WORLD RESOURCES INSTITUTE

TECHNICAL NOTE

CHINA OVERSEAS FINANCE INVENTORY DATABASE

LIHUAN ZHOU, ZIYI MA, SHUANG LIU, XINYUE MA, KEVIN ACKER, MARGARET MYERS, DEBORAH BRAUTIGAM, AND KEVIN GALLAGHER

ABSTRACT

This technical note explains the process of building the China Overseas Finance Inventory (COFI) Database, a comprehensive database covering both Chinese equity and debt foreign investments in the power-generation sector in Belt and Road Initiative (BRI) countries. The note explains how the authors addressed the challenge of matching power plants and their investments across publicly available and commercial databases and how they provided the final database as an open data source.

INTRODUCTION

China's leadership is increasingly aware of the importance of environmental sustainability as a key factor in the success of the BRI, an initiative first raised publicly by Chinese President Xi Jinping in 2013 for intercontinental development and economic cooperation (MFA 2015). The BRI puts transportation and energy infrastructure among its top priority areas, and it emphasizes the consideration of environmental and climate-change impacts in infrastructure development. The Ministry of Ecology and Environment of the People's Republic of China (PRC) initiated the BRI International Green Development Coalition to promote green development of BRI (BRIGC n.d.). Data for tracking Chinese investment in the BRI is critical to understanding how green China's BRI investments are. More comprehensive data can also enable Chinese policymakers, financial institutions, and other stakeholders to hold themselves accountable and identify how they can make the BRI green.

CONTENTS

Technical notes document the research or analytical methodology underpinning a publication, interactive application, or tool.

Suggested Citation: Zhou L., Z. Ma, S. Liu, X. Ma, K. Acker, M. Myers, D. Brautigam, and K. Gallagher. 2022. "China Overseas Finance Inventory Database." Technical Note. Washington, DC: World Resources Institute. Available online at: https://doi.org/10.46830/writn.21.00003

CHINA AFRICA RESEARCH INITIATIVE | JOHNS HOPKINS SCHOOL of ADVANCED INTERNATIONAL STUDIES | THE DIALOGUE Leadership for the Americas | BU Global Development Policy Center

Despite the significant impacts of China's overseas infrastructure investment, there is no comprehensive database of these transactions and projects. There are no official statistics on investment disclosed at the transaction or project level from the Chinese government. Existing publicly available databases have an individual focus on sectors, geographies, and investors, or lack sufficient details. There is little coverage of publicly accessible data on loans by commercial banks and of equity investments. Data from different sources are in different formats and need to be harmonized to create a comprehensive database to support analysis and decision-making.

The COFI Database (the database) is a project to leverage existing public and commercial data sources and methodologies to build a comprehensive database on China's outward investment in the infrastructure sector. This project is a collaboration among the Global Development Policy Center at the Boston University, the Inter-American Dialogue, the China Africa Research Initiative at the Johns Hopkins University, and the World Resources Institute (WRI).

SCOPE

In the database, investment is defined as equity or debt investments provided by a Chinese entity to an overseas entity. This version of the database focuses on power-generation projects and includes two major types of information: technical characteristics and financial transactions of power plants. Each type of information includes several data points. The power project was chosen as the initial focus for the database because it is one of the infrastructure sectors that receives most of the investments of Chinese companies and investors (Zhou et al. 2018). The power sector is one of the most important sectors for the low-carbon transition, and some technologies have high potential for a long-term carbon lock-in effect, since newly built power plants will continue to emit greenhouse gas for the lifetime of the fossil fuel infrastructure, which will usually last for decades (Sato et al. 2021). Therefore, the sector plays an essential role in the current decarbonization effort. The plan is to expand the database to include other energy sectors, the transportation sector, and potentially other carbon-intensive infrastructure sectors.

The following key data points are included in the database:

- Financial transactions
  - Financial instrument: equity or debt
  - Investor name
  - Amount
  - Financial close year
- Technical characteristics
  - Name
  - Installed capacity
  - Commissioning year
  - Country
  - Primary fuel type

For financial transactions, the database includes both equity and debt investments made by corporates and banks, respectively, as these are often the most important components for financing power-generation projects, especially fossil-fuel generation. Chinese official data on BRI investments confirm that the major players for energy investment are almost exclusively state-owned enterprises through either foreign direct investment or through backing by Chinese policy and commercial bank loans (Wang 2021; IEA 2020). The equity investments include greenfield investments and cross-border mergers and acquisitions (M&As). Greenfield investments establish a new entity in a foreign country, while M&As involve the acquisition of existing firms.[1] The debt investments or loans include those from Chinese banks, including the China Development Bank (CDB), the Export-Import Bank of China (China Eximbank), and commercial banks.

The database is an evolving and changing database that currently covers Chinese power-sector project investments data in BRI countries from 2000 to 2020. BRI countries are defined as those that have signed BRI Memorandums of Understanding (MoUs) with China. There were 140 BRI countries as of January 2021 (See Appendix A) (Belt and Road Portal 2021). To enable greater research possibilities, we included projects financed earlier than 2013, the year BRI was first publicly initiated.

In this initial phase, the database primarily focuses on project financing because it is one of the most common and important ways to finance infrastructure in developing countries. Project finance is generally lent through a special-purpose vehicle, and its repayment relies on cash flow of the project and therefore is less risky for the sponsors and more suitable for capital-intensive financing like infrastructure (Dentons 2013). Project finance does not include Chinese investments in balance-sheet financing, corporate bonds, collateralized debt obligations, and

central-bank swap lines with People's Bank of China.[2] These types of financing are generally less transparent than financing for stand-alone projects, and there is limited information available on them (Mawutor 2015).

The database consolidates data from several existing databases (source databases) that cover different aspects of China-related financing. Our database does not contain new data. The consolidation matched entries from source databases to present values and types of investments to power projects in which Chinese investors are involved. The consolidation also removed duplicates to avoid double counting and to resolve conflicting information.

The remainder of this technical note is structured as follows: Section 3 describes our source databases. Section 4 outlines the consolidation methodology. Section 5 discusses the results from the database. Section 6 discusses the usage of the database and its maintenance plan. Appendix A lists the BRI countries. Appendix B provides detailed variable lists and explanations of the database, and Appendix C describes the consolidation methodology in detail.

DATA SOURCES

Investment Data

We researched existing publicly available and commercial databases and found that a combination of publicly available and commercial sources provides the best coverage of China's overseas finance. We include all four publicly available databases and three commercial databases (Table 1) using the following criteria:

- Data granularity is at the project or transaction level.
- There are robust and transparent methodologies documenting the data collection and verification process.
- Data cover most years in the period 2000–2020 and is preferably updated regularly.
- The commercial databases allow sharing of data with the public domain at a reasonable cost.

The additions of commercial databases filled two major data gaps left by publicly available databases. The first

Table 1 | **A Snapshot of Source Databases Included in the Database**

DATABASE NAME	China's Global Energy Finance	Chinese Loans to Africa Database	China-Latin America Finance Database	China's Global Power Database	Refinitiv Eikon M&A	Refinitiv Eikon Loan	fDi Markets
INSTITUTION/ COMPANY	Boston University Global Development Policy Center	China Africa Research Initiative and Johns Hopkins School of Advanced International Studies, managed by Boston University as of March 2021	Inter-American Dialogue and Boston University Global Development Policy Center	This database was originally from a journal paper. Boston University has adapted it into an online tool.	Refinitiv	Refinitiv	Financial Times
DESCRIPTION	Tracks China's global energy finance provided by two Chinese policy banks, CDB and China Eximbank	Collects data on Chinese loans to African governments and African state-owned enterprises	Tracks China's finance provided by CDB and China Eximbank to Latin American governments and their state-owned enterprises	Tracks China's foreign direct investment (greenfield and M&A) in power-generation projects	Collects merger and acquisition transactions	Collects bank loans	Collects greenfield FDI data

Table 1 | **A Snapshot of Source Databases Included in the Database (Cont'd)**

DATABASE NAME	China's Global Energy Finance	Chinese Loans to Africa Database	China-Latin America Finance Database	China's Global Power Database	Refinitiv Eikon M&A	Refinitiv Eikon Loan	fDi Markets
PUBLICLY AVAILABLE	Yes	Yes	Yes	Yes	No	No	No
OBSERVATION LEVEL	Project	Project by phase/loans by tranche	Project	Unit of a power plant	Transaction	Transaction	Transaction
TIME PERIOD	2000–2020	2000–2019	2005–2020	1893–2026	2000–2020	2000–2020	2004–2019
INCLUSION STATUS DEFINITION	Loan contract signed	Loan contract signed	Loan contract signed	(expected) Power plant commissioning	Deal announcement	Loan closing	Deal announcement
DATA COLLECTION PERIOD	Ongoing	Ongoing	Ongoing	2017, updated every 2 years	Ongoing	Ongoing	Ongoing
GEOGRAPHIC COVERAGE OF HOST COUNTRY	Global	Africa	Latin America	Global	Global	Global	Global
NUMBER OF OBSERVATIONS	289	1077	94	976	189	186	202
ABBREVIATION	BU-CGEF	SAIS-CLA	IAD-GEGI	Journal-CGP	Refinitiv-M&A	Refinitiv-Loan	fDi Markets

Note: There are three major types of milestone dates included in source databases: deal announcement, loan closing or loan contract signed, and (expected) power plant commissioning. For greenfield projects, deal announcement typically happens at the very early stage of the project development phase, and loan closing or loan contract signed are typically at the end of the project development phase after a series of negotiations (Dentons 2013; DOE 2015). The target of an M&A transaction is typically a company that is already in operation. Power plant commissioning is the beginning of the operation phase, which is after the project development and construction phases (DOE 2015).

Source: Authors, based on CARI et al. 2021; Financial Times 2021; Gallagher 2021; Gallagher and Myers 2021; Li et al. 2020; Refinitiv 2021).

data gap is loans financed by Chinese commercial banks. The database uses loan data from Refinitiv Eikon to provide loans financed by these banks. The second data gap is equity investments, and we supplement publicly available sources with greenfield foreign direct investments (FDI) from fDi Markets, a division of Financial Times, and M&A data from Refinitiv Eikon. Figure 1 provides different types of equity and debt investments and the coverage of datasets included in our database.

Power Plant Characteristic Data

Power plant characteristic data rely on two databases, the Platts' UDI World Electric Power Plants Database (WEPP) and the Global Power Plant Database (GPPD) (Byers et al. 2018; Platts 2018). WEPP and GPPD supplement investment source databases with additional characteristics, including geolocation, commissioning year, installed capacity, primary fuel, and other technical information (Platts 2018; Global Energy Observatory et al. 2018). Table 2 provides information on the two power plant datasets that we used.

