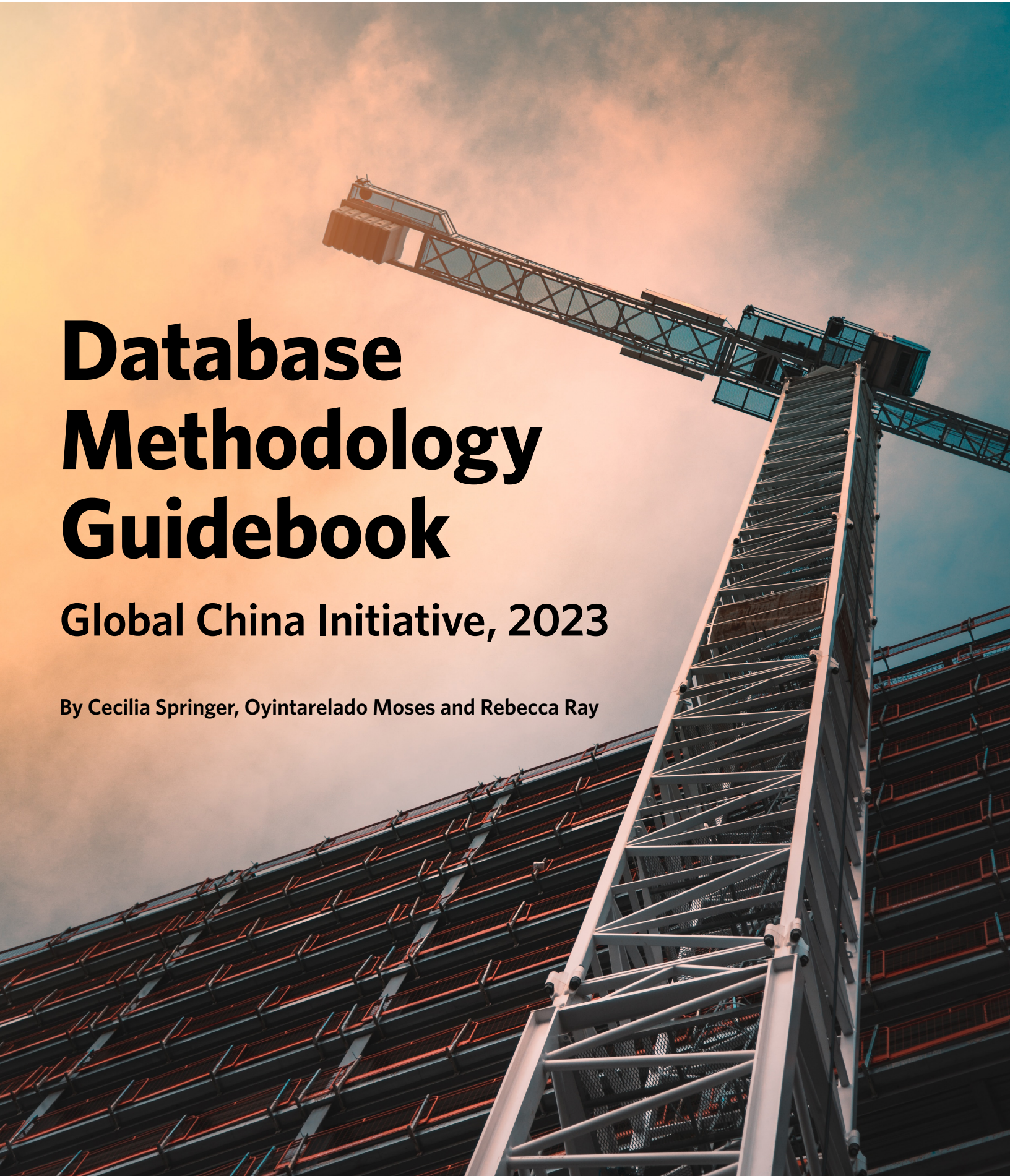


Database Methodology Guidebook

Global China Initiative, 2023

By Cecilia Springer, Oyintarelado Moses and Rebecca Ray





Cecilia Springer is the Assistant Director, Global China Initiative at the Boston University Global Development Policy Center. She earned a M.S. and Ph.D. from the Energy and Resources Group at the University of California, Berkeley and a B.S. in Environmental Science from Brown University.



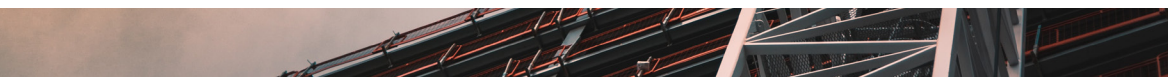
Oyintarelado Moses is the Data Analyst and Database Manager for the Global China Initiative at the Boston University Global Development Policy (GDP) Center. She holds an M.A. in China Studies and International Economics from Johns Hopkins University School of Advanced International Studies (SAIS), a certificate in Chinese and American Studies from the Hopkins-Nanjing Center and a B.A. in Political Science and Chinese Language from Duke University.



Rebecca Ray is a Senior Academic Researcher at the Boston University Global Development Policy Center. She holds a PhD in Economics from the University of Massachusetts-Amherst and an MA in International Development from the Elliott School of International Affairs at the George Washington University.

TABLE OF CONTENTS

INTRODUCTION: DATABASE OVERVIEWS AND OBJECTIVES	5
DATABASE SCOPES	5
Definitions	6
Description of Database Variable Entries	8
DATA COLLECTION METHODOLOGIES	11
Development Finance Data Sources	11
Other Loan Data Collection (CLA Database only)	14
Foreign Direct Investment Data Collection (CGP Database only)	14
Geolocation Methodology (CODF Database only)	15
Update Schedule	15
Frequently Asked Questions	17
Suggested Database Citations	18
Acknowledgments	19
Associated Papers	19
References	20





*Hong Kong, Hong Kong.
Photo by Meric Dagli
via Unsplash.*

INTRODUCTION: DATABASE OVERVIEWS AND OBJECTIVES

The Boston University Global Development Policy (GDP) Center is a University-wide center in partnership with the Frederick S. Pardee School of Global Studies. The GDP Center's mission is to advance policy-oriented research for financial stability, human well-being and environmental sustainability.

