

README

Data Package for AidData's *Geocoded Global Chinese Official Finance Dataset (Version 1.0.1)*

September 11, 2018

Official Citation

Bluhm, Richard, Axel Dreher, Andreas Fuchs, Bradley Parks, Austin Strange, and Michael Tierney. 2018. Connective Financing: Chinese Infrastructure Projects and the Diffusion of Economic Activity in Developing Countries. AidData Working Paper #64. Williamsburg, VA: AidData at William & Mary.

AidData Research and Evaluation Unit. 2017. Geocoding Methodology, Version 2.0. Williamsburg, VA: AidData at William & Mary.
<https://www.aiddata.org/publications/geocoding-methodology-version-2-0>

*Please note: Both works count as the official citation for this dataset.

Geocoded Data Coverage

Years: 2000 - 2014

Donors: 1

Total Projects: 3485

Geocoded Projects: 3485

Locations: 6190

Total Disbursements: NA

Total Commitments: \$273,611,118,916.00

Geocoded Disbursements: NA

Geocoded Commitments: \$273,611,118,916.00

Five "Quick Start" Tips

1. Use the file, **oda-like_flows.csv**, to restrict your search to only "aid" projects in the strictest sense of the term.
2. Use the file, **oof-like_flows.csv**, to restrict your search to projects that do *not* qualify as "aid" in the strictest sense of the term.

3. The file **all_flow_classes.csv** contains all Official Finance projects, including those that do qualify as “aid” and those that do not.
4. The column “**even_split_commitments**” contains an estimate of the monetary commitment that is allocable to each location of a project.
5. The column “**project_total_commitments**” contains the commitment amount for the entire project.

What data and files are covered in this README?

This data package includes five files, which are listed in the following table.

- all_flow_classes.csv
- oda-like_flows.csv
- oof-like_flows.csv
- vague_flows.csv
- project_descriptions_and_sources.csv

File Name	Description
all_flow_classes.csv	Includes locations for all projects, regardless of flow_class. (See below for field descriptions.) All fields from AidData’s Global Official Chinese Finance Version 1.0 dataset are included except for the “description”, “source”, and “factiva source” fields.
oda-like_flows.csv	Filtered version of the all_flow_classes.csv according to flow_class “oda-like”, which includes “aid” projects in the strictest sense of the term. For more information, see the “flow_class” definition below.
oof-like_flows.csv	Filtered version of the all_flow_classes.csv according to the flow_class “oof-like”, which includes projects that do not qualify as “aid” in the strictest sense of the term. For more information, see the “flow_class” definition below.
vague_flows.csv	Filtered version of the all_flow_classes.csv according to the flow_class “Vague (Official Finance).” This file includes all projects for which there is insufficient information to make “aid” or “not aid” determinations. For more information, see the

	“flow_class” definition below.
project_descriptions_and_sources.csv	To optimize file download size, project descriptions and sources of project documentation are not included in each of the files above. This file includes these fields and is organized according to projects rather than project locations.

Description

Each file — with the exception of project_descriptions_and_sources.csv — contains a merge of the ‘raw’ data used in AidData Working Paper 64, [Connective Financing: Chinese Infrastructure Projects and the Diffusion of Economic Activity in Developing Countries](#). To access the ‘raw’ data, please visit [\[INSERT LINK TO GITHUB\]](#). Some of this ‘raw’ data is also available on [GeoQuery](#).¹

Each row in the merged files contain a project location. To make it easier for users to distinguish between projects that do and do not meet the strict definition of “aid”, these files provide project location records that have been pre-filtered according to the “flow_class” variable (ODA-like, OOF-like, Vague OF). Descriptions² of these flow classes are also listed in the table below.

The geocoded data in this data package is taken from AidData’s [Global Chinese Official Finance Dataset, Version 1.0](#) (CGOF), released in October 2017. Two filters were applied prior

¹ When aggregating data, GeoQuery performs advanced calculations that attribute portions of less geographically precise aid to more geographically precise levels of aggregation. Because of this, users performing a “simple” filtering operation in a spreadsheet may arrive at a different result than users who input those same filter parameters into GeoQuery. For a full description of GeoQuery’s “boundary rasterization” process, visit the documentation for AidData’s GEO framework: <http://geoquery.org/wp-content/uploads/2017/10/IntroducingtheAidDataGeoFramework.pdf>.

² From the Readme of the source dataset, AidData’s Global Chinese Official Finance: “Chinese Official Finance refers to concessional and non-concessional sources of funding from Chinese government institutions (including central, state or local government institutions) with development, commercial, or representational intent. More specifically, it captures (a) highly concessional, Chinese development projects that meet the OECD’s criteria for ODA; and (b) officially-financed Chinese projects that lack development intent or are provided with higher interest rates and lower grant elements (i.e. projects that fall within the OECD’s criteria for “Other Official Flows”, or OOF.) Chinese ODA represents “Chinese aid” in the strictest sense of the term, but Chinese official finance (ODA and Other Official Flows) is sometimes used as a broader definition of aid. AidData’s dataset allows users to disaggregate Chinese official finance into its constituent parts and determine if they wish to use a narrow or broad definition of aid.”

to geocoding project locations: projects were only geocoded if the “recommended_for_research” field³ was set to “true” and if the “status” field⁴ was set to “completion” or “implementation”.

The main difference between the geocoded data in the merged CSV files listed above and the source dataset (without geocodes) is that each row in the merged CSV files consists of a *project location* instead of a *project*.⁵

When a project had multiple locations, the total commitment amount for that project was evenly divided by the number of project locations. The total amount for the entire project is located in the column “project_total_commitments.” This amount corresponds to the value that the “usd_defl_2014” variable takes in the source dataset. The amount attributed to each project location is listed in the “even_split_commitments” column. To facilitate analysis, non-financial project-level variables⁶ (such as “funding_agency” or “flow_class”) are carried over to each project location.