Figure 1 | **The Coverage of Source Databases Included in the Database for Different Types of Equity and Debt Investments**

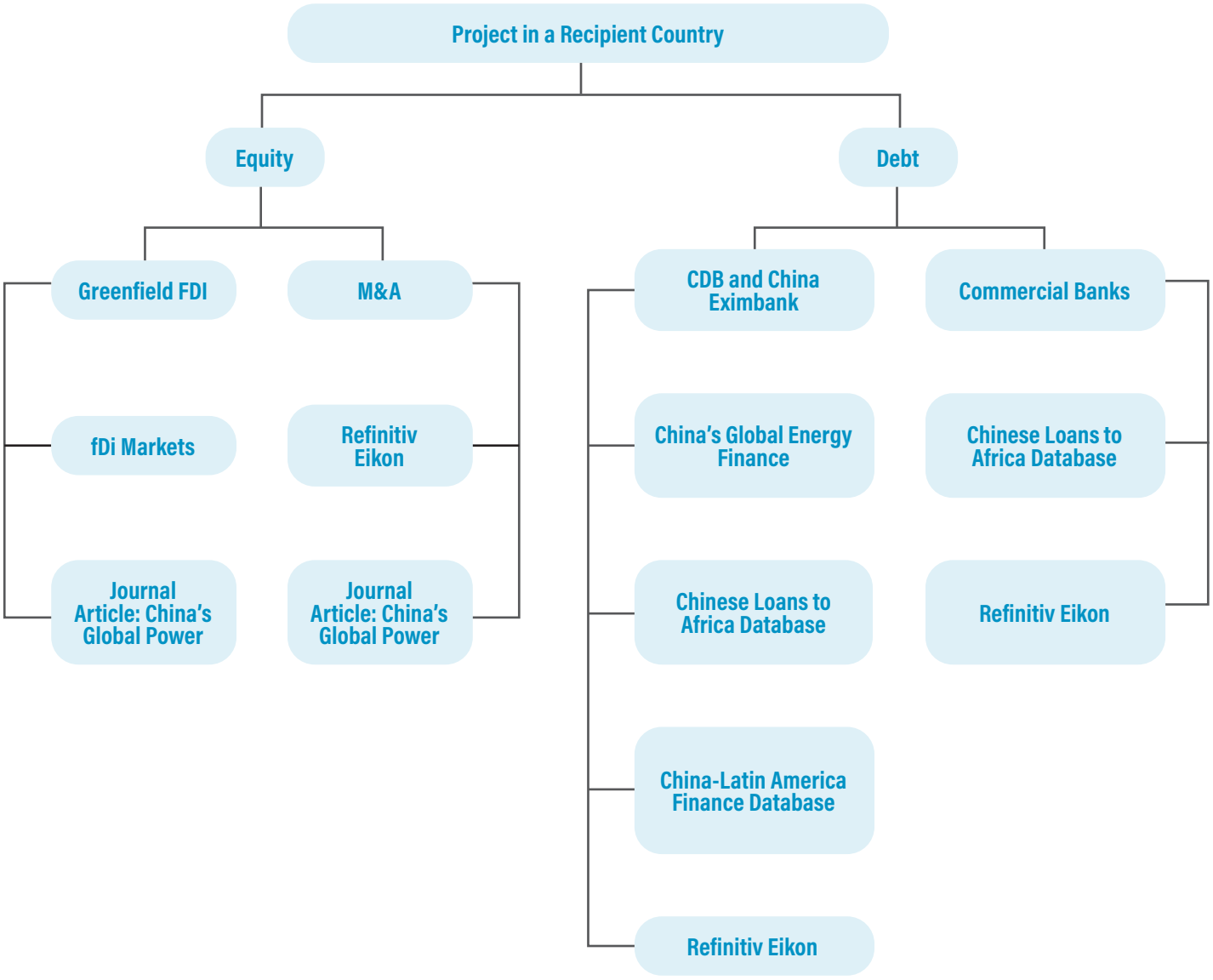


Table 2 | **A List of Power Plant Databases and Their Descriptions**

DATABASE NAME	OBSERVATION LEVEL	COMMISSIONING YEAR	DATA COLLECTION PERIOD	POWER PLANT GEOGRAPHIC COVERAGE	TYPE OF COVERAGE	NUMBER OF OBSERVATIONS	UNIT OF OBSERVATION	ABBREVIATION
UDI World Electric Power Plants Data Base	Generation unit of power plant[a]	1878-2030[b]	Ongoing	Global	Global power plants	211,829	A generation unit of a power plant	WEPP
Global Power Plant Database	Power Plant	1896-2018	Ended in 2018	Global	Global power plants	29,910	A power plant/ A unit of power plant	GPPD

Notes:

[a] Some observations were on a unit level and we consolidated them to power plant level before implementing joins. Appendix C discusses this in detail.

[b] The commissioning year is beyond 2018 because Platts also includes power plants in the planning stage or under construction. Commissioning years for these power plants are estimates.

Sources: Authors, based on Byers et al. 2018; Platts 2018.

Coverage Limitation

Because the database does not generate new data and relies on the source databases mentioned earlier for inputs, there are limitations in terms of coverage and accuracy of some data entries.

Data gaps in the source databases include the absence of all equity investment amounts and some capacity information. The three databases (fDi Markets, Journal-CGP, and Refinitiv M&A) we rely on for equity investment amounts all had missing information. Due to the difficulty of gathering equity amounts paid and estimating the amounts in an M&A deal, Journal-CGP and Refinitiv M&A do not provide this information. FDi Markets provides the total capital expenditure of a greenfield FDI project, but it is not the equity amount paid by the Chinese investors.

Data inaccuracy issues exist even when all source databases provide the necessary information for a project. The first case is the possible changes in project specifics that the source data failed to update in a timely manner. For example, the financing structure and ownership can change during the construction and operation of a project. However, these types of changes are seldom reported and difficult to track; therefore, the database may not provide up-to-date information. Data inaccuracy can also be a result of unclear syndicated loan structure. Sometimes the exact loan contribution from individual banks is not available from source databases, and sometimes lead arrangers may carry much more weight than ordinary participants. For cases like this, the database takes an average of the total loan, which will be discussed in detail in the loan estimation section.

Despite the efforts to provide a complete picture of Chinese BRI investments, the source data can still miss certain information. Because the databases from which we sourced mostly rely on secondary information, including news articles and information from government websites, without systematic disclosures directly from Chinese financial institutions and equity investors, secondary information is highly unlikely to capture every project financed by Chinese entities. For example, the database aims to cover investments by Chinese equity investment funds, such as the Silk Road Fund, through several source databases. However, it does not include

any projects financed by those entities, and Zhou et al. (2018) show that the Silk Road Fund invested in power-generation projects.

CONSOLIDATION METHODOLOGY

When combined, the source databases provide a comprehensive view of a project (like a mosaic). However, we could not simply concatenate the databases together since some projects could be included in multiple data sources. By consolidating—matching entries from different databases—we could avoid the issue of double accounting while improving the quality of coverage by piecing together different information collected by these databases.

The different data collection methodologies, classifications, and definitions used by the various source databases presented several challenges as we sought to consolidate them. Each source database used a different data collection method and quality-control standard, which resulted in varying data quality. Some relied on a mix of news reports and official government announcements, and others relied on information provided by financial institutions. Table 3 describes original data sources of the nine source databases. There is also data inconsistency among different databases, including having different names for the same power plant, defining the variable year differently, etc. For example, some databases used the term *year* as the commissioning year of a power plant while others used the same term as the year when the investment deal was signed or closed. In addition, there were missing investment amounts data from some original source databases.

We took three major steps to address these challenges and consolidate these databases into the coherent and comprehensive database. We first performed data preprocessing to standardize data. Then, we used a computer program to automatically join various databases. Finally, we did a comprehensive manual matching to ensure accuracy

Table 3 | **Original Data Sources of the Nine Source Databases**

SOURCE DATABASE	DATA SOURCES
UDI World Electric Power Plants Database	Direct surveys, vendor reference lists, power company financial and statistical reports, and the trade and business press
Global Power Plant Database	National government agencies, reports from companies that build power plants or provide their components, data from public utilities, and information from multinational organizations
China's Global Energy Finance	Official CDB and China Eximbank websites or host-country ministries, news reports, and official documents
Chinese Loans to Africa Database	Borrowing government, bank, company (including SEC company filings), and press reports in both China and the borrowing country in addition to in-country contacts and field visits
China-Latin America Finance Database	Official CDB and China Eximbank websites or host-country ministries, news reports, and official documents
China's Global Power: Estimating Chinese Foreign Investment in the Electric Power Sector	UDI World Electric Power Plants Database, government websites, company websites and annual reports, trade and business press
Refinitiv Eikon M&A	Direct-deal submissions from global banking and legal contributors, regulatory filings, corporate statements, media, and pricing wires
Refinitiv Eikon Loan	Direct-deal submissions from global banking and legal contributors, regulatory filings, corporate statements, media, and pricing wires
fDi Markets	Publicly available sources: Financial Times, media sources, agency sources, purchased data

Source: Authors, based on Byers et al. 2018; China Africa Research Initiative et al. 2021; Financial Times 2021; Gallagher 2021; Gallagher and Myers 2021; Li et al. 2020; Platts 2018; and Refinitiv 2021.

of the consolidation and resolve conflicting information for each entry. We collaborated with Sonoma Technology to implement the data preprocessing and automated join. The remainder of this section discusses these three steps in detail.

Data Preprocessing

We took three steps in preprocessing to standardize and prepare data for better matching results. The first step was to screen projects for their relevance, according to these criteria:

- The project reached financial closure between 2000 and 2020. For example, projects that were part of an MoU, a type of preliminary agreement requiring further negotiations, were excluded (Brautigam and Hwang 2020).[3]
- The projects were power-generation projects designed to connect to the grid. Power transmission and distribution projects and captive power plants were excluded.[4]
- The projects had Chinese financing in either debt or equity. For example, engineering, procurement, and construction projects undertaken by Chinese firms without financing were excluded.
- The database excluded projects that had no information source, no project specifics, and could not be verified by additional research.[5]
- The database excluded duplicated projects within and across each database.[6]

The second step was variable name standardization, which entails mapping and standardizing the variable names from all source databases for matching. Table 4 shows the common variables shared by different databases with corresponding variable names specified under the unified variable name row. Some source databases did not provide information related to these variables in stand-alone columns but, rather, in a single description column. To make use of all the available information from each database, we extracted relevant data when available.[7] Table 5 shows the variables that were directly used in our joining process and databases from which we needed to extract it.

The last step was standardizing categories for three categorical variables, because different databases used different methods and classifications for recording underlying categories. The three variables are primary fuel, country (iso3c), and equity investor name. We used standard names for fuel sources that were consistent with the database. We also used the country-code package to standardize country names and to obtain three-letter country codes defined in ISO 3166-1 for all datasets (Arel-Bundock et al. 2020). We added additional categorical variables to classify each entry in the database into certain categories, such as whether this entry was an equity, a debt investment, or both.