The Global China Initiative (GCI) at the GDP Center currently maintains a suite of five interactive public databases through the Data Analysis for Transparency and Accountability (D.A.T.A.) research group. The databases collectively track hundreds of billions of dollars in Chinese loans and investment to a variety of sectors, including energy and other types of infrastructure development. GCI databases track China's overseas development finance from China's two overseas-oriented development finance institutions (DFIs): China Development Bank (CDB) and the Export-Import Bank of China (CHEXIM). Some GCI databases track other sources of Chinese finance.

Through these databases, GCI provides transparent information to aid research, education, policymaking, reporting and accountability with regards to China's overseas development finance, foreign direct investment, other lending and projects around the world. This methodology guidebook explains how these databases were created, and how they are updated and maintained. These databases are open knowledge products, and as such, we welcome additions and corrections.

GCI databases include the China's Overseas Development Finance (CODF) Database, the Chinese Loans to Africa (CLA) Database, the Chinese Loans to Latin America and the Caribbean Database (CLLAC) Database, the China's Global Energy Finance (CGEF) Database and the China's Global Power (CGP) Database. The scope of each database can be found in the Database Scopes section.

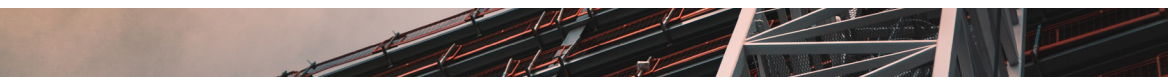
DATABASE SCOPES

The [China's Overseas Development Finance Database](#) is the first global, harmonized, validated and geolocated record of Chinese overseas development finance. As of January 2023, the interactive database covers the years 2008-2021 and includes loan commitments from China's two main development finance institutions (DFIs), the China Development Bank (CDB) and Export-Import Bank of China (CHEXIM), to governments, inter-governmental bodies, majority state-owned entities and minority state-owned entities with sovereign guarantees. The CODF Database also maps China's development finance in relation to several types of socially and ecologically sensitive territories: Indigenous peoples' lands, national protected areas and critical habitats.

The [Chinese Loans to Africa Database](#) is an interactive database tracking loan commitments from Chinese DFIs, commercial banks, government entities, companies and other financiers to African governments and state-owned enterprises. The database was created by the China Africa Research Initiative at the Johns Hopkins University School of Advanced International Studies (SAIS-CARI) and transferred to the GDP Center in March 2021.

The [Chinese Loans to Latin America and the Caribbean Database](#) is an interactive database tracking loans from two of China's DFIs to Latin American and Caribbean governments and state-owned enterprises. The database is jointly managed by the Boston University Global Development Policy Center and the Inter-American Dialogue.

The [China's Global Energy Finance Database](#) is an interactive database tracking loan commitments from China's DFIs to global energy projects, including infrastructure related to exploration and extraction, power generation, transmission and distribution, petrochemicals and energy efficiency.



The [China's Global Power Database](#) is an interactive database tracking global power plants outside of China financed by Chinese foreign direct investment and/or overseas development finance. The database also tracks and displays the deal types, Chinese investors and/or lenders, percentage of ownership by investor, amount of capacity in megawatts (MW), type of technology, operating status and the estimated annual CO₂ emissions of the power plants.

Definitions

DEVELOPMENT FINANCE INSTITUTIONS AND DEVELOPMENT FINANCE CDB and CHEXIM are China's development finance institutions (DFIs) (发展融资机构). We confine the range of "development finance (发展融资)" to loan commitments from CDB and CHEXIM due to their engagement in overseas financing.

A DFI is a stand-alone entity that primarily issues financial instruments such as loans for project-specific purposes with a public policy mandate, under a government-led strategy. They (1) are independent public institutions, (2) have a public policy mandate (i.e., development) and (3) deploy public finance and seek profits (Xu et al. 2021; Finance in Common 2023).¹

CDB was originally established as a policy financial institution in 1994.² The bank was incorporated as the CDB Corporation and officially classified as a DFI in December 2008. The bank is under direct jurisdiction of the State Council, and the Governor of CDB is a cabinet minister-level official. CDB also functions similarly to an export or overseas investment credit agency as a result of its

Table 1: Database Basic Information and Scope

Database	First Published	Time Period Covered	Updates	Geographical Scope	Observation Type*	Financial Scope
China's Overseas Development Finance Database	2020	2008-2021	Annual	Global	Loan	Development finance
Chinese Loans to Africa Database	2016	2000-2020	Annual	Africa	Loan	Development finance, other loan types
Chinese Loans to Latin America and the Caribbean Database	2012	2005-2022	Annual	Latin America, Caribbean	Loan	Development finance
China's Global Energy Finance Database	2018	2000-2021	Annual	Global	Loan	Development finance
China's Global Power Database	2020	2000-2032**	Every two years	Global	Power Plant	Development finance, Foreign Direct Investment

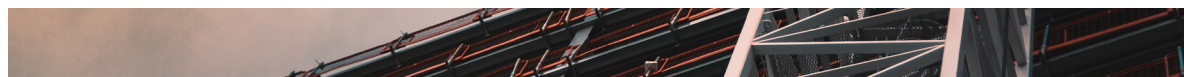
Source: Boston University Global Development Policy Center, 2023.

Note: * "Loan" refers to loan commitments, not debt or disbursement amounts.

** CGP tracks the year or projected year of commission for electric power generating units.

¹ CDB and CHEXIM are referred to as "policy banks" in previous Boston University Global Development Policy Center publications. However, the government of China classifies CDB as a commercial bank that is also a DFI, rather than a policy bank. Based on characteristics, both CDB and CHEXIM fit the definition of a DFI used by Finance in Common. For this reason, both CDB and CHEXIM are henceforth labelled as DFIs.

² CDB is defined as a DFI based on its function, although in recent years it has been designated as a commercial bank by the Chinese government. For CDB history, see CDB's 2021 Annual Report.



active involvement in supporting Chinese firms in China's overseas trade and investment finance (CDB 2022a).

CHEXIM, an export credit agency (ECA) dedicated to supporting exports and investments of Chinese companies, was established in 1994. CHEXIM can be considered a DFI, since its characteristics fit the Finance in Commons definition. Like CDB, the bank is under direct jurisdiction of the State Council.