All data from the Global Chinese Official Finance Dataset (Version 1.0) were collected using [AidData’s Tracking Underreported Financial Flows \(TUFF\) methodology, version 1.3](#). TUFF is a rigorous and replicable methodology that synthesizes and standardizes open-source information in order to create project-level official finance data for opaque donors and lenders.

Since AidData began tracking underreported financial flows in 2013, it has benefited from the generous support of Humanity United, the John D. and Catherine T. MacArthur Foundation, the William and Flora Hewlett Foundation, the Academic Research Fund of Singapore’s Ministry of Education, the United Nations University World Institute for Development Economics Research (UNU-WIDER), and the College of William and Mary.

Field Names and Definitions

³ The “recommended_for_research” field identifies projects that AidData recommends including for most types of research & financial analysis. In the CGOF dataset, it is equal to “TRUE” for all projects except for those that are marked as umbrella projects, pledge projects, cancelled projects, and suspended projects. Projects marked as “FALSE” were excluded from this dataset.

⁴ The “status” field tracks the progression of a project. Labels from the CGOF dataset include: Pipeline: Pledge, Pipeline: Commitment, Implementation, Completion, Suspended, Cancelled. Projects that were not marked “Implementation” or “Completion” were excluded from this dataset.

⁵ Each row in the project_descriptions_and_sources.csv file still contains project-level data. This data includes the project descriptions and the sources used to create the dataset, AidData’s Global Chinese Official Finance, Version 1.0.

⁶ Full project descriptions and sources used to compile the original project record in the Global Chinese Official Finance dataset were omitted from the pre-merged project location files to economize file size.

Field Name	Definition
ad_purpose_codes	<p>AidData has developed a granular system of sector coding, which expands the OECD's purpose code scheme. For more information, please visit: https://www.aiddata.org/pages/data-user-guide</p> <p>Note: This dataset is not purpose coded, so this column is blank.</p>
ad_purpose_names	<p>AidData has developed a granular system of sector coding, which expands the OECD's purpose code scheme. For more information, please visit: https://www.aiddata.org/pages/data-user-guide</p> <p>Note: This dataset is not purpose coded, so this column is blank.</p>
all_recipients	<p>From the CGOF readme: "The recipient entity or entities receiving official finance from a donor entity. A project record can have multiple recipients in cases where a donor entity is providing financing, goods or services to more than one country through a single project. This field provides a pipe-delimited list of all recipients specified as receiving official finance from a donor entity."</p>
amount	<p>From the CGOF readme: "Monetary amount pledged or committed by the donor entity for the completion of a project."</p> <p>Note: See "disambiguation" section below. In the CGOF, "amount" refers to commitments in varying currencies. For the total amount committed to an entire project deflated to 2014 USD, please refer to the field "project_total_commitments". For an estimation of amounts committed to a project location, please refer to the field "even_split_commitments".</p>
cofinancing_agency	<p>Co-financing agencies or other agencies involved in the project</p>

contacts	From the CGOF readme: “Name and position of people who were involved in the project if available.”
currency	<p>From the CGOF readme: “Currency associated with the monetary amount for a project.”</p> <p>Note: “Currency” only refers to the value in the “amount” column. It does not refer to monetary amounts listed in other columns, which denominated in USD.</p>
crs_sector_code	From the CGOF readme: “3-digit sector classification based on OECD sector codes.”
crs_sector_name	Name of the 3-digit sector classification based on OECD sector codes sources Links to all the sources attached to one project.
deflators_used	<p>From the CGOF readme: “Deflator value calculated based on AidData methodology (which mimics OECD deflation methodology) using World Bank sources. Deflator is calculated using donor country inflation and change in buying power.”</p> <p>This conversion was applied to the value in the “usd_current” column in order to calculate the value listed in the “usd_defl_2014” column, which provides the source for the “project_total_commitments” in this data package.</p>
description	<p>1-2 paragraph comprehensive summary of the activities supported by the project. For a detailed explanation of how this variable is recorded, please see page 15 of the TUFF methodology found here: https://www.aiddata.org/methods/tracking-underreported-financial-flows</p> <p>Note: In this data package, the “description” field only appears in the file “description_and_sources.csv”. In order optimize the overall size of this data package, this field was not merged with project locations.</p>
end_actual	From the CGOF readme: “The actual date that a project was completed on.”

end_actual_isodate	Standardized conversion of the value in the “end_actual” column to the isodate format.
end_planned	The announced completion date for a project.
even_split_commitments	This estimate of the amount of the project that was allocated to each location of the project is taken by dividing the value in the “project_total_commitments” column by the number of locations that correspond to a particular project. (i.e. the number of rows with a project_location_id that includes the project_id.)
even_split_disbursements	Total disbursements divided by number of project locations. Since this dataset only tracks commitments, this column is also blank.
exchange_rates_used	<p>From the CGOF readme: “Exchange rate used to change the specified currency of the original amount into USD.”</p> <p>This rate was applied to the value in the “amount” column to arrive at the value in the “usd_current” column.</p>
field_completeness	<p>From the CGOF Readme: “Rates projects based on the number of missing data fields critical to understanding the project. It also penalized projects with “Vague” makers like “Vague Official Finance” for flow class or “Vague TBD” for flow type. The field completeness measure varies from 0 to 9; higher values indicate that a higher percentage of the “fields” (i.e. variables such as transaction amount, flow type, and commitment year) for a given project record are complete. For more information, see http://china.aiddata.org/content/improving_the_data.”</p>
flow	Details on how financial flows, goods or services are transferred from the donor to recipient for a project. Flow types include: Debt Forgiveness, Debt Rescheduling, Export Credits, Foreign Direct Investment, Freestanding Technical Assistance, Grant, Loan, Joint Venture with Recipient, Scholarships/Training in Donor Country, and Strategic/Supplier Credit.
flow_class	TUFF Coders are also instructed to assign all projects to one of the following flow class categories: “ODA-like”, “OOF-like”, “Vague (Official Finance)”. More details on these