Table 4 | **Source Databases and Corresponding Field Names Mapping**[a]

FIELD NAMES/ DATASETS	BU-CGEF	SAIS-CLA	FDI MARKETS	IAD-GEGI	JOURNAL-CGP	WEPP	REFINITIV-LOAN	REFINITIV-M&A	GPPD
power_plant_name	description	purpose		purpose		plant	use_of_proceeds_notes	synopsis	name
country	country	country	destination_country	country	country	country	issuer_borrower_nation	target_nation	country_long
country_iso3c	country_code								country
province	location	destination_state				state			
city	location	destination_city				city			
installed_capacity	mw_if_applicable	purpose	description		capacity_mw	mw	use_of_proceeds_notes		capacity_mw
commissioning_year			description		online_year	year			commissioning_year
primary_fuel	energy_source	purpose	fuel_source		technology	fuel	use_of_proceeds_notes	synopsis	primary_fuel
latitude									latitude
longitude									longitude
total_investment_amount[b]	amount_m	usd_m	capital_investment	amount					
equity_investor_name_1			investing_company		major_chinese_shareholder	company	acquirer_name		
equity_investor_amount_1			capital_investment					deal_size_m_usd	
equity_investment_type					type_of_investment			form_of_the_transaction	
equity_investment_year			project_date					date_effective	

Table 4 | **Source Databases and Corresponding Field Names Mapping[a] (Cont'd)**

FIELD NAMES/ DATASETS	BU-CGEF	SAIS-CLA	FDI MARKETS	IAD-GEGI	JOURNAL-CGP	WEPP	REFINITIV-LOAN	REFINITIV-M&A	GPPD
parent_company_of_investor			parent_company			parent		acquirer_ultimate_parent	
debt_investment_year	date	year		year			loan_dates_facility_closing_date		
debt_investment_amount	amount	US$		amount					
number_of_lenders	Lender	Lender		Lender			all_managers		
china_development_bank	Lender	Lender		Lender			loan_manager_commitment_amount		
export_import_bank_of_china	Lender	Lender		Lender			loan_manager_commitment_amount		
Industrial_and_Commercial_Bank_of_China		Lender		Lender			loan_manager_commitment_amount		
China_Construction_Bank		Lender					loan_manager_commitment_amount		
Bank_of_China		Lender		Lender			loan_manager_commitment_amount		

Table 4 | **Source Databases and Corresponding Field Names Mapping[a] (Cont'd)**

FIELD NAMES/ DATASETS	BU-CGEF	SAIS-CLA	FDI MARKETS	IAD-GEGI	JOURNAL-CGP	WEPP	REFINITIV-LOAN	REFINITIV-M&A	GPPD
Agricultural_ Bank_of_China		Lender					loan_manager_commitment_amount		
China_Citic_ Bank_Corp				Lender			loan_manager_commitment_amount		
China_Merchants_Bank							loan_manager_commitment_amount		
China_Minsheng_Banking_Corp							loan_manager_commitment_amount		
China_ Zheshang_ Bank							loan_manager_commitment_amount		
China_Everbright_Bank							loan_manager_commitment_amount		
cross-checking IDs	bu_id	sais_id	fdi_id	iad_id	unit_id in Platts	comp_id; location_id; unit_id	r_id	rma_id	gppd_id

Notes: [a] All field names were converted to lowercase for mapping.

[b] Total investment amount measures the total investment cost of a power plant.

Source: Authors.

Table 5 | **Extracted Variable Names from Corresponding Datasets**

VARIABLE NAMES	EXTRACTED DATASETS
power_plant_name	BU-CGEF, SAIS-CLA, Refinitiv-M&A, Refinitiv-Loan
capacity	SAIS-CLA, fDi Markets, Refinitiv-Loan
primary_fuel	SAIS-CLA, Refinitiv-M&A, Refinitiv-Loan

Source: Authors.

Automated Join

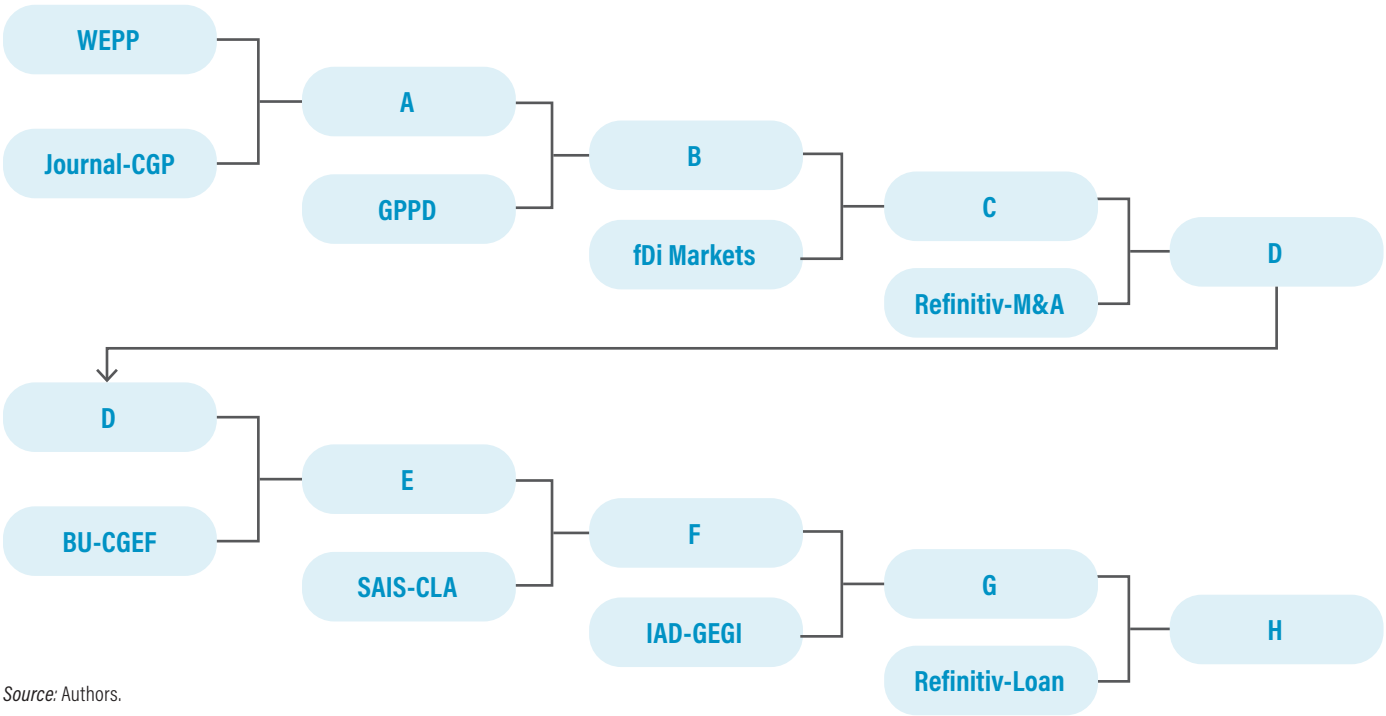
A major challenge of matching records was the lack of a common unique identifier in most source databases, which prevented us from matching them definitively. We used R studio (R Core Team 2020) to match and join databases two at a time. Figure 2 below shows an overview of the join steps. Appendix C includes detailed join plan for each step.

We used *equal join*, which matches strings on field equivalence, to match databases and variables if they shared a common unique identifier. This applies to the first three databases we joined; that is, WEPP, Journal-CGP, and GPPD. We joined WEPP and Journal-CGP at the generation unit level using a unique identifier, UNITID, and then converted the join result into the power-plant level to further join with GPPD using a unique identifier LOCATIONID. UNITID is a power-plant unit identification number, and LOCATIONID is a plant identification number. We also used *equal join* for two variables, country_iso3c (ISO 3166-1 alpha-3 country codes) and primary_ fuel. We standardized these two variables across databases and could check if values from two databases were equal.

For the rest of the databases, we used fuzzy join (also called fuzzy matching), a technique of identifying non-exact matches, to join source databases together (Figure 3) (Robinson et al. 2020). The fuzzy join method joined two databases together based on similarities of shared variables.

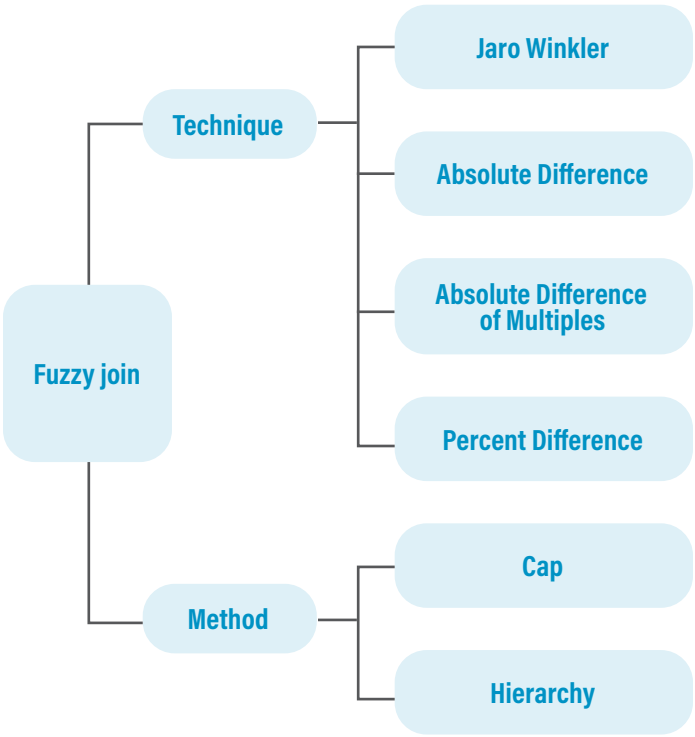
We used Jaro-Winkler Distance to compare similarity between texts, such as power-plant names. The shorter the distance, the more similar the two strings are. However, the Jaro-Winkler score is calculated as the inversion of the distance (1 – Jaro-Winkler Distance) and therefore is scaled between 0 (not similar at all) and 1 (exact match) (Cohen et al. 2003). The remaining fuzzy join techniques are for numeric, the absolute difference ($|x - y|$), the absolute difference of multiples (min ($|x - y|$, $|2*x - y|$, $|3x - y|$, … , $|2y - x|$, $|3y - x|$, …)), and the percent difference $(x - y)/x$.

Figure 2 | **Join Steps**



Source: Authors.

Figure 3 | **Techniques and Methods Used for Joining Datasets**



Source: Authors.

In addition to these quantitative metrics, we used two constraints called cap and hierarchy to increase join accuracy. A cap is the maximum value of distance for which to return a match; for example, we capped Jaro-Winkler Distance at 0.2 (which would mean a score of 0.8) so that we would not match two power plants with very different names. Numeric caps differed by variables, and we discuss them in detailed join steps in Appendix C. The hierarchy is the order of match importance across variables being matched. We ranked variables higher and gave them higher priorities for join when we had more confidence in their quality and accuracy.

Several limitations in the automated join and source data result in a low confidence in the accuracy of the automated join result. The first limitation is limited and inconsistent project-level information across source databases. Some important variables for automated join are not available in some source databases. Even for those variables that are available, inconsistencies among databases create difficulties for a successful match. For example, equity investor names are sometimes inconsistent because of differences between parent and subsidiary companies or changes in company's name.