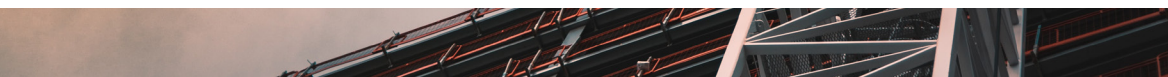
Before 2005, CHEXIM was China's only major overseas capital provider in energy projects and equipment exports (CHEXIM 2021). Prior to 2021, CHEXIM distributed two types of "preferential loans"/两优贷款, the concessional loan and the preferential export buyer's credits, but the bank also dedicates a large proportion of its credit operations to non-concessional loans (CHEXIM 2013). In 2021, China International Development Cooperation Agency (CIDCA), China's aid agency, became the administer of concessional loans.

BORROWERS Loans tracked in the five GCI databases are international sovereign loans, meaning the recipients are public entities, majority-public entities or minority-public entities with sovereign guarantees. Common borrowers of CDB and CHEXIM financing in the databases include:

- Recipient country governments, such as ministries of finance or central banks or financial institutions appointed and guaranteed by the sovereign government.
- State-owned enterprises (SOEs) of the recipient country. In these cases, companies sign contracts directly with Chinese banks.
- Joint ventures and special purpose entities with partial state ownership. These borrowers are included when the state has a majority ownership share or when a sovereign guarantee is issued for the loan.

LOAN TYPES The CODF, CLA, CLLAC and CGEF Databases track the following types of loan commitments to borrowing recipients.

- Export buyer's credits (EBCs)/买方信贷: EBCs are non-concessional loans that assist recipient countries in purchasing goods and services from Chinese firms (CHEXIM 2018).
- Concessional loans (CLs)/援外优惠贷款: CLs, one of the two types of preferential loans, are often considered official development assistance (aid). Intergovernmental agreements are required for this loan. CLs typically have a 2-3 percent fixed interest rates, long-term maturities of 15-20 years (grace period + repayment) and renminbi denomination (CIDCA 2021; Brautigam and Hwang 2022; CHEXIM 2013). Prior to 2021, CHEXIM distributed CLs. In 2021, China's aid agency, China International Development Cooperation Agency (CIDCA), became the administer of concessional loans and now classifies them as aid.
- Preferential export buyer's credits (PEBCs)/优惠出口买方信贷: PEBCs are another type of preferential loan and are the result of China's reforms to foreign assistance in the 1990s. They function similarly to export buyer's credit, but the Chinese government provides subsidies to lower the interest rate to a fixed, lower level. An intergovernmental agreement is not required. These loans are regulated by the International Cooperation Department and Department of Finance. PEBCs typically have 2-3 percent fixed interest rates, long-term maturities of 15-20 years and USD denomination (Brautigam and Hwang 2022; CHEXIM 2013).
- Interest Free Loans (IFLs)/无息贷款: IFLs are considered aid and are administered by the China International Development Cooperation Agency. They were previously administered



by China's Ministry of Commerce. IFLs accrue no interest and typically have long-term maturities of 10-30 years (Hwang and Moses 2022).

- Non-concessional loans: Non-concessional loans are loans with interest rates and terms comparable to the market rate or specified as a commercial loan/商业贷款 (Brautigam and Hwang 2022). The CDB, CHEXIM or other lenders can distribute these loans.

When collecting development finance information, GCI databases do not consider umbrella framework agreements as valid loan records (Brautigam and Hwang 2022). Unless specific projects are mentioned and evidence supports that the projects are in motion, statements pledging funds to general sectors are not recorded as development finance.³ The following types of economic cooperation involving CDB and CHEXIM are not recorded, even when covered by media:

- Memorandums of Understanding (MoU), as they are not considered confirmed loans.
- Financial agreements with other banks.
- Lines of credit, which are not loans by themselves.
- Proposed projects that have not met financial closure.
- Loans to projects that have been cancelled.
- Loans provided by other banks that are falsely reported as CDB or CHEXIM. These records will be deleted if there is further evidence indicating that these DFIs are not involved.
- Grants and technical assistance, which are unlikely to be issued by CDB and CHEXIM.

The GDP Center tracks co-financing between CDB, CHEXIM and other lenders, where applicable. Databases include co-financing information between these DFIs and other financing institutions when co-financing information is detected during source collection. When co-financing data is available, the CODF, CLLAC and CGEF Databases note the share of the total amount from CDB and/or CHEXIM. The CGP Database notes when power generating units have been co-financed by loans from CDB and/or CHEXIM, as well as FDI from Chinese companies.

In rare cases, CDB and CHEXIM loans are granted to the same project separately. If these loans are covered in different news reports, they are recorded as separate entries, as the loans may finance different phases of a larger project. If a loan is found to include multiple phases of disbursement, a revision or note in the previous entry will be added regarding the installment. In most cases, the percentage of contribution is not revealed. Loans are not assumed to be evenly supported by co-financiers.

Description of Database Variable Entries

The following tables describe the information displayed in each database's interactive website.

³ In rare cases, multiple projects are recorded in one report (with relatively high credibility, such as a Chinese embassy). However, the projects are not widely covered. In such cases, these projects count as one entry in a database.



Table 2: Variables Included in Each Database

Category	Variable	CODF	CLA	CLLAC	CGEF	CGP
Basic Information	BU ID	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
	Year	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
	Commission Year					<input checked="" type="checkbox"/>
	Project Name	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	
	Sector	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	
	Project Status					<input checked="" type="checkbox"/>
	Unit Name					<input checked="" type="checkbox"/>
	Plant Name					<input checked="" type="checkbox"/>
	Source	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
Geographic Information	Country	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
	Region	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>		<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
	State/Province					<input checked="" type="checkbox"/>
	BRI Country				<input checked="" type="checkbox"/>	
	Footprint Type	<input checked="" type="checkbox"/>				
	Precision Level	<input checked="" type="checkbox"/>				
Financial Information	Loan Amount (USD M)	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	
	Borrower	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
	Lender	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
	Other Lender	<input checked="" type="checkbox"/>				
	Chinese Policy Bank?					<input checked="" type="checkbox"/>
	FDI?					<input checked="" type="checkbox"/>
	Deal Type					<input checked="" type="checkbox"/>
	Investing Company					<input checked="" type="checkbox"/>
	Investing Company Ownership %					<input checked="" type="checkbox"/>
Energy and Environmental Information	Energy Source				<input checked="" type="checkbox"/>	
	Energy Subsector				<input checked="" type="checkbox"/>	
	Capacity					<input checked="" type="checkbox"/>
	Technology					<input checked="" type="checkbox"/>
	Estimated Annual CO ₂ Emissions from Power Generation					<input checked="" type="checkbox"/>
	Sensitive Territory Overlap				<input checked="" type="checkbox"/>	

Source: Boston University Global Development Policy Center, 2023.