	<p>flow categorizations can be found in TUFF Methodology 1.3 document located here:</p> <p>If one's goal is to capture Chinese Official Development Assistance (ODA) or aid in the strictest sense of the term, include only projects that are classified as ODA-like. However, if one's goal is to capture the universe of Chinese Government-financed projects provided on concessional and non-concessional terms, include projects that are classified as ODA-like, OOF-like and Vague (Official Finance).</p>
funding_agency	Donor agency that provided the funding.
gazetteer_adm_code	<p>An array of unique identifiers that represents the hierarchy of a GeoNames place taken from various sources in the GeoNames Gazetteer. For more information on GeoNames data source, visit:</p> <p>https://www.geonames.org/data-sources.html</p>
gazetteer_adm_name	<p>An array of names associated with the gazetteer_adm_code that represents the hierarchy of a place. For more information on GeoNames data sources, visit:</p> <p>https://www.geonames.org/data-sources.html</p>
geographic_exactness	<p>A two-point, ordinal scale taken from the International Aid Transparency Initiative (IATI) standard that captures how precisely a record has been geocoded, where 1 is "exact" and 2 is "approximate". For more information, visit:</p> <p>http://reference.iatistandard.org/202/codelists/GeographicExactness/</p>
geoname_id	<p>The unique identifier of a place taken from the "GeoNames" catalogue. For more information about GeoNames, visit</p> <p>http://www.geonames.org/</p>
grace_period	Interval to first repayment of capital in years.
grant_element	<p>From the CGOF readme: "Grant element of the loan, in percent. Uses grant element reported by project sources, unless it can be calculated independently with the OECD grant element calculator (assumes 10 percent rate of discount, 2 payments per annum, and equity principal payment)."</p>

implementing_agency	Agency that implemented the project. Could be from the donor, recipient or other parties.
intent	From the CGOF Readme: “Perceived intent of the finance provider. Categories covered by “donor intent” include: Development, Commercial, Representational, Mixed (encompassing any combination of the other three intents).”
interest_rate	Reported interest rate of the project loan, in percent.
is_cofinanced	From the CGOF readme: “Denotes projects with multiple countries acting as donor (including the donor of investigation), where the financial amount provided by each donor separately could not distinguished.”
is_ground_truthing	From the CGOF readme: “A binary marker that indicates when a project has been verified with in-country field research.”
latitude	The assigned latitudinal coordinate assigned to a project coordinate.
line_of_credit	Denotes a project or financial arrangement where the donor extended a line of credit to a recipient entity. This credit may or may not be used in its entirety by the recipient.
loan_type	From the CGOF readme: “Represents the general nature of the loan. Categories covered include: interest free; concessional; non-concessional; no information and some information.”
location_class	An index of location types used by the International Aid Transparency Initiative (IATI) standard. For more information, visit: http://reference.iatistandard.org/202/codelists/GeographicLocationClass/
location_details	The subnational locations where projects occurred.
location_type_code	United States National Geospatial-Intelligence Agency (US NGA) Feature Designation Codes. For more information visit: http://reference.iatistandard.org/202/codelists/LocationType/ http://geonames.nga.mil/namesgaz/

location_type_name	The name associated with a location_type_code. For more information on location types, visit: http://reference.iatistandard.org/202/codelists/LocationType/ http://geonames.nga.mil/namesgaz/
longitude	The assigned latitudinal coordinate assigned to a project coordinate.
maturity	Reported duration of the project loan, in years.
place_name	The name of a place taken from the in the “GeoNames” catalogue. For more information about GeoNames, visit http://www.geonames.org/
precision_code	An eight-point, ordinal scale that captures how precisely a record has been geocoded, where 1 is the highest level of spatial precision and 8 is the lowest level of spatial precision. For a complete description of AidData’s geocoding methodology, visit the geocoding methodology documents found at: https://www.aiddata.org/methods/geocoding
project_id	Unique identification number assigned to every project in the dataset. Within the dynamic, online version of the dataset (at china.aiddata.org), projects can be accessed with their unique identification numbers by using the “Access project by ID” function.
project_location_id	The unique identification number assigned to every project in the raw dataset. The component of the project_location_id that is located before the “_” corresponds to the project with which a location is associated.
project_total_commitments	This amount represents the total amount committed to the project, deflated to 2014 USD. This amount is taken from the “usd_defl_2014” column in AidData’s Global Chinese Official Finance Dataset. (Related columns to this deflation are also included in this dataset for reference.) In the raw data file, level_1a.csv, this field name is referred to as “total_commitments”.
project_total_disbursements	This amount represents the total amount disbursed to the project, deflated to 2014 USD. This dataset only tracks commitments, so this column is blank.