Another limitation of automated join is the lack of a metric to ensure that an automated match result is correct. We use a confidence score to rank the possible join results, but it is difficult to determine a threshold of confidence score above which the result is definitively correct. This limitation can further hamper efforts to expand the automated join method to include other energy and transportation sectors, as ultimately, we will likely need manual checking to ensure the accuracy of results.

Manual Join

We performed manual join in the same sequence as fuzzy join to check for accuracy of the automated join and make corrections. Two bilingual analysts who are proficient in both English and Chinese individually matched each database and resolved discrepancies based on extensive research and discussion. We conducted additional research to check the accuracy of the original source and to gather more information on a project, which ultimately improved the overall quality of the database. The following steps summarize the process of manual join:

- Information: We mainly relied on six types of information to determine if two projects were a match:
  - country
  - fuel source of the power plant
  - name of the power plant (additional research)
  - investor names (both debt and equity) (additional research)
  - installed capacity (10% variation)
  - detailed location (city and province)
- Criteria: We determined that it was a correct match when both entries were in the same country using the same fuel source, and at least three of the remaining four variables above were consistent in two databases. A few power plants met the above matching criteria but were in different phases of a project (some power projects consisted of different plants built in different years, differentiated by phase 1 or 2) and had to be recorded separately, but most of the instances

were successful matches. We allowed a roughly 10 percent variation in capacity measured in megawatts (MW) when matching to account for different definitions of capacity.[8]

- Additional research: We also conducted additional research in both English and Chinese to identify investors and power-plant names. Each database sometimes recorded investors differently, with some using the name of the company that made the investment and some used the name of the parent company.
- Unmatched: For the rest of the projects that did not yield a match, we kept them as separate entries.

We used two databases to resolve two types of conflicting information that we found during matching. Different databases could cover the same type of information but provide conflicting information for a project due to their varying collection and classification methodologies.

- For conflicting information related to power plant characteristics, such as the name of a power plant and installed capacity, we used WEPP as the standard because it provided the most comprehensive design information for power plants at the global level, and it used a combination of direct survey and secondary information collection (Platts 2018).
- For conflicting amounts of debt investment, we used the information from SAIS-CLA and Refinitiv Loan as the standard, when possible, as they provided more detailed loan terms and have a more rigorous loan collection methodology.

Although manual join can ensure accuracy of results, this process is time- and labor-consuming, and makes it difficult to replicate for other databases and researchers. It involves using external sources other than source databases to confirm existing information and find additional information. This process is not completely objective, and there could be errors even with multiple reviewers.

Estimation of Individual Loan Contributions

Information on investment value by individual investor provides useful details to analyze financial flows. However, this type of information is not always available. All 220 projects with equity investment do not have reported investment values.[9] We kept these values as missing.

Debt investments in general have better coverage. Of the 253 debt investments, 222 have investment amounts by individual investors (88 percent). The rest of the debt investments, or 12 percent, have total investments available.

We extracted or estimated individual loan contributions from total investment based on the level of investment details available in source databases. There were two types of situations:

- If individual loan contributions were missing but both the total amount of the loan and the number of participants were available, we used the average loan contribution as an estimate for individual loan contributions.[10] For example, some projects in BU-CGEF were financed by both CDB and China Eximbank without their individual contribution. We divided the total amount evenly and assigned it to the two banks.
- When an entry in a source database contained financing for multiple power plants with one aggregate investment amount, we calculated the investment amount for each power plant weighted by its installed capacity.

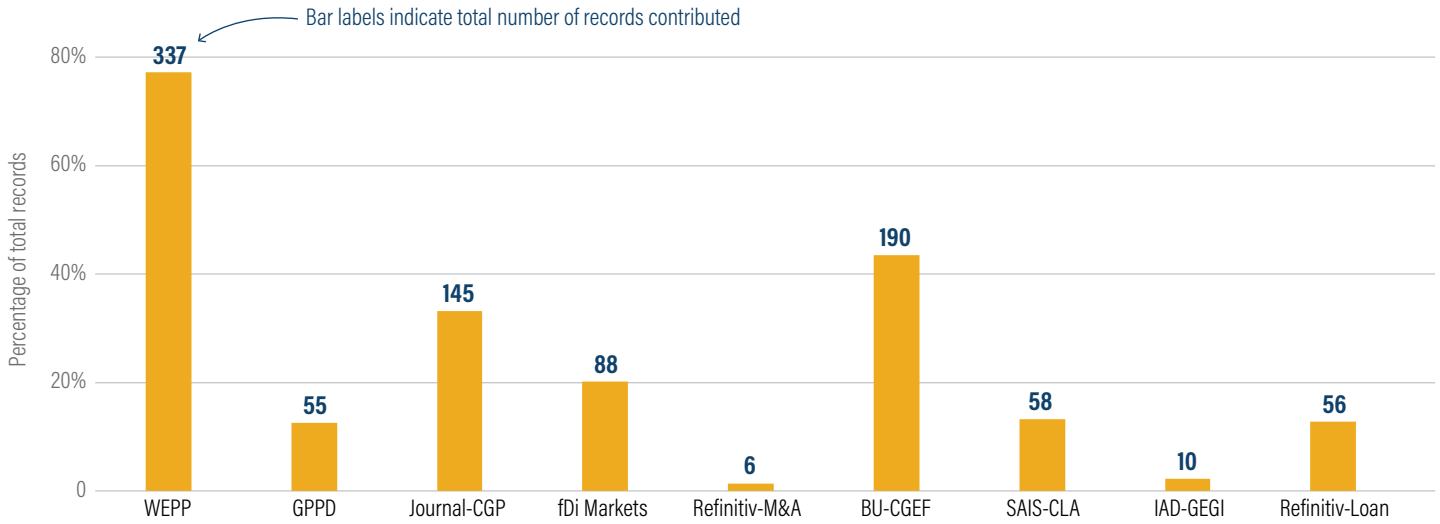
RESULTS

The final database has 443 observations for 430 power plants after the automated join and manual join. Each observation of the database includes details of power-plant characteristics, equity and/or debt investor names, and investment amount (if available) for a power plant. The database includes 473 equity and debt investments (253 debt investments and 220 equity investments). The number of investments is more than the number of observations because some observations record both a debt and an equity investment made to the same project.

In addition, some power plants have received multiple debt investment transactions in different years.[11] The database records these transactions as separate observations since each observation can only include debt investments made in a single year.[12] In the database, two power plants have three debt investments in three different years, and eight power plants have two debt investments in two different years. As a result, the two power plants have six observations, and the eight power plants have 16 observations, or a total of 22 observations (for 10 power plants).

The database contains nine source databases, and their contributions to the database vary, with WEPP contributing the most (Figure 4). There are significant overlaps among different investment databases. Ninety-six records, or 22 percent of the total record, are from multiple investment sources as shown in Figure 5. The most

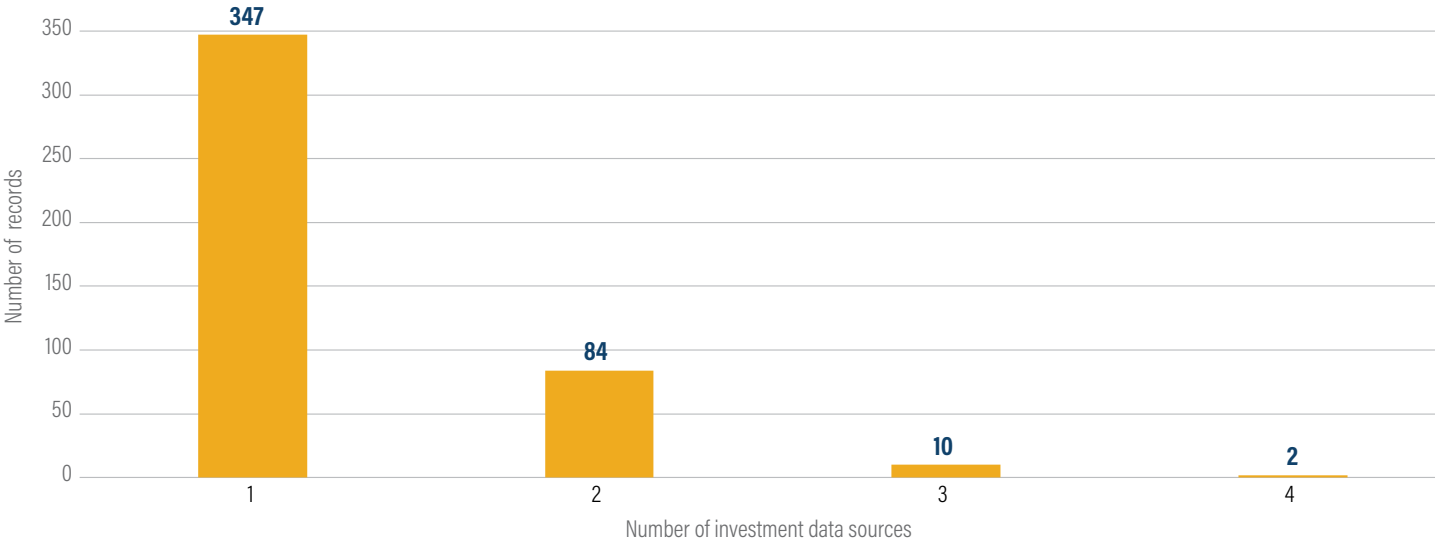
Figure 4 | **Data Source Contributions to Final Records[a] (N = 443)**



Note: [a] As one observation might have multiple source databases contributing to it, the sum of these number will be larger than N = 443.

Source: Authors.

Figure 5 | **The Number of Records and the Investment Data Sources Contributing to Each Record**