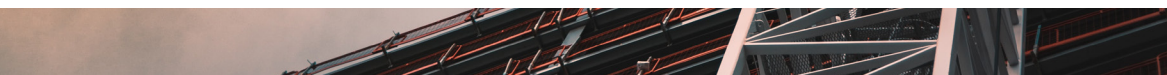


Table 3: Variable Descriptions

Variable	Description
BU ID	An ID code used to identify each project and find the same project in multiple databases.
Year	The year the loan contract was signed. In cases where reports of a certain project only includes information that the project is “to be financed by certain banks,” the date of the loan signing will be verified in a later iteration of the database.
Commission Year	Year a power plant entered or is scheduled for commission.
Project Name	Name or description of the project supported by the loan.
Loan (USD Million)	Loan amount in millions of US dollars at the current value of the contract, regardless of periods of distribution. In some cases, the loan is transacted in another currency and is converted using the annual average exchange rate from the International Monetary Fund of the respective year of the loans.
Country	The official name of the recipient country.
Region	The region where the project takes place. Regions are influenced by United Nations region classifications.
Borrower	The primary recipient of loans. If the loan is granted to an unspecified public agency, the borrower is recorded as “government.” Other recipients may include entities with majority public ownership and private entities that receive a government guarantee for the loan.
Lender	The provider of the finance. This is a Chinese DFI for most databases or a commercial/government lender for the CLA Database. In some cases where co-financing takes place, there are two DFIs designated as lenders and/or other co-financing institutions included.
Other Lender	Other lenders that provided loans to the project that are not DFI lenders. This category only applies to the CODF Database.
Sector	The destination sector of the project (energy, financial sector, health, industry and trade/ services, information and communication technology, social protection, transportation, non-energy mining, public administration, water/sanitation/waste, other). Sectors are influenced by World Bank sectors.
Source	The public sources used to confirm the loan signing.
Unit Name	Name of the electric power plant generating unit.
Plant Name	Name of the power plant, which could have more than one unit.
Capacity	Gross generating capacity of the unit in megawatts (MW). If the capacity of a pipeline or oil/ gas field is recorded, it is recorded using the unit of measurement of the source.
Technology	Primary fuel used for electricity generation (coal, gas, solar, wind, hydropower, nuclear, biomass, oil, waste, geothermal).
Estimated Annual CO ₂ Emissions from Power Generation (1000 ton) ⁴	For every fossil-based power project, we calculate its expected annual carbon dioxide emissions in kilotons.
Project Status	Project status (operational, under construction, or under planning).
State/Province	The province or state where the project takes place.
Chinese DFI?	Shows if a Chinese DFI is involved.
FDI?	Shows if foreign direct investment (FDI) is involved.

⁴ The estimation approach and original data can be found in the CGP Database 2022 policy brief ‘Understanding China’s Global Power: 2022 Update.’ <https://www.bu.edu/gdp/2022/10/19/understanding-chinas-global-power-2022-update/>



Variable	Description
FDI Deal Type	The type of FDI in the form of mergers and acquisitions (M&A) or greenfield investment.
Investing Company	The company that invested in the project.
Investing Company Ownership %	The percentage of ownership the Chinese company has in the investing company if known.
BRI Country	A country participating in the Belt and Road Initiative (BRI). The CGEF Database updates the list of BRI countries according to the latest reporting of countries that have signed the BRI Cooperation MOUs with China. ⁵
Deal Type	The type of financial arrangement: FDI (M&A, greenfield investment), development finance or involves both FDI and DFIs.
Energy Source	The primary energy source: biomass, coal, gas/LNG, geothermal, hydropower, nuclear, oil, solar, wind, unspecified. The energy source is usually specified in exploration projects and power generation projects. Transmission projects with a specified source, such as transmission lines from a dam, are recorded accordingly. Electricity distribution projects such as power grids that have no specified energy sources are listed as “unspecified electricity.”
Energy Subsector	The type of energy sector activity. This includes exploration and extraction; power generation; transmission and distribution; petrochemicals; energy efficiency; and multipurpose. The first three of these categories are the commonly used sub-sectors in the energy supply sector. Transport-related facilities for oil and gas are categorized in the transmission sector.
Footprint Type	The shape of projects with geolocation footprints.
Precision Level	The level of geolocation accuracy for projects with geolocated footprints. These levels include 1: Exact Location, 2: Within 25km; 3: second order administrative division; 4: first order administrative division; 5: Multiple first order administrative divisions, and 6: Nation.
Sensitive Territory Overlap	Based on spatial mapping, whether a project overlaps with Indigenous peoples’ lands (IPLs), critical habitats (CH) and/or national protected areas (PA). ⁶

Source: Boston University Global Development Policy Center, 2023.

DATA COLLECTION METHODOLOGIES

Development Finance Data Sources

The web scraping algorithmic and manual data collection processes are both utilized for the CODF, CLA, CLLAC and CGEF Databases as seen in Figure 1.