project_title	Title of project.
recipient_agencies	From the CGOF readme: “Recipient agencies that are involved in the project.”
recipient_agencies_count	Number of recipient agencies that are involved in the project.
recipient_condensed	From the CGOF readme: “Recipient country. If multiple countries were identified as recipients, this field identifies the smallest regional grouping that all specified countries belong to.”
recipient_count	From the CGOF readme: “Number of recipient <i>countries</i> identified” (emphasis added).
recipient_cow_code	<p>From the CGOF readme: “Code of recipient country according to COW standard”.</p> <p>For more information on the <i>covariates of war</i> (cow) standard, visit: http://www.correlatesofwar.org/data-sets/cow-country-codes</p>
recipient_iso3	Code of recipient country according to ISO3 standard
recipient_iso2	Code of recipient country according to ISO2 standard
recipient_oecd_code	Code of recipient country according to OECD standard
recipient_oecd_name	Name of recipient country according to OECD standard
recipient_un_code	Code of recipient country according to UN standard
recipient_imf_code	Code of recipient country according to IMF standard
recipient_region	From the CGOF Readme: “Region of the recipient country.”
round_coded	This dataset includes locations that were geocoded in three rounds: (1) Chinese Government-financed projects in Africa , which were coded in 2013; (2) Chinese Government-financed projects in three ecologically sensitive regions , which were coded in 2016; and (3) all remaining projects (at the implementation or completion stage) from AidData's Global Chinese Official Finance Dataset, Version 1.0 , which were coded in 2017. The first

	<p>dataset was coded using the 1.1. version of AidData's geocoding methodology. Precision scores were assigned by geocoders, indicating how close to a particular set of latitude and longitude coordinates a project occurred. The second and third datasets were coded using the 2.0 version of AidData's geocoding methodology. In this methodology, AidData adopted the International Aid Transparency Initiative (IATI) standard for describing the location class and geographic exactness of a given location. For more information on AidData's geocoding methodology, along with a detailed crosswalk of AidData precision codes to the IATI standard, please refer to documentation found here: https://www.aiddata.org/methods/geocoding</p>
source_triangulation	<p>From the CGOF readme: "Rates projects based on the types of sources (media, academic, donor/recipient government, other official, NGO). The source triangulation measure varies from 0 to 19, with higher values representing a project record that draws upon a diverse set of information sources. For more information, see http://china.aiddata.org/content/improving_the_data."</p>
sources_count	<p>From the CGOF readme: "Number of sources attached to one project."</p>
start_actual	<p>From the CGOF readme: "The actual date that the implementation of a project began on."</p>
start_actual_isodate	<p>Standardized conversion of the value in the "start_actual" column to the isodate format.</p>
start_planned	<p>From the CGOF readme: "The announced start date for a project."</p>
status	<p>Tracks the progression of a project. Labels from the CGOF dataset include: Pipeline: Pledge, Pipeline: Commitment, Implementation, Completion, Suspended, Cancelled.</p> <p>Note: For this dataset, projects were only geocoded if this field contained a value of "Implementation" or "Completion".</p>
title	<p>Short phrase describing the nature of a given project.</p>

transactions_end_year	<p>This refers to whichever is later: a) the listed end date of a project, or b) the last transaction recorded for a project. This information is drawn source information for a project and is used in GeoQuery to enable users to filter project records by date.</p> <p>Note: “transactions_end_year” replicates the “year” column in the CGOF dataset.</p>
transactions_start_year	<p>This refers to whichever is earlier: a) the listed start date of a project, or b) the first transaction recorded for a project. This information is drawn source information for a project and is used in GeoQuery to enable users to filter project records by date.</p> <p>Note: “transactions_start_year” replicates the “year” column in the CGOF dataset.</p>
umbrella	<p>This field identifies projects where some agreement was signed between two countries but the funds were not allocated at the original agreement date, only to be allocated through smaller projects later on. Given the nature of TUFF data collection, it is likely that the smaller projects that relate to these umbrella projects were captured elsewhere in the dataset. To avoid double counting, we suggest users do not include records that are marked as umbrella projects in financial analysis. All general lines of credit are also marked as umbrella projects.</p>
usd_current	<p>From the CGOF readme: “Nominal USD value of the amount specified for the project.”</p> <p>Note: The “current” year refers to the the commitment amount in the year that the project was committed. This is listed in the “year” column.</p>
usd_defl_2014	<p>From the CGOF readme: “Deflated monetary equivalent of reported monetary amount in reported currency to 2014 U.S. dollars.”</p> <p>Note: This column is replicated in the column, “project_total_commitments.”</p>

year	The year in which an agreement was reportedly made between a supplier of development finance and a recipient for a project. If available, the agreement year is the year of the formal signed commitment for a project; if commitment year is unavailable, or if a commitment has not been made, the year of the informal pledge is the agreement year for a project.
year_uncertain	From the CGOF readme: ‘Marker for projects without any sources reporting a specific pledge or commitment year. In these cases, the year of the earliest media report serves is coded as “agreement year.”’

Disambiguation of Key Terms

“Donor”

For a list of Chinese agencies that provided official finance, please refer to the field **“funding_agency”**. The merged CSV files in this data package do not contain a column labeled “donor”.

In AidData’s [Global Chinese Official Finance Dataset, Version 1.0](#) (CGOF), released in October 2017, “donor” refers to “[the] entity providing assistance to recipient country. This can be a sovereign state or multilateral organization providing assistance to a recipient country. For this [CGOF] dataset, the donor will always be China”.

The CGOF dataset provided a field labeled “funding_agency” which listed specific agencies that provided official finance. In the ‘raw’ data used in Working Paper 64, the “donor” column in level_1a.csv and projects.csv refers to these agencies listed in the CGOF “funding_agency” column.

Monetary Amounts

Monetary amounts for each *project* and for each *project location* are each denominated in 2014 USD. The amount per project is listed in the **“project_total_commitments”** column, and the amount per project *location* is listed in the **“even_split_commitments”** column. To calculate the amount per project location, the total amount for the project was divided evenly by the number of project locations associated with that project.