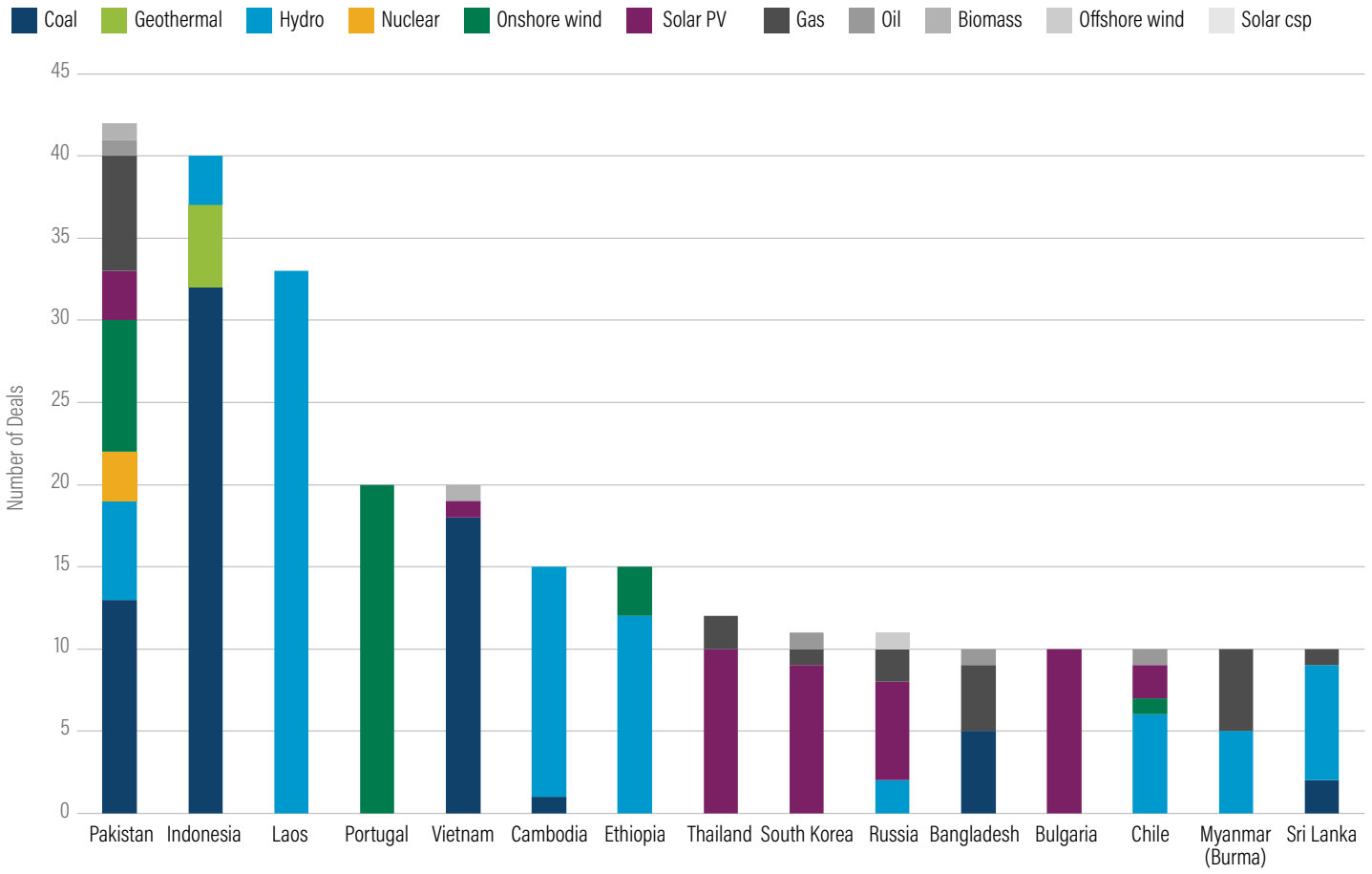


Source: Authors.

common types are projects with two contributing investment data sources. This is true for 84 of the 96 projects. Projects with three or four overlapping investment data sources are relatively rare. The variation in contribution is partly due to the non-power-generation projects in some of our source databases such as SAIS-CLA, which included more transportation projects.

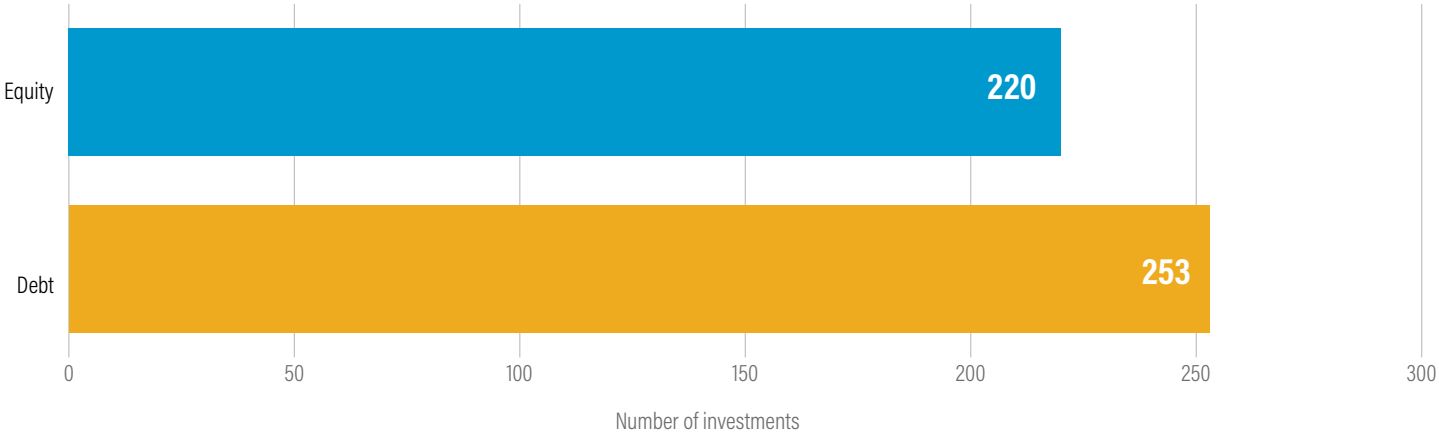
The database's 473 investments are located in 76 BRI countries (Figure 6). The top five countries—Pakistan, Indonesia, Laos, Portugal, and Vietnam—by number of investments combined received 155 investments or a third of the total number of investments recorded in the database. Fifteen countries received at least 10 investments and 269 investments combined, or 58 percent of the total investments. There are 253 debt investments and 220 equity investments in the database (Figure 7).

Figure 6 | **Number of Investments by Host Country from 2000 to 2020**



Source: Authors.

Figure 7 | **Number of Investments by Financing Type from 2000 to 2020**



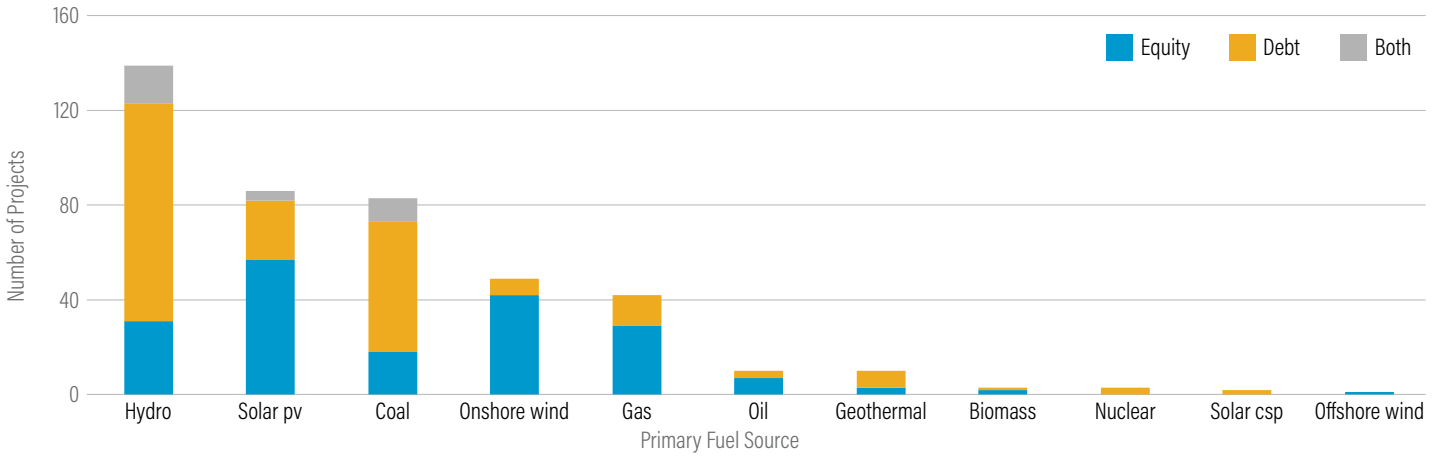
Source: Authors.

The database includes 430 power plants using 11 types of generation technology (Figure 8). There are 136 fossil fuel power plants, 293 non-fossil fuel power plants, and 1 power plant with an unknown fuel source. Hydropower, coal, solar PV, onshore wind, and gas are the most invested power plants in terms of number of plants with at least 40 power plants for each fuel type. These five types of power plants combined have 399 plants, or 93 percent of the total power plants.

Because of the missing information on equity investment amounts and capacity from source databases, the database result shows that 20 power plants are missing capacity information, and only the debt investment amount is available. We present the result on debt investment and capacity in Figure 10, but these results should be evaluated with caution, as it is not the complete picture of Chinese overseas investments in power generation.

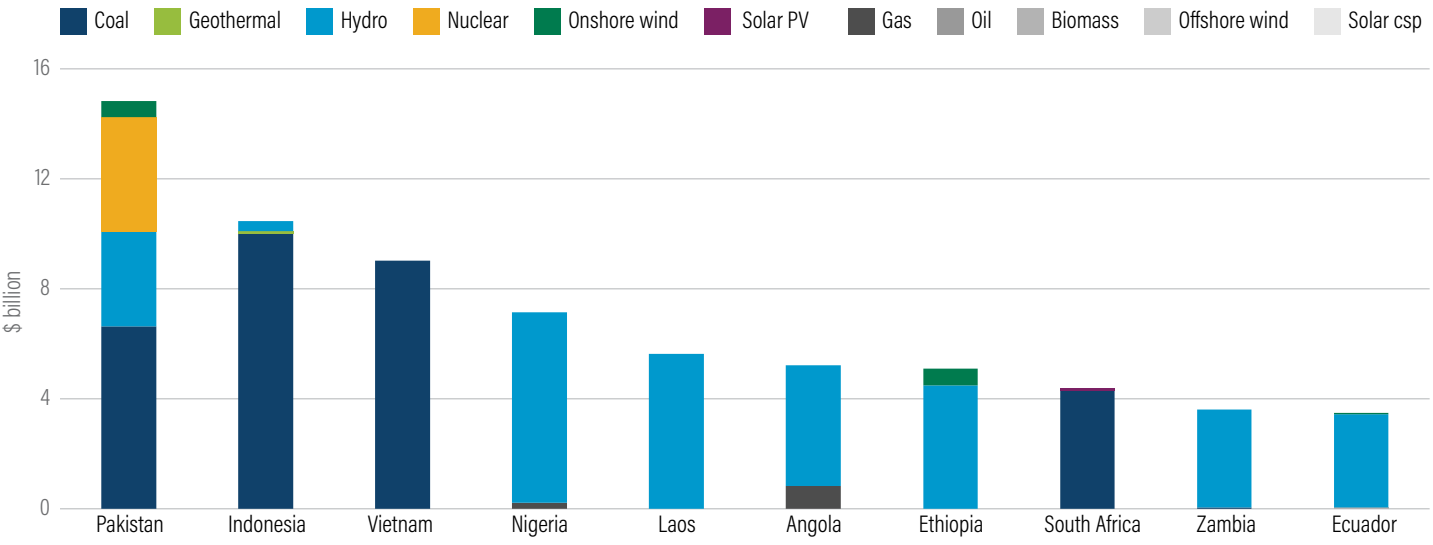
The database result shows that from 2000 to 2020, Chinese investors made in total $106.35 billion debt investments in BRI countries, with most investments in Pakistan, Indonesia, Vietnam, Nigeria, and Laos (Figure 9). In terms of capacity by primary fuel, the most invested capacity by China is coal but the largest amount of debt investment goes into hydropower sector (Figure 10).

