exact geographic coordinates for some projects, other projects are more challenging to localize. For instance, the MEA supported the project "[sending] English Teachers in Laos, Philippines, Vietnam" in 2010. Here, it was neither reasonable to expect nor feasible to find any information on subnational locations, so we geocoded the project at the national level. In other cases, the location information that we gathered spans multiple provinces. By way of illustration, the MEA-financed "Terai Road" project connects the Nepalese capital with the southern Terai region near the Indian border. In cases such as this one (see Figure 3), we created line features on the map that follow the projected course of infrastructure projects such as roads and railways to identify all regions that were affected by such infrastructure projects.

The data release includes the following datasets by increasing level of aggregation:

File	Details
ind_aid_global_locations_releaseV1.dta ind_aid_global_locations_releaseV1.xlsx	Full project-location level dataset. Precision codes provide geographic granularity information as described.
ind_aid_global_adm2regions_releaseV1.dta ind_aid_global_adm2regions_releaseV1.xlsx	Second administrative unit aggregates. For project-locations at precision level 3 or lower. Collapsed from the location level by admin 2 region and year.
ind_aid_global_adm1regions_releaseV1.dta ind_aid_global_adm1regions_releaseV1.xlsx	First administrative unit aggregates. For project-locations at precision level 4 or lower. Collapsed from the location level by admin 1 region and year.
ind_aid_global_country_releaseV1.dta ind_aid_global_country_releaseV1.xlsx	Country level aggregates. Collapsed from the location level by country and year.

In addition, we released the project titles and long descriptions as provided by the respective funding agency in the separate data files "ind_aid_global_longdescription.dta" and "ind_aid_global_longdescription.xlsx." This information can be merged to the data sets via AidData's project ID (aiddata_project_id). The Stata file "ind_aid_global_adm1regions_releaseV1.dta" was used by the authors in "Does India Use Development Finance to Compete with China? A Subnational Analysis" (AidData Working Paper #110, September 2021). Note that we use *agencyname* and not the *flowclass* variable to distinguish official Indian flows (see variable definitions below).
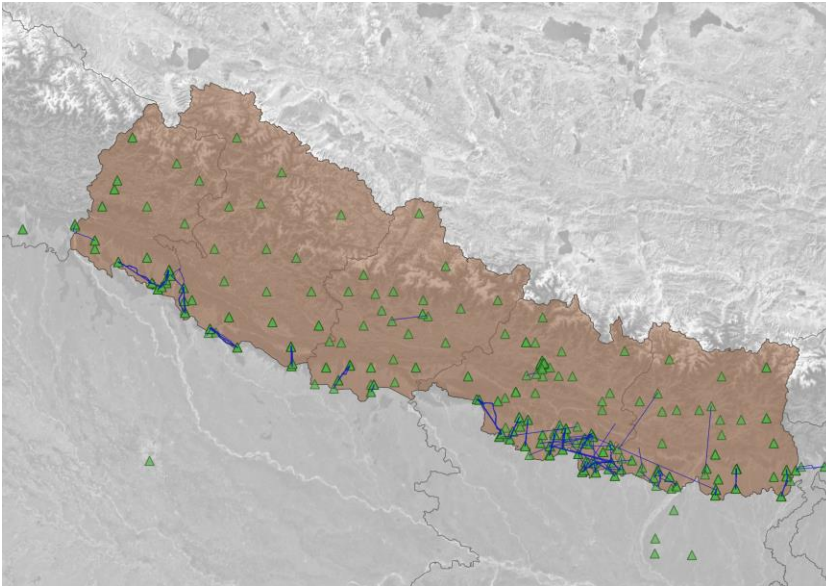
Figure 2: Illustration of the locations of Indian development projects in Nepal, 2007-2014. The blue lines represent line features of Indian development projects, including the "Terai Road" project. The green triangels represent the locations of all Indian development projects in our database. Source: Asmus et al. (2021).

<u>*Precision codes*</u>

We assigned precision codes ranging from 1 to 8 to each project, where project location(s) were identified at the following spatial levels with decreasing precision:

1: an exact location (such as a building or a populated place),
2: a location within a 25 km radius,
3: a second-order administrative (ADM2) region (such as a district or municipality),
4: a first-order administrative (ADM1) region (such as a province),
5: estimated coordinates (e.g. between places, along roads and borders, more than 25 km from a specific location), or large topographical features spanning several administrative boundaries,
6: a location only related to an independent political entity (such as locally dispersed country-wide projects),
8: a location can only be related to an independent political entity where the central government will be only direct beneficiary (such as capacity building, budget support, technical assistance).

This allows aggregating the data to the level of spatial aggregation required for the respective analysis. Note that the coordinates latitude and longitude refer to administrative centroids for precision codes higher than 2, and are precise for precision code 1 only. For the other precision codes, the coordinates are the centeroids of the administrative areas.

<u>*List of variables*</u>

Variable	Definition
project_location_id	Unique project location ID
aiddata_project_id	AidData project ID
iso3c	ISO3 country code of project location
recipientname	Recipient name of the project
year	Reported year of commitment
name_0	Name of ADM0 region (GADM2.8)
name_1	Name of ADM1 region (GADM2.8)
id_0	ID of ADM0 region (GADM2.8)
id_1	ID of ADM1 region (GADM2.8)
latitude	Latitude of project-specific location
longitude	Longitude of project-specific location

geoname_id	Geonames ID of project location (http://www.geonames.org/)
place_name	Name of project location
project_locas	Number of project locations
project_count	Count of projects by country/adm1/adm2 region and year (for aggregated datasets) Dummy for full locations dataset
project_title	Project title as provided by the agency
longdescription	Project long description as provided by the agency
agencyname	Name of Indian government agency
precision_code	Geographic precision code (1-8, see description above)
flow_class	Categories: ODA(-like), OOF(-like)[1]
crs_sector	3-digit sector classification based on OECD sector codes
crs_sector_name	Name of the 3-digit sector classification based on OECD sector
usd_commitment_con	Project location (even-split) commitment amount (constant 2014 USD)
usd_commitment_pt_con	Project total commitment amount (constant 2014 USD)
usd_disbursment_con	Project location (even-split) disbursement amount (constant 2014 USD)
usd_disbursment_pt_con	Project total disbursement amount (constant 2014 USD)

Authors' details

Gerda Asmus	Heidelberg University
	✉ gerda.asmus[at]awi.uni-heidelberg.de
Vera Eichenauer	Swiss Economic Institute KOF, ETH Zurich
	✉ eichenauer[at]kof.ethz.ch
Andreas Fuchs	University of Göttingen, Kiel Institute for the World Economy
	✉ mail[at]andreas-fuchs.net
Brad Parks	AidData, College of William & Mary
	✉ bparks[at]aiddata.wm.edu

---

[1] See the definition on pages 19-20 in Custer, S., Dreher, A., Elston, T.B., Fuchs, A., Ghose, S., Lin, J., Malik, A., Parks, B.C., Russell, B., Solomon, K., Strange, A., Tierney, M.J., Walsh, K., Zaleski, L., and Zhang, S. (2021). Tracking Chinese Development Finance: An Application of AidData's TUFF 2.0 Methodology. Williamsburg, VA: AidData at William & Mary. https://docs.aiddata.org/ad4/pdfs/AidData_TUFF_methodology_2_0.pdf (last accessed January 26, 2022). On the difficulties in coding the flow class variable, see discussion at https://www.aiddata.org/blog/negotiating-the-blurred-lines-between-official-development-assistance-and-other-official-flows (last accessed October 1st, 2021).