- $\text{“project_total_amount”} / (\text{number of project locations}) = \text{“even_split_commitments”}$

In the raw data used in AidData Working Paper 64, this “project_total_column” column is listed as “total_commitments” in the file, “level_1a.csv.” For clarity, this column was renamed in the merged files in this data package. This column was sourced from the “usd_defl_2014” column in AidData’s [Global Chinese Official Finance Dataset, Version 1.0](#) (CGOF).

Also included are columns from the CGOF that were used to calculate the “usd_defl_2014” column. These columns are located in the project_ancillary.csv file in the ‘raw’ data used in AidData Working Paper 64. The value for “usd_defl_2014” was calculated by converting the nominal amount into USD and then deflating to a consistent year, 2014.

- “amount” / “exchange_rate_used” = “usd_current”
- “usd_current” / “deflators_used” = “usd_defl_2014”

For more information about exchange rates and deflators used in the CGOF dataset, visit: <http://aiddata.org/data/chinese-global-official-finance-dataset>

As noted elsewhere in this document, this dataset only tracks commitments and not disbursements, as defined by the OECD. For more information, visit: <http://www.oecd.org/dac/financing-sustainable-development/development-finance-data/dac-glossary.htm>

Dates and Years

This data package draws from the ‘raw’ data used in AidData Working Paper 64, which itself draws from AidData’s [Global Chinese Official Finance Dataset, 2000-2014, Version 1.0](#) (CGOF). Because AidData’s geocoded research releases generally include fields to handle both commitment and disbursements, there are additional naming conventions. These columns are included in the merged files in this data package for reference.

- The columns “**start_actual_isodate**” and “**end_actual_isodate**” replicate to the “start_actual” and “end_actual” columns in the CGOF.
- “**transaction_start_year**” and “**transaction_end_year**” replicate the “**year**” column in the CGOF. Because this geocoded dataset is comprised of only commitments (and not disbursements) the starting and ending transaction years are the same. In the ‘raw’ data used in Working Paper 64, these are listed in the “transaction_year” column in transactions.csv.

Columns With Uniform Values Not Included in the Merged CSV Files

recommended_for_research: true

is_geocoded: true

start_type_actual: start actual

end_type_actual: end actual

As noted above, projects were only geocoded if the “recommended_for_research” field in AidData’s [Global Chinese Official Finance Dataset, Version 1.0](#) (CGOF) was set to “true”. Similarly, all of the records in this data package are geocoded, so “is_geocoded” is always set to “true”.

The “recommended_for_research” field identifies projects that AidData recommends including for most types of research & financial analysis. In the CGOF dataset, it is equal to “TRUE” for all projects except for those that are marked as umbrella projects, pledge projects, cancelled projects, and suspended projects. Projects marked as “FALSE” were excluded from this dataset.

The fields “start_type_actual” and “end_type_actual” refer to a IATI standard field, Activity Date Type. For this dataset, projects were always geocoded based on “actual” start and end dates. “Actual start” equals the IATI activity type code, “2”. “Actual end” equals the IATI activity type code “4”. For more information about the International Aid Transparency Initiative (IATI) standard activity date type field, visit:

<http://reference.iatistandard.org/203/codelists/ActivityDateType/>

For more information about how AidData’s geocoding methodology crosswalks with the IATI standard, please visit the documentation found at: <https://www.aiddata.org/methods/geocoding>

Guides for using our geocoded data in GIS and in a Spreadsheet

For further information on using data in GIS and spreadsheets, reference:

https://github.com/AidData-WM/public_datasets/blob/master/documentation/using_geocoded_data.md

Tracking Underreported Financial Flow Sources and Methodology

This dataset draws from AidData’s Global Chinese Official Finance (Version 1.0) dataset. The following information in this section is taken from the readme for that dataset.

AidData uses the Tracking Underreported Financial Flows (TUFF) methodology to collect detailed financial, operational, and locational information about China’s overseas investments. The TUFF methodology involves a systematic, transparent, and replicable process of triangulating information from four primary sources: (1) English, Chinese, and local-language news reports; (2) Chinese ministries, embassies, and economic and commercial counselor offices; (3) the aid and debt information management systems of finance and planning ministries

in counterpart countries; and (4) case studies and field research undertaken by scholars and NGOs. A dedicated research team synthesizes and standardizes the information from these various sources into discrete project records.

In total, AidData drew upon more than 15,000 distinct information sources to assemble its Global Chinese Official Finance (Version 1.0) dataset. On average, each project record is informed and confirmed by three independent sources. Human data collectors were also assisted by a machine learning algorithm (called the TUFF Robot) to help automate the early stages of the data collection process.

China has formally opted out of international reporting systems, including the OECD's Creditor Reporting System and the International Aid Transparency Initiative (IATI). Beijing has published aggregated numbers on foreign aid in the State Council White papers of 2011 and 2014, but it is unclear how they define and measure foreign aid. China does not publish a bilateral, country-by-country breakdown of its international development finance activities. Nor does it publish project-level data or detailed information about the less concessional and more commercially oriented financing that it provides. AidData provides details on all three.

Please see the [AidDataTUFF_methodology_1.3.pdf](#) document for a detailed description of the data collection and quality control procedures that AidData use to collect data on Chinese Official Finance.

Data Use Recommendations

Users should pay extra attention to debt forgiveness/rescheduling projects. We have included debt forgiveness/rescheduling projects in the 'recommended_for_research' field, but there may be some overlap if the original loan is captured elsewhere in the dataset. Users may choose to exclude these projects if they so desire, as there is insufficient information to identify whether forgiven debts are already captured by other loan project rows.

Familiarize yourself with the key fields in the dataset and the various categorizations they offer. These fields include: donor, intent, flow, flow class, status, umbrella, sector, and recommended for research.

Glossary

Chinese Official Finance: Concessional and non-concessional funding from Chinese government institutions (including central, state or local government institutions) with development, commercial, or representational intent.