ALGORITHMIC DATA COLLECTION (CODF, CLLAC, CGEF AND CLA DATABASES POST-2022)

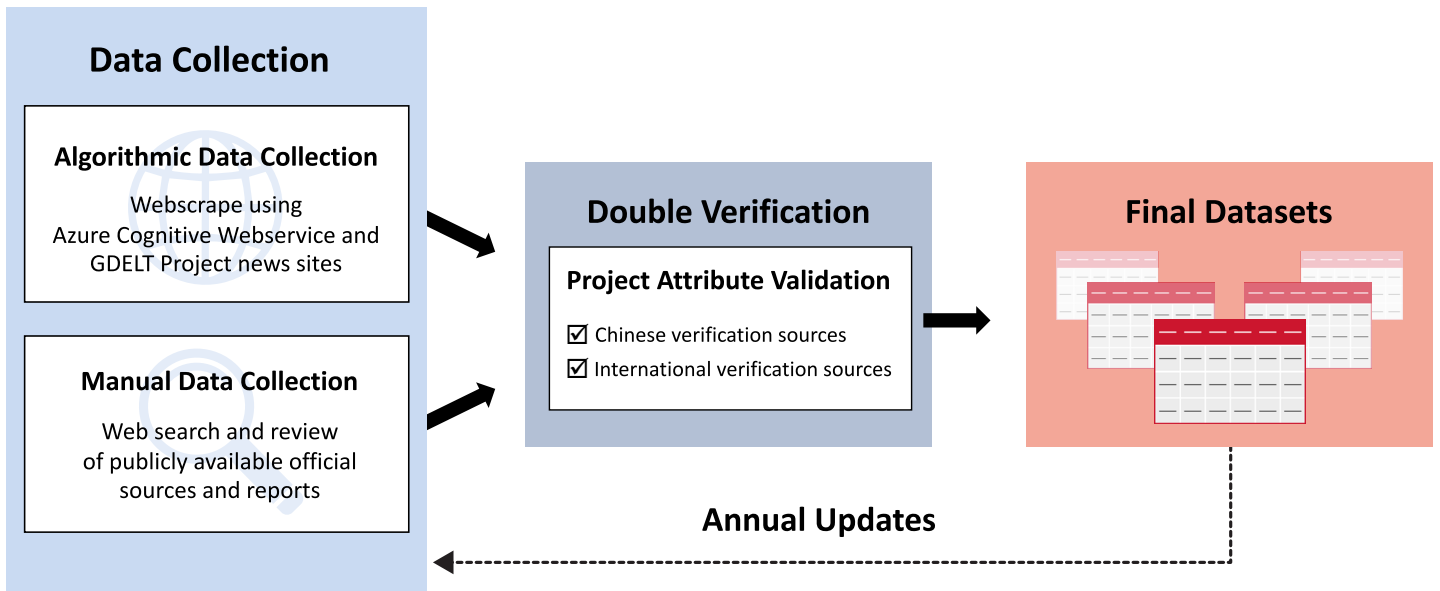
To algorithmically collect data, news aggregators and historical news services typically provide an Application Programming Interface (API) to facilitate programmatic access for web scraping. This allows many thousands of individual searches (e.g., “China Development Bank loan” in “English” for “August-1-2015” yielding 50 results) to be performed by an algorithm. An algorithm collecting this data may take several hours to days to collect the entire corpus of search results for complex projects but will extend the capacity of human researchers by increasing coverage and speed.

⁵ The list of BRI participating countries changes over time, and GCI databases will adjust the variable annually according to the list on the official BRI website (Belt and Road Portal).

⁶ For detailed definitions of these layers, see the journal article associated with the CODF Database published in *Nature’s Scientific Data* (Ray et. al. 2021).



Figure 1: Data Collection Method



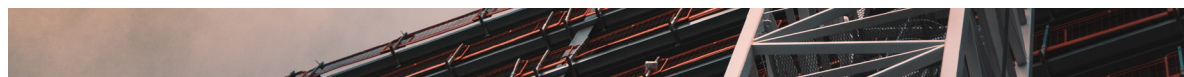
Source: Boston University Global Development Policy Center, 2023.

Two main data sources are used for the web scrape. The first, Azure Cognitive Webservice, is provided by Microsoft and is accessed by an API. One weakness of this API is the coverage, which is strongest around English articles. The GDELT Project acts as a supplement, as it has much greater coverage, especially for non-English languages, however, it can yield significantly higher false positive matches. False positives must be vetted manually, so a primary reliance on this data source is untenable.

The following search terms are used in web scraping:

- "China" or "Chinese" and
- "Development Bank," "CDB," "Export-Import," "Export Import," "ExIm," "Ex-Im," "Ex Im," or "Eximbank" and
- "Loan," "Lending," "Loaned," "Lends," "Credit," "Credits," "Finance," or "Financed," in English, Arabic, Bengali, Burmese, Cambodian, French, Indonesian, Lao, Mongolian, Portuguese, Russian, Spanish, Urdu or Vietnamese
- The name of at least one borrowing country, in English or in other languages commonly spoken in each, in noun or adjective form (for example "Hungary," "Magyarország," "Hungarian," or "Magyar," for Hungary, or "Bangladesh," "Bangladeshi," "বাংলাদেশ," or "বাংলাদেশী" for Bangladesh). For languages that assign genders to adjectives depending on the gender of the noun they are describing, both masculine and feminine versions of the adjective are included. For example, Brazil is included through the search terms: Brazil, Brazilian, Brasil, Brasileiro, Brasileira.

After initial record collection, additional algorithms reduce the data by filtering to remove duplicates, poor news sources and to remove articles that do not mention the key terms above in close proximity with one another. The records are then manually scanned for relevancy and further reduced. Then,



researchers manually check and read each of the remaining records against existing records in the GCI databases, resolving any conflicts in project attributes. Where conflicts arise between verification sources, government sources are given top priority, followed by academic sources, civil society sources and private press sources.

MANUAL DATA COLLECTION (CODF, CLLAC, CGEF AND CLA DATABASES POST-2022) To manually collect data, official sources and reliable sectoral sources are checked during every update cycle. This allows for double verification of each project and fills any research gaps from the algorithmic webscrape collection. Chinese and recipient country official sources are checked for the CODF, CLLAC and CLA databases. For the CGEF databases, energy-related official government websites of recipient countries and strong energy-related news aggregator websites are checked. The manual methodology is inspired by the China-Africa Loans Database Research Guidebook of the China Africa Research Initiative at the Paul Nitze School of Advanced International Studies (SAIS-CARI).

The following domains are included for China-based manual collection (in no particular order): (1) gov.cn; (2) cdb.com.cn; (3) Chinese embassies to foreign countries; (4) xinhuanet.com; (5) en.people.cn; (6) cgtn.com; (7) chinaplus.cri.cn; (8) chinanews.com; (9) cmw.cn and (10) china-daily.com.cn. The level of credibility of Chinese sources is generally ranked as follows: (1) Official websites of CDB, CHEXIM, and MOFCOM; (2) Websites of local embassies and involved companies; (3) China's official media such as Xinhua News Agency; (4) Academic journals (which typically use the above sources for evidence, but sometimes use interviews and other exclusive information); (5) Open reports from industrial associations, such as the China Electricity Council; (6) Commercial media (which usually forward important announcements from the above sources) and (7) Other. News released on forums and blogs is considered the least reliable source.