Figure 8 | **Number of Power Plants by Fuel Source and Investment Type from 2000 to 2020**



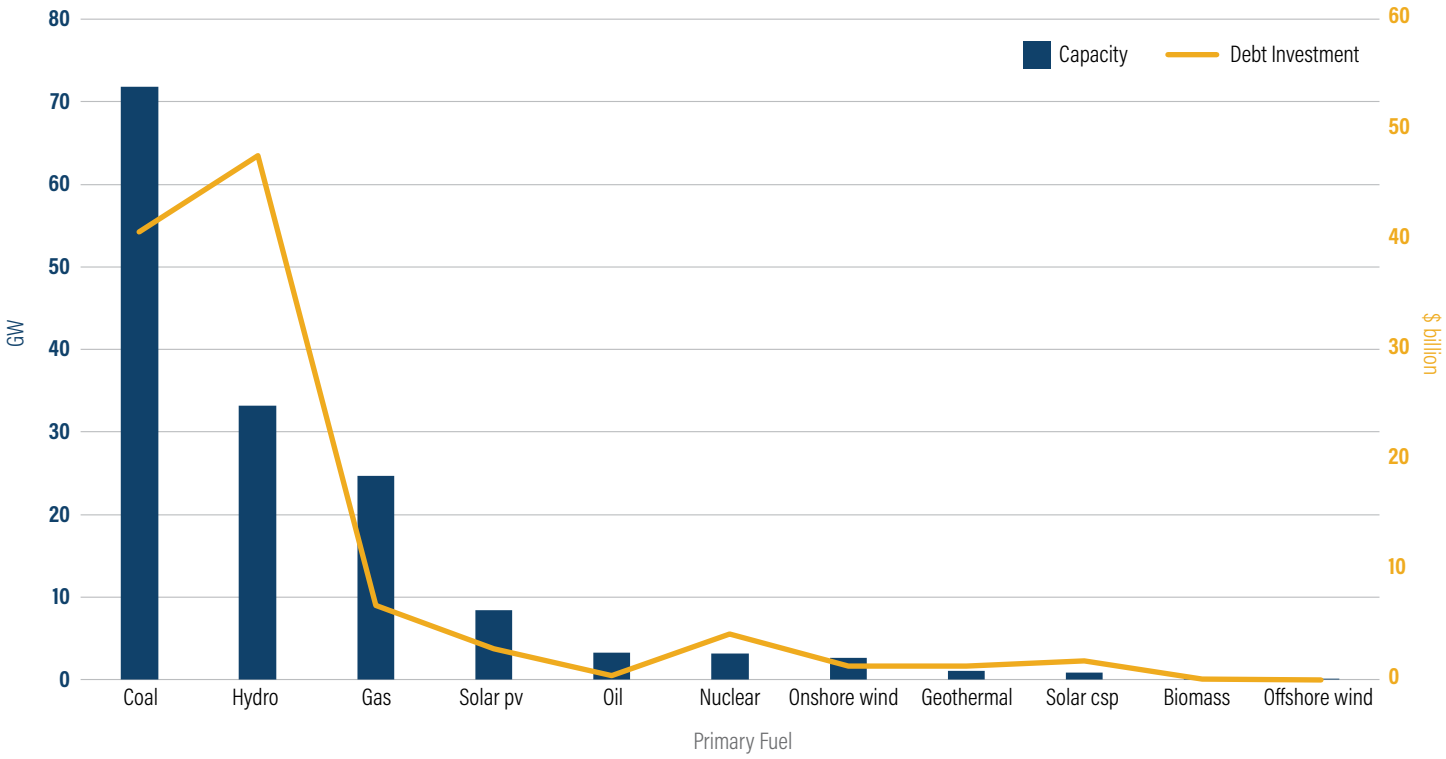
Source: Authors.

Figure 9 | **Top 10 Chinese Debt Investment Destinations by Fuel Source, 2000–2020**



Source: Authors.

Figure 10 | **Capacity Invested by Primary Fuel and Debt Investment Amount by Primary Fuel 2000–2020**



Source: Authors.

DATA USAGE AND MAINTENANCE PLAN

Various types of research can be conducted based on or inspired by the database. Analysis of Chinese overseas coal reserves and debt stress would help inform investors about how to manage their coal stranded assets risks. Trend analysis in oil and gas investment, as well as renewables in different countries, would be valuable to identify needs and gaps in the energy development pathways in BRI countries. The database can also encourage transparency of data across the BRI countries.

We will update the database annually. The annual update frequency aligns with several source databases. The annual update will include new projects for each newly added year and corrections for existing projects in previous years.

Four of the nine source databases are non-publicly available (commercial) databases, including WEPP, fDi Markets, Refinitiv-M&A, and Refinitiv-Loan. We have obtained the right to disclose investment information to the public from fDi Markets, Refinitiv-M&A, and Refinitiv-Loan. The copyright restriction of WEPP does not allow us to share design information for power plants, but users can use unique identifiers in this database to cross-reference records in WEPP. Some of the databases require monthly or annual subscriptions, and updates can be severely limited if subscriptions expire.

We are considering expanding and improving the database to include additional sectors and new databases and to develop methods to address missing values, such as improving the auto-join algorithms and exploring other source databases.

APPENDIX A. BRI COUNTRIES AS OF JANUARY 2021

AFRICA	ASIA	EUROPE	OCEANIA	NORTH AMERICA	SOUTH AMERICA
Sudan	South Korea	Cyprus	New Zealand	Costa Rica	Chile
South Africa	Mongolia	Russia	Papua New Guinea	Panama	Guyana
Senegal	Singapore	Austria	Samoa	El Salvador	Bolivia
Sierra Leone	East Timor	Greece	Niue	Dominica	Uruguay
Cote d'Ivoire	Malaysia	Poland	Fiji	Trinidad and Tobago	Venezuela
Somalia	Myanmar	Serbia	Federated States of Micronesia	Antigua and Barbuda	Suriname
Cameroon	Cambodia	Czech Republic	Cook Islands	Dominica	Ecuador
South Sudan	Vietnam	Bulgaria	Tonga	Grenada	Peru
Seychelles	Laos	Slovakia	Vanuatu	Barbados	
Guinea	Brunei	Albania	Solomon Islands	Cuba	
Ghana	Pakistan	Croatia	Kiribati	Jamaica	
Zambia	Sri Lanka	Bosnia and Herzegovina			
Mozambique	Bangladesh	Montenegro			
Gabon	Nepal	Estonia			
Namibia	Maldives	Lithuania			
Mauritania	UAE	Slovenia			
Angola	Kuwait	Hungary			
Djibouti	Turkey	North Macedonia (formerly Macedonia)			
Ethiopia	Qatar	Romania			
Kenya	Oman	Latvia			
Nigeria	Lebanon	Ukraine			
Chad	Saudi Arabia	Belarus			
Republic of Congo	Bahrain	Moldova			
Zimbabwe	Iran	Malta			
Algeria	Iraq	Portugal			
Tanzania	Afghanistan	Italy			
Burundi	Azerbaijan	Luxembourg			

AFRICA	ASIA	EUROPE	OCEANIA	NORTH AMERICA	SOUTH AMERICA
Cape Verde	Georgia				
Uganda	Armenia				
Gambia	Kazakhstan				
Togo	Kyrgyzstan				
Rwanda	Tajikistan				
Morocco	Uzbekistan				
Madagascar	Thailand				
Tunisia	Indonesia				
Libya	Philippines				
Egypt	Yemen				
Equatorial Guinea					
Liberia					
Lesotho					
Comoros					
Benin					
Mali					
Niger					
Congo (DRC)					
Botswana					

Note: The regional classification is from the Belt and Road Portal website that observes the seven-continent model.

Source: Belt and Road Portal 2021.

APPENDIX B. DETAILED VARIABLE LISTS AND DESCRIPTIONS

VARIABLE NAME	VARIABLE DESCRIPTIONS
Unique_id	The unique identification number of the entry
Power_plant_name	Name of power plant
Country	Country in which plant is located
Country_iso3c	Three letter country code in ISO 3166-1
Region	Five regions classified by the United Nations
Subregion	17 subregions classified by the United Nations
Province	State or province in which plant is located
City	City in which plant is located
Installed_capacity	Gross generating capacity of plant (MW)
Commissioning_year	Year plant entered or is scheduled to enter operation
Primary_fuel	Primary fuel used by plant
Latitude	Latitude of plant
Longitude	Longitude of plant
Duplicate_power_plant	"Y' or 'N'. "Y' means the power plant of this entry is a duplicative power plant due to multiple investments in different years for the same power plant.
Investment_type	Types of investment involved, including equity, debt, or both
Total_investment_estimated	'Y' or 'N'. 'N' means the total investment value is reported by a source in fDi Markets, not estimated. fDi Markets records this value under the variable capital investment.
Investment_cofinanced_BU	'Y', 'N', or 'NA'. This is the column, "cofinanced with unknown proportion," from the original BU dataset. "Y' means there is additional unknown portion of finance from investors other than CDB or Eximbank.
Debt_investment_averaged	'Y' or 'N'. "Y' means the loan contributions by individual banks are estimated by dividing the total loan amount by the total number of lenders. For records without information on the number of lenders, the total debt amount only captures the Chinese portion and was divided by the number of Chinese banks.
Debt_investment_weighted	'Y' or 'N'. "Y' means the one debt transaction corresponds to multiple power projects and their individual contribution is weighted by capacity
Equity_investment_weighted	'Y' or 'N'. "Y' means the one equity transaction corresponds to multiple power projects and their individual contribution is weighted by capacity
Total_investment_amount	Total investment amount of plant ($ million)
Equity_investor_name_1	Name of the first equity investor
Equity_investor_amount_1	Amount invested by the first equity investor ($ million)
Equity_investment_type	Types of equity investment type involved
Equity_investment_year	The year of equity investment made
Parent_company_of_investor	The parent company of the majority equity Investor

VARIABLE NAME	VARIABLE DESCRIPTIONS
Debt_investment_year	The year of debt investment made
Debt_investment_amount	Total debt investment of plant ($ million)
Number_of_lenders	Total number of lenders, including non-Chinese banks, only in the Refinitiv loan database. The column is left blank for loan investments from other source databases since they do not collect the total number of lenders.
Bank_of_China	The amount invested by Bank of China ($ million)
China_Development_Bank	The amount invested by CDB ($ million)
export_import_bank_of_china	The amount invested by China Eximbank ($ million)
Industrial_and_Commercial_Bank_of_China	The amount invested by Industrial and Commercial Bank of China ($ million)
China_Construction_Bank	The amount invested by China Construction Bank ($ million)
Bank_of_Communications	The amount invested by Bank of Communications ($ million)
Agricultural_Bank_of_China	The amount invested by Agriculture Bank of China ($ million)
China_Citic_Bank_Corp	The amount invested by China Citic Bank ($ million)
China_Merchants_Bank	The amount invested by China Merchants Bank ($ million)
China_Minsheng_Banking_Corp	The amount invested by China Minsheng Bank($ million)
China_Zheshang_Bank	The amount invested by China Zheshang Bank ($ million)
China_Everbright_Bank	The amount invested by China Everbright Bank ($ million)
Compid	The 'COMPID' variable in the Platts power plant database. A unique ID assigned by Platts to the parent company of the power plant
Unit_ID	The 'UNITID' variable in the Platts power plant database. A unique ID assigned by Platts to a generation unit of a power plant
Location_ID	The 'LOCATIONID' variable in the Platts power plant database. A unique ID assigned by Platts to a plant (location)
GPPD_ID	The 'gppd_idnr' variable in the Global Power Plants Database. A 10- or 12-character identifier for the power plant
J_ID	The 'Unit ID' in Platts variable in the journal database, China's Global Power_ Estimating Chinese Foreign Investment in the Electric Power Sector. Identical to unit_id
Fdi_id	A unique ID assigned by fDi Markets to an investment
RMA_ID	The 'deal number' variable in the Refinitiv M&A database. A unique ID assigned by Refinitiv to an investment
BU_ID	A unique ID generated by authors to identify each project in the BU CGEF database
SAIS_ID	The 'Loan ID' variable in the SAIS China Africa Loan database. A unique ID assigned by SAIS to a loan
IAD_ID	A unique ID generated by authors to identify each project in the BU CGEF database
R_ID	The 'Deal PermID' variable in the Refinitiv loan database. A unique ID assigned by Refinitiv to a tranche of a loan.