Official Development Assistance (ODA): A project that is primarily aimed at development, has a concessionality level (e.g. grant element) of at least 25%, and the recipient is ODA eligible (as

defined by the OECD, based on country income level). This definition is aligned with the OECD Development Assistance Committee Directives. For a project to be categorized as ODA in our database, the intent field must be set to “Development”, and the flow type field must be a grant, grant-like (e.g. scholarships, technical cooperation), or a concessional loan. In our dataset, these flows are categorized as “ODA-like”.

Other Official Flows (OOF): A project that does not meet the criteria for ODA, because it is not primarily focused on development, it is insufficiently concessional (e.g. grant element less than 25%), or is being provided to a country that is not eligible for ODA. This definition is aligned with the OECD Development Assistance Committee 2014-2015 Directives. Projects in this category include those where the intent is primarily commercial or representational, and/or projects that are not concessional enough to qualify as ODA. This category includes export credits. In our dataset, these flows are categorized as “OOF-like”.

Vague Official Finance: This flow class captures officially financed projects where there was not enough information about the project’s concessional level or intent to make a clear ODA or OOF determination. The OECD-DAC does not require this type of residual category because its members systematically report the intent and concessional level of their officially-financed projects.

Health of Record Score: An assigned value that rates the completeness and verifiability of each record in the database. The score is used internally to prioritize subsequent rounds of research and quality checks, and can also be used by external users to identify the relative reliability of each record. It is comprised of two individual scores: (1) The Source Triangulation Score captures our level of confidence in the information in a project record based on the type and number of resources used to create the project record, and (2) The Data Field Completeness Score indicates the level of completeness for a given record, based on the availability of information on high-value fields. To get to a Health of Record Score for each project record, we add the Source Triangulation Score and the Data Field Completeness Score together. We consider a score between 0-8 low, and target those projects during our quality assurance procedures to bolster the record. A score of 8-15 is considered good, and a score above 15 is considered very good.

Licensing Agreement

DATA LICENSE AND ATTRIBUTION

ODC Attribution License (ODC-By)

Preamble

The Open Data Commons Attribution License is a license agreement intended to allow users to freely share, modify, and use this Database subject only to the attribution requirements set out in Section 4. Databases can contain a wide variety of types of content (images, audiovisual material, and sounds all in the same database, for example), and so this license only governs the rights over the Database, and not the contents of the Database individually. Licensors may therefore wish to use this license together with another license for the contents. Sometimes the contents of a database, or the database itself, can be covered by other rights not addressed here (such as private contracts, trademark over the name, or privacy rights / data protection rights over information in the contents), and so you are advised that you may have to consult other documents or clear other rights before doing activities not covered by this License.

----- The Licensor (as defined below) and You (as defined below) agree as follows:

1.0 Definitions of Capitalised Words

"Collective Database" – Means this Database in unmodified form as part of a collection of independent databases in themselves that together are assembled into a collective whole. A work that constitutes a Collective Database will not be considered a Derivative Database.

"Convey" – As a verb, means Using the Database, a Derivative Database, or the Database as part of a Collective Database in any way that enables a Person to make or receive copies of the Database or a Derivative Database. Conveying does not include interaction with a user through a computer network, or creating and Using a Produced Work, where no transfer of a copy of the Database or a Derivative Database occurs.

"Contents" – The contents of this Database, which includes the information, independent works, or other material collected into the Database. For example, the contents of the Database could be factual data or works such as images, audiovisual material, text, or sounds.

"Database" – A collection of material (the Contents) arranged in a systematic or methodical way and individually accessible by electronic or other means offered under the terms of this License.

"Database Directive" – Means Directive 96/9/EC of the European Parliament and of the Council of 11 March 1996 on the legal protection of databases, as amended or succeeded.

"Database Right" – Means rights resulting from the Chapter III ("sui generis") rights in the Database Directive (as amended and as transposed by member states), which includes the Extraction and Re-utilisation of the whole or a Substantial part of the Contents, as well as any similar rights available in the relevant jurisdiction under Section 10.4.

"Derivative Database" – Means a database based upon the Database, and includes any translation, adaptation, arrangement, modification, or any other alteration of the Database or of a Substantial part of the Contents. This includes, but is not limited to, Extracting or Re-utilising the whole or a Substantial part of the Contents in a new Database.

"Extraction" – Means the permanent or temporary transfer of all or a Substantial part of the Contents to another medium by any means or in any form.

"License" – Means this license agreement and is both a license of rights such as copyright and Database Rights and an agreement in contract.

"Licensor" – Means the Person that offers the Database under the terms of this License.

"Person" – Means a natural or legal person or a body of persons corporate or incorporate.

"Produced Work" – a work (such as an image, audiovisual material, text, or sounds) resulting from using the whole or a Substantial part of the Contents (via a search or other query) from this Database, a Derivative Database, or this Database as part of a Collective Database.

"Publicly" – means to Persons other than You or under Your control by either more than 50 percent ownership or by the power to direct their activities (such as contracting with an independent consultant).

"Re-utilisation" – means any form of making available to the public all or a Substantial part of the Contents by the distribution of copies, by renting, by online or other forms of transmission.

"Substantial" – Means substantial in terms of quantity or quality or a combination of both. The repeated and systematic Extraction or Re-utilisation of insubstantial parts of the Contents may amount to the Extraction or Re-utilisation of a Substantial part of the Contents.

"Use" – As a verb, means doing any act that is restricted by copyright or Database Rights whether in the original medium or any other; and includes without limitation distributing, copying, publicly performing, publicly displaying, and preparing derivative works of the Database, as well as modifying the Database as may be technically necessary to use it in a different mode or format.