The following website types are used for recipient-based manual collection (in order of credibility): (1) Official websites of borrower ministries and embassies; (2) Official websites of recipients and involved companies; (3) Securities and Exchange Commission filings, International Monetary Fund Article IV consultations; (4) Surveying agencies and organizations with their own data; (5) Mainstream, state and industry specific news media; (6) Open source database and academic research (which typically uses the above sources for evidence) or NGOs and (7) Other. News reports that only appear in online forums without further evidence are taken as invalid. If the only evidence of a project is from a governmental source or global institution (such as the World Bank), the project is included, but only if details of the project are provided.

The following domains are used to manually search for energy-related projects (in no particular order): (1) Ministry of Energy and other energy related official websites; (2) power-technology.com; (3) renewablesnow.com; (4) pv-magazine.com; (5) hydroreview.com; (6) thewindpower.net; and (7) pv-tech.org.

DOUBLE VERIFICATION Projects are recorded if confirmed by one Chinese source and one external report (from recipient governments or international sources), indicating a loan signed by either CDB, CHEXIM or other Chinese lenders. In some cases, if a single highly official source such as a government budget document, SOE annual report, financial statement or stakeholder interview does not have an accompanying source, the highly official source is considered sufficiently authoritative to substitute for double verification. In this case, a loan with a highly authoritative source will also be supported with another source in the source column of the database, if a source is available. If a report can confirm a project but cannot confirm that the lender provided a loan to support the project, a loan is not recorded. For example, if a news report suggests that a loan was signed, this is not official enough to confirm the loan, therefore it will not be included. In the case that a project is officially canceled, the entry is deleted.



The GDP Center works with the Inter-American Dialogue to further verify loan commitments and projects for the CLLAC Database through their in-house expertise and connections with the region.

Other Loan Data Collection (CLA Database only)

The CLA Database collects data on all Chinese loans to Africa. Beyond China's DFIs, the CLA Database also tracks loans from Chinese commercial banks, Chinese state-owned enterprises (SOEs) or companies that provide supplier's credits or investment loan financing, the Chinese Ministry of Commerce and later the China International Development Cooperation Agency and the People's Bank of China. Prior to 2023, collection for these loans strictly followed the methodology from the China-Africa Loans Database Research Guidebook of SAIS-CARI (Brautigam and Hwang 2022). From 2023 onward, the data collection still follows these manual methodology practices and is supplemented by algorithmic webscrape data collection.

Foreign Direct Investment Data Collection (CGP Database only)

The core approach to compiling the CGP Database is to adapt the Standard and Poor's (S&P) Global World Electric Power Plants Database (WEPP) to account for the foreign ownership of overseas power plants invested in by Chinese companies. WEPP is an electric power generating unit-level database and its first global version was published in the early 1990s. The latest version available through the end of the calendar year is reviewed ahead of each CGP Database update. WEPP includes a comprehensive list of all power plants in the world derived from a variety of sources including direct surveys, power company financial and statistical reports, vendor reference lists and trade and business press. For every power generating unit, WEPP provides information regarding its name, operational status, capacity (MW), year of commission (coming online), primary fuels, pollution control equipment, owner company, location data and additional attributes (Platts, 2015). While WEPP includes the name of the company that has the majority ownership in a particular plant, it does not indicate the country of origin of the firm. In addition, there are a variety of subsidiaries or affiliates of the same holding companies and WEPP makes limited recording of them in a clear and consistent manner. An initial project list is developed using WEPP through three steps:

- 1) Generation of company list: A list of Chinese power companies that may invest globally is generated and updated by surveying a wide variety of online databases and public reports. Combining these sources, an initial list of companies considered to be potential investors in foreign power markets is created. Detailed lists of companies can be found in the 2020 CGP Database Methodology Note. It is important to note that this list may not be exhaustive, as there could be missing companies, especially small- to medium-sized private companies.
- 2) Direct matching with WEPP: After compiling the list of firms that invest globally in the power sector, the names on the company list are matched with companies listed in WEPP. A simple matching algorithm is used to search the WEPP database for names of firms on the company list and variations thereof. For every company, one to three keywords that are representative of all possible names used by WEPP is generated. Power plants are tagged with the name of a holding company when at least one of the keywords is found in WEPP indicating it is the original owner. All power plants obtained via this matching process are compiled in a new list.
- 3) Project status verification: It can take as long as a decade for an announced power plant project to begin operating. Many announced projects end up canceled or delayed and an incorrect recording of foreign outward investment could wrongly include these "zombie deals." The CGP Database relies on the project status information provided by WEPP, which records power generating units as in operation, under construction, under planning, deferred, canceled or retired.



Only projects that are in operation, under construction or under planning are included as valid investments. Investments at other project stages are excluded to avoid zombie deals. In addition, investments are validated by internet searches to avoid false matches (see Double Verification section). Website links are included in the database. Double counting is avoided by tracing investments to the power generating unit level and assigning unique project IDs.

Geolocation Methodology (CODF Database only)

The CODF Database interactive includes loan commitments tied to specific projects with physical footprints such as highways and power lines, as well as commitments for general financial support. Each loan-financed project is evaluated for geolocation. Projects with physical footprints are located through querying Google Maps, Open Street Maps and Open Route Services. Projects that are in planning or construction phases and do not yet have a finished footprint are geolocated via environmental impact assessments, feasibility studies, firm reports, and similar publications. These new and upcoming projects are reviewed in future years to capture any changes.

Each geolocated observation is then classified as a point (or collection of points), line (or collection of discontinuous lines), or polygon (or collection of discontinuous polygons). Points are used for individual buildings or installations. Lines are used for linear infrastructure including roads, rails, power distribution, wired communications networks and pipelines. Polygons show projects with footprints that are larger than single buildings or installations, with well-defined boundaries, including dam reservoirs, oil and gas fields, and clusters of buildings such as housing or stadium complexes.