Source: Authors.

APPENDIX C. DETAILED AUTOMATED JOIN PLAN

There are eight steps in total for the join operation and we always join the new dataset with the previous joined results datasets. Figure 3 gives an overview of the eight steps we took in joining. Detailed tables on each step will be discussed.

Table C1 | Step 1: A_pp

- **A_unit = WEPP + Journal-CGP at unit level**
- **A_pp = Sum of A_unit at power plant level**

UNIT EQUIVALENT JOIN	UNITID== UNITID
Variables	**Aggregation Standard across Column**
installed_capacity	sum
commissioning_year	max

Since both the WEPP and Journal-CGP datasets provided information for power plants on a unit level and each had a unique identifier unitid, and the same plant-level identifier locationid, we divided the first step into two substeps to obtain the power plant-level result. We first used locationid of WEPP and Journal-CGP to equal-join on a unit level to a dataset A_unit, and then aggregated the unit to the plant level using locationid and fuel type to dataset A_pp.

Table C2 | Step 1: B

- **Aggregate GPPD to plant level**
- **B = A_pp + GPPD**

EQUIVALENT JOIN	LOCATIONID == WEPP_ID
Variables	**Aggregation Standard across Column**
latitude/longtitude	average
GPPD_id	concatenated

Since the GPPD database had a mix of plant-level and unit-level data, we aggregated the data by WEPP_ID and fuel type to the plant level. WEPP_ID in the GPPD is mapped to locationid in WEPP. Step B then equal-joined the A_pp dataset and GPPD by locationid and WEPP_ID.

Table C3 | Step 3: C = B + fDi Markets

EQUIVALENT JOIN	FUZZY JOIN	
	HIERARCHY	**TECHNIQUE**
Country, primary_fuel	equity_investor_name_n	Jaro-Winkler/cap 0.2
	commissioning_year	Absolute difference /cap 1
	installed_capacity	Absolute difference between multiples/cap 10
	city	Jaro-Winkler/ no cap
Unique Additions from Datasets	**Variables**	
B	compid, united, locationid, power_plant_name, latitude, longitude	
fDi Markets	equity_investment_year, equity_investor_amount_n, parent_company_of_investor, total_investment_amount	

Step three created dataset C by joining B and fDi Markets. The exact join variables were country and primary fuel. The fuzzy join variable with the highest importance was equity_investor_name_n. After comparing parent_investor_name, we matched it with equity_investor_name since it fit better in terms of accuracy of company names as some parent companies owned multiple subsidiaries that could create a mismatch in names.

Table C4 | **Step 4: D = C+ Refinitiv-M&A**

EQUIVALENT JOIN	FUZZY JOIN	
	HIERARCHY	TECHNIQUE
country	equity_investor_name_n	Jaro-Winkler / cap 0.2
	equity_investment_year	Absolute difference / cap 15 years
	parent_company_of_investor	
	equity_investment_type	
Unique Additions from Datasets	**Variables**	
C	commissioning_year, primary_fuel, installed capacity, city, province, compid, united, locationid, power_plant_name, latitude, longitude	
Refinitiv-M&A	equity_investment_type	

Step 4 created dataset D by joining C and Refinitiv-M&A. Since Refinitiv-M&A only included investments that are M&A, its investment type was mutually exclusive to FDI investment type (greenfield). Even though C and Refinitiv-M&A shared the same column equity_investment_amount_n, this variable only appeared in the fDi Markets dataset, and therefore the amount in dataset C only included the amount for greenfield investment. We treated equity_investment_amount_n as additional information.

The variable for equal join was country. However, a problem for the Refinitiv-M&A transaction was that it was on a company level and usually only showed a Chinese company acquiring or merging another company in a foreign country without detailed information on the types or the number of power plants owned by the company. In most of the cases, the company owns more than one power plant, with different primary fuel (wind and coal, for example). This created a problem for matching with our other datasets whose unit of observation was a power plant and lowered the probability of successful matches significantly. The problem was fixed by manual matching.

Debt Investments

Table C5 | **Step 5: E = D + BU-CGEF**

EQUIVALENT JOIN	FUZZY JOIN	
	HIERARCHY	TECHNIQUE
Country, primary_fuel	power_plant_name	Jaro-Winkler / cap 0.2
	installed_capacity	Absolute difference between multiples / cap 300 mw
Unique Additions from Datasets	**Variables**	
D	compid, united, locationid, latitude, longitude, equity_investment_type, equity_investor_name_n, equity_investment_amount_n, equity_investment_year, parent_company_of_investor	
BU-CGEF	debt_investment_year, china_development_bank, export-import_bank_of_china	

Step 5 started adding debt investments information to our database. This step mostly added new information on debt finances made by the CDB and China Eximbank and also identified any project with both equity and debt finance.

Table C6 | Step 6: F = E + SAIS-CLA

EQUIVALENT JOIN	FUZZY JOIN	
	HIERARCHY	TECHNIQUE
Country	debt_investment_year	Absolute difference / cap 15 years
	power_plant_name	Jaro-Winkler / cap 10
	installed_capacity	Absolute difference between multiples / cap 300 mw
	Debt investment amounts (china_development_bank, export-import_bank_of_china)	Percent difference /cap 10%
Unique Additions from Datasets	**Variables**	
E	compid, united, locationid, latitude, longitude, province, city, primary_fuel, equity_investment_type, equity_investor_name_n, equity_investment_amount_n, equity_investment_year, parent_company_of_investor	
SAIS-CLA	industrial_bank_of_china, china_construction_bank, bank_of_china, agriculture_bank_of_china	

Since BU-CGEF only has debt investment amount data on CDB and China Eximbank, the debt investment amount was joined separately for CDB and China Eximbank. We put the debt investment amount in the last of the hierarchy because we were uncertain about the investment amount documented. Sometimes the lenders were combined for some cofinanced projects with only a total investment amount available. In this case, we relied on power_plant_name and installed_capacity for successful joins.

Table C7 | Step 7: G = F + IAD-GEGI

EQUIVALENT JOIN	FUZZY JOIN	
	HIERARCHY	TECHNIQUE
Country	debt_investment_year	Absolute difference / cap 15 years
	Debt investment amounts (all Chinese banks)	Percent difference /cap 10%

The IAD-GEGI dataset and F dataset shared the same columns for debt investment banks, and therefore the join technique was the same as the previous step. By step 7, all the necessary columns were complete, and therefore IAD-GEGI only added novel deals and values for missing data in matched deals.

Table C8 | Step 8: H = G + Refinitiv Loan

EQUIVALENT JOIN	FUZZY JOIN	
	HIERARCHY	TECHNIQUE
Country	debt_investment_year	Absolute difference / cap 15 years
	power_plant_name	Jaro Winkler
	Debt investment amounts (all Chinese banks)	Percent difference /cap 10%

The last step joined the dataset Refinitiv Loan, which also only recorded debt investments. But the individual debt investment amount in Refinitiv Loan was mostly incomplete, and therefore we had low confidence in matching based on this variable. So, we extracted power_plant_name from descriptions in the Refinitiv Loan dataset.

After all the joining steps were completed, we finalized the results. First, we removed the rows with no investment data, and calculated the total_investment_amount for each deal. Then, we updated investment types based on their ID values. For example, if all debt source IDs were NA, the type will be equity, and if none of the IDs were NA, the type was both. Finally, we ordered the column according to our final data variable lists for final presentation.

LIST OF ABBREVIATIONS

BRI	Belt and Road Initiative
BU-CGEF	Boston University's China Global Energy Finance
CARI	China-Africa Research Initiative
China Eximbank	Export-Import Bank of China
FDI	Foreign Direct Investment
GPPD	Global Power Plant Database
IAD	Inter-American Dialogue
M&A	Mergers and Acquisitions
MoU	Memorandum of Understanding
PBoC	People's Bank of China
SAIS-CLA	School of Advanced International Studies-Chinese Loans to Africa
WEPP	World Electric Power Plants

ENDNOTES

1. We analyze greenfield FDI and M&A as separate equity investments because they represent different investment strategies and involve very different levels of risk for Chinese investors. Greenfield FDI is riskier than M&A because the former involves establishing a new entity in a foreign country.

2. Balance-sheet financing is typically a loan or a line of credit provided by financial institutions directly to a parent company, instead of a special-purpose vehicle established for an infrastructure project. Corporate bonds are bonds issued by a corporation in order to raise financing from a wide range of investors. Collateral debt obligations are a kind of asset-backed security that hold different kinds of debt, such as bank loans or corporate bonds, and sold to investors. Foreign central banks can use People's Bank of China swap lines to borrow Chinese Yuan, mainly to address short-term liquidity issues and trade settlement (Xinhua News Agency 2018).

3. Financial closure usually means the actual execution of the financing agreement for the power project, and the financing amount will be available to the project.

4. We excluded stand-alone power transmission and distribution projects but included those projects that were designed to connect a power plant to the grid, which we combined with the power plant. A captive power plant is a power plant that provides a localized power to a user, typically industrial facilities (Clarke Energy n.d.).

5. We determine a project without project specifics when two of the three pieces of information on power plant name, capacity, or location are missing.

6. Projects are thought to be duplicated if information on power-plant characteristics and investment information are identical. All duplicated projects were in the Refinitiv loan database where a project loan with multiple tranches was repeated by the number of tranches.

7. Besides information directly related to variables in the database, some additional variables helpful for organizing and tidying the original databases were also extracted, particularly from r and rm&a. We extracted information on financing type from the r database to help us better locate project finance entries; and we extracted deal type information from rm&a to better understand whether a deal acquires a company, a power plant, or multiple power plants. For both datasets, we extracted information on the energy subsector that classifies whether the project is generation, transmission, or distribution.

8. For example, nameplate capacity is the maximum rated output of a generator, prime mover, or other electric power-production equipment under specific conditions designated by the manufacturer. Net capacity is the capacity of the plant after plant parasitic loads have been considered, which are subtracted from nameplate capacity (US EIA n.d.; NREL 2013).