"You" – Means a Person exercising rights under this License who has not previously violated the terms of this License with respect to the Database, or who has received express permission from the Licensor to exercise rights under this License despite a previous violation. Words in the singular include the plural and vice versa.

2.0 What this License covers

2.1. Legal effect of this document.

This License is: a. A license of applicable copyright and neighbouring rights; b. A license of the Database Right; and c. An agreement in contract between You and the Licensor.

2.2 Legal rights covered. This License covers the legal rights in the Database, including:

a. Copyright. Any copyright or neighbouring rights in the Database. The copyright licensed includes any individual elements of the Database, but does not cover the copyright over the Contents independent of this Database. See Section 2.4 for details. Copyright law varies between jurisdictions, but is likely to cover: the Database model or schema, which is the structure, arrangement, and organisation of the Database, and can also include the Database tables and table indexes; the data entry and output sheets; and the Field names of Contents stored in the Database;

b. Database Rights. Database Rights only extend to the Extraction and Re-utilisation of the whole or a Substantial part of the Contents. Database Rights can apply even when there is no copyright over the Database. Database Rights can also apply when the Contents are removed from the Database and are selected and arranged in a way that would not infringe any applicable copyright; and

c. Contract. This is an agreement between You and the Licensor for access to the Database. In return you agree to certain conditions of use on this access as outlined in this License.

2.3 Rights not covered.

a. This License does not apply to computer programs used in the making or operation of the Database;

b. This License does not cover any patents over the Contents or the Database; and

c. This License does not cover any trademarks associated with the Database.

2.4 Relationship to Contents in the Database. The individual items of the Contents contained in this Database may be covered by other rights, including copyright, patent, data protection, privacy, or personality rights, and this License does not cover any rights (other than Database Rights or in contract) in individual Contents contained in the Database. For example, if used on a Database of images (the Contents), this License would not apply to copyright over individual images, which could have their own separate licenses, or one single license covering all of the rights over the images.

3.0 Rights granted.

3.1 Subject to the terms and conditions of this License, the Licensor grants to You a worldwide, royalty-free, non-exclusive, terminable (but only under Section 9) license to Use the Database for the duration of any applicable copyright and Database Rights. These rights explicitly include commercial use, and do not exclude any field of endeavour. To the extent possible in the relevant jurisdiction, these rights may be exercised in all media and formats whether now known or created in the future.

The rights granted cover, for example:

- a. Extraction and Re-utilisation of the whole or a Substantial part of the Contents;
- b. Creation of Derivative Databases;
- c. Creation of Collective Databases;
- d. Creation of temporary or permanent reproductions by any means and in any form, in whole or in part, including of any Derivative Databases or as a part of Collective Databases; and
- e. Distribution, communication, display, lending, making available, or performance to the public by any means and in any form, in whole or in part, including of any Derivative Database or as a part of Collective Databases.

3.2 Compulsory license schemes. For the avoidance of doubt:

- a. Non-waivable compulsory license schemes. In those jurisdictions in which the right to collect royalties through any statutory or compulsory licensing scheme cannot be waived, the Licensor reserves the exclusive right to collect such royalties for any exercise by You of the rights granted under this License;
- b. Waivable compulsory license schemes. In those jurisdictions in which the right to collect royalties through any statutory or compulsory licensing scheme can be waived, the Licensor waives the exclusive right to collect such royalties for any exercise by You of the rights granted under this License; and,
- c. Voluntary license schemes. The Licensor waives the right to collect royalties, whether individually or, in the event that the Licensor is a member of a collecting society that administers voluntary licensing schemes, via that society, from any exercise by You of the rights granted under this License.

3.3 The right to release the Database under different terms, or to stop distributing or making available the Database, is reserved. Note that this Database may be multiple-licensed, and so You may have the choice of using alternative licenses for this Database. Subject to Section 10.4, all other rights not expressly granted by Licensor are reserved.

4.0 Conditions of Use

4.1 The rights granted in Section 3 above are expressly made subject to Your complying with the following conditions of use. These are important conditions of this License, and if You fail to follow them, You will be in material breach of its terms.

4.2 Notices. If You Publicly Convey this Database, any Derivative Database, or the Database as part of a Collective Database, then You must:

- a. Do so only under the terms of this License;

b. Include a copy of this License or its Uniform Resource Identifier (URI) with the Database or Derivative Database, including both in the Database or Derivative Database and in any relevant documentation;

c. Keep intact any copyright or Database Right notices and notices that refer to this License; and

d. If it is not possible to put the required notices in a particular file due to its structure, then You must include the notices in a location (such as a relevant directory) where users would be likely to look for it.

4.3 Notice for using output (Contents). Creating and Using a Produced Work does not require the notice in Section 4.2. However, if you Publicly Use a Produced Work, You must include a notice associated with the Produced Work reasonably calculated to make any Person that uses, views, accesses, interacts with, or is otherwise exposed to the Produced Work aware that Content was obtained from the Database, Derivative Database, or the Database as part of a Collective Database, and that it is available under this License.

Please use this notice:

Bluhm, Richard, Axel Dreher, Andreas Fuchs, Bradley Parks, Austin Strange, and Michael Tierney. 2018. Connective Financing: Chinese Infrastructure Projects and the Diffusion of Economic Activity in Developing Countries. AidData Working Paper #64. Williamsburg, VA: AidData at William & Mary.

AidData Research and Evaluation Unit. 2017. Geocoding Methodology, Version 2.0. Williamsburg, VA: AidData at William & Mary.

<https://www.aiddata.org/publications/geocoding-methodology-version-2-0>

*Please note: Both works count as the official citation for this dataset.