The level of precision for geolocation data is then identified using different precision level requirements:

1. Exact project footprint,
2. Within 25 km, based on sources labeling a project as “near” a mapped place. This category also includes projects known at the second-order administrative division where that division has a radius of less than 25 km,
3. Second-order administrative division, such as municipality or county,
4. First-order administrative division, such as state or province,
5. Spanning multiple first-order administrative divisions,
6. Country,
7. Unknown.

Data on the geographic footprints of loan-financed projects can be accessed through submitting a data download form on the CODF Database interactive website and submitting a data use agreement to the GCI Database Manager at the GDP Center.

Update Schedule

Data updates for the CODF, CLLAC and CGEF Databases are published every year with a one-year time lag. For example, database updates published in 2022 would generally have data up to 2021. Updates to the CLA Database are also published every year, but with a two-year time lag. The CGP Database is published every two years with data collected into the end of the year prior to the year of release. Data collection for the databases occurs at the beginning of the calendar year, followed by data verification. During this time, new projects are added, dates of postponed projects are updated, repetitive and cancelled entries are removed and other considered revisions are made.

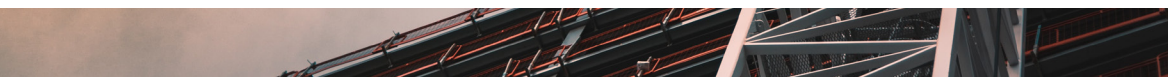




Photo by Jacek Dylag
via Unsplash.

FREQUENTLY ASKED QUESTIONS

How do I download the data?

Excel files for each database can be downloaded by request on the respective interactive websites by clicking the 'About/Data Download' button in the top right corner:

- [CODF Database](#)
- [CLA Database](#)
- [CLLAC Database](#)
- [CGEF Database](#)
- [CGP Database](#)

See suggested citations below for crediting GCI databases.

How can I submit a suggestion or update?

Suggestions and updates from contributors are welcome. To submit suggestions on methodology or specific loan or projects, please email the GCI Database Manager, Oyintarelado Moses, at odm@bu.edu.

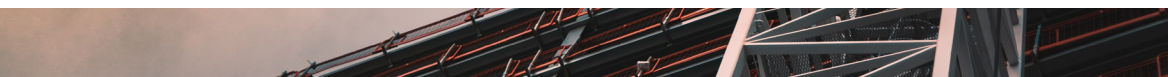
How do GCI databases compare to other databases tracking China's overseas finance and/or projects?

A concerted global effort is needed to build transparent knowledge and understanding of China's overseas involvement and the Belt and Road Initiative. GCI provides open-source information on projects in its databases and uses a double verification methodology to confirm the existence of loan commitments for each project. Unique attributes applicable to all GCI databases and specific to some GCI databases distinguish GCI databases from other databases tracking China's overseas finance and/or projects.

GCI databases mainly focus on tracking loan commitments from China's development finance institutions, CDB and CHEXIM. Finance from commercial, company and other government entities are tracked in the CLA Database. GCI databases only track loans to governments (sovereign), meaning that borrowers are public entities, majority-public owned entities or minority-public entities with sovereign guarantees. This scope allows for comparability with loan data from other development finance institutions, including the World Bank IBRD and IDA loan data. GCI databases do not track grants or scholarships. Additionally, each database update has an associated user-friendly interactive and policy brief, and in some cases an additional bulletin, contextualizing data trends.

The CODF Database provides precise geolocated information for projects receiving Chinese overseas development finance. Each project is evaluated for geolocation using a high level of verification. Through the CODF Database interactive, users can observe the loan-financed projects' geographic footprints and their overlaps with sensitive territories. Users can also gain access to a footprint map that provides a network view of projects by submitting a data use agreement (DUA) to the GCI Database Manager. The DUA is accessible through the data download page of the CODF Database interactive website.

The CGP Database tracks capacity (in gigawatts) and CO₂ emissions of overseas electric power plant generating units financed by China's overseas development finance and equity investments. It provides specific attributes for each project including energy sources, number of units and investment information for power plants financed by greenfield or mergers and acquisition equity investment.



Does the double verification methodology lead to systematic undercounting of China's overseas finance?

No. The double verification methodology ensures each loan signing is credibly verified based on publicly available information. The methodology also ensures that loans with only one identifiable official source are included in the database, since it is a credible source. Additionally, given that deals are frequently announced and later canceled, the double verification methodology ensures that canceled projects are not included in GCI databases. There is often a delay between the announcement and initiation of a project. In annual updates, confirmed loans and projects from prior years are added when updated information is available and verifiable, such as project completion.

How are lines of credit handled in the GCI databases?

A line of credit is noted when it is initially committed to projects within the scope of the databases, but not all instances where the line of credit is used or drawn upon are noted. Data on the line of credit is adjusted if it is later found and verified that it was never used, partially used or extended.

How can I better understand the data beyond analyzing the numbers?

Each dataset has an accompanying policy brief (in some cases an additional bulletin) that contextualizes the data and trends. These publications can be accessed online, specifically in the 'About/ Data Download' section of each database interactive.

How are users notified of changes made to databases in each cycle of publication?

The GCI Database Methodology Guidebook will be regularly updated to reflect methodological updates to GCI databases. Changes to a database will be reflected in the Appendix section of the accompanying policy brief or bulletin.

SUGGESTED DATABASE CITATIONS

Boston University Global Development Policy Center. 2023. China's Overseas Development Finance Database. Retrieved from <http://www.bu.edu/gdp/chinas-overseas-development-finance/>.

Boston University Global Development Policy Center. 2022. Chinese Loans to Africa Database. Retrieved from <http://bu.edu/gdp/chinese-loans-to-africa-database>.

Ray, Rebecca and Margaret Myers (2023) "Chinese Loans to Latin America and the Caribbean Database," Washington: Inter-American Dialogue.

Boston University Global Development Policy Center. 2022. China's Global Energy Finance Database. Retrieved from <http://www.bu.edu/cgef>.

Boston University Global Development Policy Center. 2022. China's Global Power Database. Retrieved from <https://www.bu.edu/cgp/>.



ACKNOWLEDGMENTS

Our efforts are inspired by the work of Deborah Brautigam and her team at the China-Africa Research Initiative in the Paul Nitze School of Advanced International Studies (SAIS-CARI).