9. fDi Markets does not provide the equity investment amount but only includes the capital investment, which is typically the total investment cost of the project. Without additional information, such as equity ratio, it is not possible to calculate the equity investment amount.

10. The average loan contribution is dividing the total loan amount by the number of participants. In a small number of cases, one or two participants out of all participants had individual loan contributions, and the rest did not. We subtracted the known amount and divided the remaining amount evenly among all other participants.

11. This could happen if the initial investment is not enough for completing the project due to cost overruns and the project's need for additional investments to complete.

12. We created a variable in the database to document extra observations for the same power plant.

REFERENCES

Arel-Bundock, V., C.J. Yetman, and N. Enevoldsen. 2020. *Countrycode: Convert Country Names and Country Codes* (version 1.2.0). https://CRAN.R-project.org/package=countrycode. Accessed April 20, 2021.

Belt and Road Portal. 2021. "已同中国签订共建'一带一路'合作文件的国家一览 [Countries That Have Signed Co-Developing Belt and Road Collaboration documents with China]." 2021. https://www.yidaiyilu.gov.cn/gbjg/gbgk/77073.htm. Accessed March 31, 2021.

Brautigam, D., and J. Hwang. 2020. "SAIS-CARI Research Guidebook." SAIS China-Africa Research Initiative. https://static1.squarespace.com/static/5652847de4b033f56d2bdc29/t/5efd04572cbaa4329d32be1e/1593640026011/SAIS-CARI+Research+Guidebook+2020.07.01.pdf.

BRIGC. n.d. "BRIGC Overview." http://en.brigc.net/About_us/Overview/202009/t20200928_102502.html. Accessed May 10, 2021.

Byers, L., J. Friedrich, R. Hennig, A. Kressig, X. Li, L.M. Valeri, and C. McCormick. 2018. "A Global Database of Power Plants," May. https://www.wri.org/research/global-database-power-plants.

CARI (China Africa Research Initiative), Boston University Global Development, and Boston University Global Development Policy Center. 2021. (Database.) *Chinese Loans to Africa Database, Version 2.0.* 2021. https://chinaafricaloandata.org/. Accessed March 31, 2021.

Clarke Energy. n.d. "Captive Power Plant." https://www.clarke-energy.com/applications/captive-power-plants/. Accessed April 19, 2021.

Cohen, W., P. Ravikumar, and S. Fienberg. 2003. "A Comparison of String Metrics for Matching Names and Records." *Kdd Workshop on Data Cleaning and Object Consolidation* 3: 73–78.

Dentons. 2013. "A Guide to Project Finance." Dentons. https://www.dentons.com/~/media/6a199894417f4877adea73a76caac1a5.ashx. Accessed May 12, 2021.

DOE (U.S. Department of Energy). 2015. "The Five-Step Process Framework for Project Development." https://www.energy.gov/sites/default/files/2015/08/f25/5-Step%20Project%20Development%20Overview.pdf.

Financial Times. 2021. "FDi Markets." 2021. https://www.fdimarkets.com/. Accessed March 31, 2021.

Gallagher, K.P. 2021. "China's Global Energy Finance." http://www.bu.edu/cgef/#/intro. Accessed March 31, 2021.

Gallagher, K.P., and M. Myers. 2021. (Database.) *China-Latin America Finance Database*. https://www.thedialogue.org/map_list/. Accessed March 31, 2021.

Global Energy Observatory, Google, KTH Royal Institute of Technology in Stockholm, Enipedia, and World Resources Institute. 2018. (Database.) *Global Power Plant Database Version 1.2.0*. https://datasets.wri.org/dataset/globalpowerplantdatabase. Accessed March 31, 2021.

Hilman, J., and A. Tippett. 2021. "The Climate Challenge and China's Belt and Road Initiative." Council on Foreign Relations. Blog. March 31, 2021. https://www.cfr.org/blog/climate-challenge-and-chinas-belt-and-road-initiative.

IEA (International Energy Agency). 2020. "Energy Financing and Funding—World Energy Investment 2020—Analysis." IEA. https://www.iea.org/reports/world-energy-investment-2020/energy-financing-and-funding. Accessed August 30, 2021.

Li, Z., K.P. Gallagher, and D.L. Mauzerall. 2020. "China's Global Power: Estimating Chinese Foreign Direct Investment in the Electric Power Sector." *Energy Policy* 136 (January): 111056. doi:10.1016/j.enpol.2019.111056.

Mawutor, J.K.M. 2015. "Role of Project Finance in Emerging Economies." SSRN Scholarly Paper ID 2258834. Rochester, NY: Social Science Research Network. doi:10.2139/ssrn.2258834.

MFA (Ministry of Foreign Affairs). 2015. "Vision and Actions on Jointly Building Silk Road Economic Belt and 21st-Century Maritime Silk Road." https://www.fmprc.gov.cn/mfa_eng/zxxx_662805/t1249618.shtml. Accessed May 10, 2021.

NREL (National Renewable Energy Laboratory). 2013. "Solar Energy and Capacity Value." NREL. https://www.nrel.gov/docs/fy13osti/57582.pdf.

Platts. 2018. (Database.) *UDI World Electric Power Plants Database*. S&P Global. https://www.spglobal.com/platts/en/products-services/electric-power/market-data-power. Accessed March 31, 2021.

R Core Team. 2020. (Database.) *R: A Language and Environment for Statistical Computing* (version 4.0.3). Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/. Accessed April 22, 2021.

Refinitiv. 2021. *Refinitiv Eikon*. https://www.refinitiv.com/en/products/eikon-trading-software. Accessed March 31, 2021.

Robinson, D., J. Bryan, and J. Elias. 2020. (Database.) *Fuzzyjoin: Join Tables Together on Inexact Matching* (version 0.1.6). https://CRAN.R-project.org/package=fuzzyjoin. Accessed April 22, 2021.

Sato, I., B. Elliott, and C. Schumer. 2021. "What Is Carbon Lock-in and How Can We Avoid It?" https://www.wri.org/insights/carbon-lock-in-definition. Accessed August 23, 2021.

US EIA (U.S. Energy Information Administration). n.d. "Glossary—U.S. Energy Information Administration (EIA)." https://www.eia.gov/tools/glossary/index.php. Accessed April 26, 2021.

Wang, C.N. 2021. *China Belt and Road Initiative (BRI) Investment Report H1 2021—Green Belt and Road Initiative Center*. https://green-bri.org/china-belt-and-road-initiative-bri-investment-report-h1-2021/. Accessed September 10, 2021.

Xinhua News Agency. 2018. "中巴两国央行续签双边本币互换协议 [The Central Banks of China and Pakistan Renew the Bilateral Local Currency Swap Agreement]." http://www.gov.cn.qingcdn.com/xinwen/2018-05/24/content_5293454.htm.

Zhou, L., S. Gilbert, Y. Wang, M.M. Cabré, and K.P. Gallagher. 2018. *Moving the Green Belt and Road Initiative: From Words to Actions*. Washington, DC: World Resources Institute. https://files.wri.org/s3fs-public/moving-green-belt-and-road-initiative-from-words-to-actions.pdf.

ACKNOWLEDGMENTS

We are pleased to acknowledge our funders, who provided grants to support the project: the Children's Investment Fund Foundation and the ClimateWorks Foundation. We are grateful for the insightful comments from our WRI colleagues: Dimitrios Mentis, Laura Malaguzzi Valeri, Logan Byers, Peishan Wang, Ye Wang, and Zhe Liu. Our sincere thanks also go to the external experts Christoph Nedopil, Cecilia Han Springer, Jiajun Xu, and Liu Xu, whose critical reviews and suggestions have been a tremendous help. We would also like to express our gratitude to Nathan Pavlovic and Cari Gostic from Sonoma Technology, without whose expertise and help on using machine learning to automate join our data, this paper would not have been possible. While the contributions from reviewers are greatly appreciated, the views presented in this paper reflect those of the authors alone. We also wish to acknowledge the support from Renee Pineda and WRI's Research and Development team for their time and effort during the publishing process; and Caroline Taylor, Romain Warnault, and Billie Kanfer for editing, design, and all the final touches to make the publication the way it is.

ABOUT THE AUTHORS

Lihuan Zhou is an Associate for Finance Center at WRI.
Contact: lihuan.zhou@wri.org

Ziyi Ma is a Research Assistant for Finance Center at WRI.
Contact: ziyi.ma@wri.org

Shuang Liu is a China Finance Senior Associate and Acting Global Director for Finance Center at WRI.
Contact: shuang.liu@wri.org

Xinyue Ma is the China Research and Project Leader at the Boston University Global Development Policy Center (GDP Center).
Contact: xma@bu.edu

Kevin Acker was a Research Manager at the China Africa Research Initiative at Johns Hopkins University's School of Advanced International Studies (SAIS).

Margaret Myers is the Director for the Asia and Latin America Program at Inter-American Dialogue.
Contact: mmyers@thedialogue.org

Dr. Deborah Brautigam is the Bernard L. Schwartz Professor of Political Economy and Director of the China Africa Research Initiative at Johns Hopkins University's School of Advanced International Studies (SAIS).
Contact: dbrautigam@jhu.edu

Kevin P. Gallagher is a Professor of Global Development Policy at Boston University's Pardee School of Global Studies and Director of the Boston University Global Development Policy Center.
Contact: kpg@bu.edu

ABOUT WRI

World Resources Institute is a global research organization that turns big ideas into action at the nexus of environment, economic opportunity, and human well-being.

Our Challenge

Natural resources are at the foundation of economic opportunity and human well-being. But today, we are depleting Earth's resources at rates that are not sustainable, endangering economies and people's lives. People depend on clean water, fertile land, healthy forests, and a stable climate. Livable cities and clean energy are essential for a sustainable planet. We must address these urgent, global challenges this decade.

Our Vision

We envision an equitable and prosperous planet driven by the wise management of natural resources. We aspire to create a world where the actions of government, business, and communities combine to eliminate poverty and sustain the natural environment for all people.

ABOUT SAIS CHINA AFRICA RESEARCH INITIATIVE

Launched in 2014, the SAIS China Africa Research Initiative (SAIS-CARI) is based at the Johns Hopkins University School of Advanced International Studies in Washington D.C. SAIS-CARI was set up to promote evidence-based understanding of the relations between China and African countries through high quality data collection, field research, conferences, and collaboration. Our mission is to promote research, conduct evidence-based analysis, foster collaboration, and train future leaders to better understand the economic and political dimensions of China Africa relations and their implications for human security and global development.

ABOUT THE INTER-AMERICAN DIALOGUE

The Inter-American Dialogue engages our network of global leaders to foster democratic governance, prosperity, and social equity in Latin America and the Caribbean. Together, we work to shape policy debate, devise solutions, and enhance cooperation within the Western Hemisphere.

ABOUT THE BOSTON UNIVERSITY GLOBAL DEVELOPMENT POLICY CENTER

The Boston University Global Development Policy Center is a policy-oriented research center working to advance financial stability, human well-being, and environmental sustainability across the globe. Learn more at bu.edu/gdp.