4.4 Licensing of others. You may not sublicense the Database. Each time You communicate the Database, the whole or Substantial part of the Contents, or any Derivative Database to anyone else in any way, the Licensor offers to the recipient a license to the Database on the same terms and conditions as this License. You are not responsible for enforcing compliance by third parties with this License, but You may enforce any rights that You have over a Derivative Database. You are solely responsible for any modifications of a Derivative Database made by You or another Person at Your direction. You may not impose any further restrictions on the exercise of the rights granted or affirmed under this License.

5.0 Moral rights.

5.1 This section covers moral rights, including any rights to be identified as the author of the Database or to object to treatment that would otherwise prejudice the author's honour and reputation, or any other derogatory treatment:

a. For jurisdictions allowing waiver of moral rights, Licensor waives all moral rights that Licensor may have in the Database to the fullest extent possible by the law of the relevant jurisdiction under Section 10.4;

b. If waiver of moral rights under Section 5.1 a in the relevant jurisdiction is not possible, Licensor agrees not to assert any moral rights over the Database and waives all claims in moral rights to the fullest extent possible by the law of the relevant jurisdiction under Section 10.4; and

c. For jurisdictions not allowing waiver or an agreement not to assert moral rights under Section 5.1 a and b, the author may retain their moral rights over certain aspects of the Database. Please note that some jurisdictions do not allow for the waiver of moral rights, and so moral rights may still subsist over the Database in some jurisdictions.

6.0 Fair dealing, Database exceptions, and other rights not affected.

6.1 This License does not affect any rights that You or anyone else may independently have under any applicable law to make any use of this Database, including without limitation:

a. Exceptions to the Database Right including: Extraction of Contents from non-electronic Databases for private purposes, Extraction for purposes of illustration for teaching or scientific research, and Extraction or Re-utilisation for public security or an administrative or judicial procedure. b. Fair dealing, fair use, or any other legally recognised limitation or exception to infringement of copyright or other applicable laws.

6.2 This License does not affect any rights of lawful users to Extract and Re-utilise insubstantial parts of the Contents, evaluated quantitatively or qualitatively, for any purposes whatsoever, including creating a Derivative Database (subject to other rights over the Contents, see Section 2.4). The repeated and systematic Extraction or Re-utilisation of insubstantial parts of the Contents may however amount to the Extraction or Re-utilisation of a Substantial part of the Contents.

7.0 Warranties and Disclaimer

7.1 The Database is licensed by the Licensor "as is" and without any warranty of any kind, either express, implied, or arising by statute, custom, course of dealing, or trade usage. Licensor specifically disclaims any and all implied warranties or conditions of title, non-infringement, accuracy or completeness, the presence or absence of errors, fitness for a particular purpose, merchantability, or otherwise. Some jurisdictions do not allow the exclusion of implied warranties, so this exclusion may not apply to You.

8.0 Limitation of liability

8.1 Subject to any liability that may not be excluded or limited by law, the Licensors are not liable for, and expressly excludes, all liability for loss or damage however and whenever caused to anyone by any use under this License, whether by You or by anyone else, and whether caused by any fault on the part of the Licensors or not. This exclusion of liability includes, but is not limited to, any special, incidental, consequential, punitive, or exemplary damages such as loss of revenue, data, anticipated profits, and lost business. This exclusion applies even if the Licensors have been advised of the possibility of such damages.

8.2 If liability may not be excluded by law, it is limited to actual and direct financial loss to the extent it is caused by proved negligence on the part of the Licensors.

9.0 Termination of Your rights under this License

9.1 Any breach by You of the terms and conditions of this License automatically terminates this License with immediate effect and without notice to You. For the avoidance of doubt, Persons who have received the Database, the whole or a Substantial part of the Contents, Derivative Databases, or the Database as part of a Collective Database from You under this License will not have their licenses terminated provided their use is in full compliance with this License or a license granted under Section 4.8 of this License. Sections 1, 2, 7, 8, 9 and 10 will survive any termination of this License.

9.2 If You are not in breach of the terms of this License, the Licensors will not terminate Your rights under it.

9.3 Unless terminated under Section 9.1, this License is granted to You for the duration of applicable rights in the Database.

9.4 Reinstatement of rights. If you cease any breach of the terms and conditions of this License, then your full rights under this License will be reinstated:

- a. Provisionally and subject to permanent termination until the 60th day after cessation of breach;
- b. Permanently on the 60th day after cessation of breach unless otherwise reasonably notified by the Licensors; or
- c. Permanently if reasonably notified by the Licensors of the violation, this is the first time You have received notice of violation of this License from the Licensors, and You cure the violation prior to 30 days after your receipt of the notice.

9.5 Notwithstanding the above, Licensors reserves the right to release the Database under different license terms or to stop distributing or making available the Database. Releasing the Database under different license terms or stopping the distribution of the Database will not

withdraw this License (or any other license that has been, or is required to be, granted under the terms of this License), and this License will continue in full force and effect unless terminated as stated above.

10.0 General

10.1 If any provision of this License is held to be invalid or unenforceable, that must not affect the validity or enforceability of the remainder of the terms and conditions of this License and each remaining provision of this License shall be valid and enforced to the fullest extent permitted by law.

10.2 This License is the entire agreement between the parties with respect to the rights granted here over the Database. It replaces any earlier understandings, agreements or representations with respect to the Database.

10.3 If You are in breach of the terms of this License, You will not be entitled to rely on the terms of this License or to complain of any breach by the Licensor.

10.4 Choice of law. This License takes effect in and will be governed by the laws of the relevant jurisdiction in which the License terms are sought to be enforced. If the standard suite of rights granted under applicable copyright law and Database Rights in the relevant jurisdiction includes additional rights not granted under this License, these additional rights are granted in this License in order to meet the terms of this License.