The GDP Center thanks the Charles Stewart Mott Foundation, the Rockefeller Brothers Fund, the ClimateWorks Foundations, Carnegie Corporation, the William and Flora Hewlett Foundation, the Climate and Land-Use Alliance and the Packard Foundation for their generous support.

The CODF Database benefitted from foundational research of Rebecca Ray, Kevin P. Gallagher, William Kring, Joshua Pitts and B. Alexander Simmons and from valuable research assistance provided by Dora Betts, Alvin Camba, Siqi Chen, Hua-Ke (Kate) Chi, Victoria Chonn Ching, Jessica DiCarlo, Laura Gormley, Xinyi Han, Eleanor Horvath, Anjali Kini, Yudong (Nathan) Liu, Hsuan Lo, Yaxiong Ma, Ivan Nikitovic, Henry Nossiter, Andrea Pareja Villegas, Mackenzie Paskerian and Yiyang Xu. The GDP Center also gratefully acknowledges Oyintarelado (Tarela) Moses for data support.

The CLA Database, from 2022 onward, benefitted from the work of research managers and research assistants, including Kevin Acker, Jyhjong Hwang, Xinyue Ma, Oyintarelado Moses, Lucas Engel, Violet Barton, Victoria Bien-Aime, Samuel W. Croy, Yulian Fan, Yumao Gao, Zhenkun Huang, Abena Oduro, Reed Piercy, Sobia Shadbar, Keyi Tang, Ava Wang, Yinxuan Wang, Jiaming Liang and Hao Lyu. The GDP Center also acknowledges the work of the many SAIS-CARI research managers and research assistants in building the CLA database since 2007.

The CLLAC Database is the product of collaboration between the Inter-American Dialogue and the GDP Center.

The CGEF Database benefitted from the research management of Xinyue Ma and Cecilia Springer, with research assistance from Honest Shao, Junda Jin, Xintong Bu, Yanan Guo, Kehan Wang and Siqi Chen.

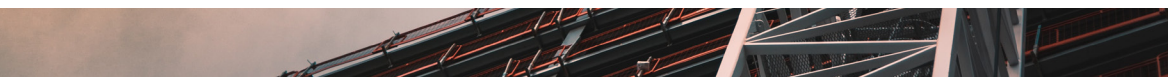
The CGP Database benefitted from the assistance of Professor Denise L. Mauzerall, Zhongshu Li and Xu Chen at the Department Civil and Environmental Engineering and the School of Public and International Affairs, Princeton University. The GDP Center also gratefully acknowledges Hua-Ke (Kate) Chi, Junda Jin, Yangsiyu Lu, Amanda Pareja and Cecilia Springer for research management and assistance.

ASSOCIATED PAPERS

B. A. Simmons, R. Ray, H. Yang, K. P. Gallagher., 2021. China can help solve the debt and environmental crises. *Science* 371: 466-468.

Chen, X., Gallagher, K.P. and Mauzerall, D.L., 2020. Chinese overseas development financing of electric power generation: A comparative analysis. *One Earth*. Available at <https://www.bu.edu/gdp/2020/10/08/chinese-overseas-development-financing-of-electric-power-generation-a-comparative-analysis/>.

Li, Z., Gallagher, K.P. and Mauzerall, D.L., 2020. China's global power: Estimating Chinese foreign direct investment in the electric power sector. *Energy Policy*, 136, p.111056. Available at <https://www.bu.edu/gdp/2020/10/23/chinas-global-power-estimating-chinese-foreign-direct-investment-in-the-electric-power-sector-2/>.



Ray, R., Gallagher, K.P., Kring, W., Pitts, J., and Simmons, B., 2021. Geolocated dataset of Chinese overseas development finance. *Nature Scientific Data*. Available at <https://www.nature.com/articles/s41597-021-01021-7>.

REFERENCES

Belt and Road Portal. https://eng.yidaiyilu.gov.cn/info/iList.jsp?cat_id=10076&cur_page=1

China Development Bank (CDB). 2022a. 2021 Annual Report. http://www.cdb.com.cn/English/gykh_512/ndbg_jx/2021_jx/

China Development Bank (CDB). 2022b. Export Buyer's Credit. http://www.cdb.com.cn/cpfw/gjyw/dkrz_434/201602/t20160219_2641.html

Deborah Brautigam, Jyhjong Hwang. 2022. China-Africa Loan Database Research Guidebook. <https://static1.squarespace.com/static/5652847de4b033f56d2bdc29/t/58ac6353f7e0ab024bcc665c/1487692628411/guidebook+draft+v.26.pdf>

Export-Import Bank of China (CHEXIM). 2013. "Export Import Bank of China: Two Preferential Loans Business Introduction." <https://www.chinca.org/cms/html/files/2013-12/16/20131216102948872930302.pdf>

Finance in Common. 2023. <https://financeincommon.org/why-finance-in-common>

Li Zhongshu, Ma, Xinyue, Gallagher, Kevin. 2020. "China's Global Power (CGP) Database Methodological Note." Boston University Global Development Policy Center. <https://www.bu.edu/gdp/2020/10/19/china-global-power-database-methodology-note/>.

Gallagher, K.P., Kamal, R., Jin, J., Chen, Y. and Ma, X., 2018. Energizing development finance? The benefits and risks of China's development finance in the global energy sector. *Energy Policy*, 122, pp.313-321. Available at <https://www.sciencedirect.com/science/article/abs/pii/S0301421518303975>.

The Export-Import Bank of China (CHEXIM). 2018. Export Buyer's Credit. https://web.archive.org/web/20221202145758/http://www.eximbank.gov.cn/business/credit/extrade/201806/t20180610_3513.html

Xu, Jiajun, Régis Marodon and Xinshun Ru. 2021. "Mapping 500+ Development Banks: Qualification Criteria, Stylized Facts, and Development Trends." Peking University Institute for New Structural Economics and French Development Agency. https://financeincommon.org/sites/default/files/2021-10/Mapping%20DFIs%20Worldwide_0.pdf

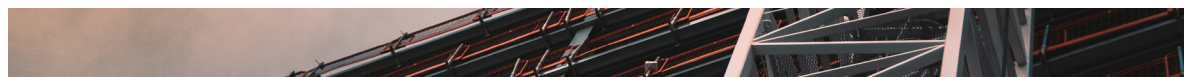




Photo by Guillaume Techer via Unsplash.