Data Article

Improved database of public-private partnerships from World Bank with imputed economic, institutional and conflict data.

Marta Fernández-Olmos [a,e], Jorge Fleta-Asín [a,e,*],
Talía Gómez-Aguas [b,e], Fernando Muñoz [c,e], Carlos Sáenz-Royo [d,e]

[a] *IEDIS. Departamento de Dirección y Organización de Empresas, Facultad de Economía y Empresa, Gran Vía, 2, 50005, Zaragoza, Spain*
[b] *Departamento de Dirección y Organización de Empresas, Facultad de Economía y Empresa, Gran Vía, 2, 50005, Zaragoza, Spain*
[c] *IEDIS. Departamento de Contabilidad y Finanzas, Facultad de Economía y Empresa, Gran Vía, 2, 50005, Zaragoza, Spain*
[d] *Departamento de Dirección y Organización de Empresas, Facultad de Ciencias Sociales y del Trabajo, Violante de Hungría, 23, 50009, Zaragoza, Spain*
[e] *Expert from the SIP Foundation, Zaragoza, Spain*

ARTICLE INFO

ABSTRACT

The World Bank's database on private participation in infrastructure (PPI) projects provides detailed information on these initiatives. However, the original dataset includes imputed macro-level data for the countries that is outdated, lacks assigned ISO country codes, and is not linked to other standard country-level variables necessary for proper analysis and control by territory. In the improved version of the database, 10,958 project observations from 1900 to 2021 have been supplemented with ISO 2 and 3 country codes, enabling accurate integration with other databases. Additionally, 49 new variables related to economic, institutional, and conflict data are incorporated by country and year. This enhanced database ensures that researchers can retain critical

---

* Corresponding author.
E-mail addresses: maferno@unizar.es (M. Fernández-Olmos), jorge.fleta@unizar.es (J. Fleta-Asín), 758097@unizar.es (T. Gómez-Aguas), fmunoz@unizar.es (F. Muñoz), csaenz@unizar.es (C. Sáenz-Royo).

Specifications Table

Subject	Business, Management and Decision Sciences.
Specific subject area	Public-Private Partnerships analysis for policymaking, strategy, and international business.
Type of data	Table (.dta format) Supporting materials (Codebook) .
Data collection	The database contains 10,958 projects from 130 territories registered with the World Bank from 1900 to 2021. For each country and year, data on GDP per capita, population, GDP growth rate, as well as variables related to financial development and economic openness have been included, along with institutional variables and conflict data. In this way, 51 variables are added to 10,958 observations.
Data source location	130 territories.
Data accessibility	Repository name: Zenodo Data identification number: 10.5281/zenodo.13522682 Direct URL to data: https://zenodo.org/records/14786273 **Instructions for accessing these data**: The data are openly accessible and can be downloaded directly from the provided URL. No special permissions or credentials are required for access.
Related research article	Fleta-Asín, J., Muñoz, F., & Sáenz-Royo, C. (2024). Unravelling the influence of formal and informal institutions on the duration of public concessions. Socio-Economic Planning Sciences, 101966. https://doi.org/10.1016/j.seps.2024.101966

1. Value of the Data

- The attribution to each observation of its ISO-3 and ISO-2 country codes allows data to be added in a massive way with Stata "merge" type orders or similar.
- The degree of development of the country where each project is located is considered the year in which it was carried out through GDP per capita. In the original database of the World Bank, all projects in a country are classified statically and invariably, without recognizing their specific level of development when the project was carried out.
- Macroeconomic variables necessary to control analyses for each country and year of the project, such as population, unemployment, and GDP growth in percentage terms, have been added. Institutional variables from the Governance Indicators imputed by country and year are added. Added conflict and terrorism variables by country and year from the Global Terrorism Database.
- The country variables are, already lagged one year with respect to the project start date to facilitate analysis of how contexts influence the characteristics of the projects.
- It is necessary to complete and leave this database in a repository because in World Bank updates, the number of projects per year may decrease and prevent the verification of the analyses carried out with said data.
- This dataset offers significant value to a diverse range of stakeholders, from researchers and industry professionals to policymakers and the broader public. For academics and researchers, it provides a valuable resource to fill existing knowledge gaps and test hypotheses, particularly in public-private partnerships. Its comprehensive nature makes it a crucial tool for advancing research, improving understanding of key phenomena, and supporting new studies. In addition, this data can inform the development of new theories and practical applications, enhancing the overall body of knowledge in public-private partnerships.

- For practitioners and policymakers, the data offers practical insights that can guide decision-making and improve strategies in different public procurement sectors. By integrating these insights, professionals can refine practices, optimize resource allocation, and develop more effective solutions to current challenges. Furthermore, the dataset supports evidence-based policy development, enabling governments and organizations to make informed decisions that address societal issues such as development and poverty reduction.

2. Background

The analysis of public-private partnerships (PPPs) is of great interest due to the significant social impact and welfare they generate in the regions where they are implemented. As a result, the World Bank has been registering various types of projects and their characteristics across a wide number of countries since 1900. Numerous studies have utilized this data to explore factors influencing project duration [1], success [2,3], investment volume [4–7], private sector participation [8–10], and project type [11]. However, the current version of the database poses limitations for conducting these analyses among others.

To enable comprehensive analyses, it is necessary to incorporate control variables at the country level, as well as contextual variables [1–12]. Moreover, all previous versions of the database classify each project within a country using a static and qualitative measure of economic development without accounting for the specific level of development in each corresponding country-year. Additionally, when analysing project characteristics, the contextual variables affecting these projects should be lagged by at least one period to account for their potential impact.

3. Data Description

There are two files, a Codebook in Word format and a World Bank database to which 45 variables have been incorporated as follows:

- The database of 10,958 projects from the database originating from the World Bank in Stata dta format. It includes different project variables whose glossary can be consulted on its website [13].
- To the original base have been added GDP in millions of current US dollars, GDP per capita in current US dollars, GDP growth in percentage terms, total population in thousands, and the percentage of unemployment according to a source other than the World Bank, for each country and year. Additionally, economic openness variables such as exports and imports of goods and services, foreign direct investment inflows, and domestic credit to the private sector (all as percentages of GDP or in current US dollars) have also been included. All these variables are lagged one year with respect to the beginning of the registered project. This allows regression analysis where the contextual variables are the independent ones. In addition, it allows the country to be classified by degree of economic development - through GDP per capita - more correctly than the original base, which classifies them invariably from 1900 to 2021.
- The Governance Indicators are also imputed to the original base, for each country and year of each project [14,15] lagged by a period with respect to the execution of the project. The Governance Indicators are a set of six indicators developed by the World Bank that measure different aspects of governance in a wide number of countries [16]. These indicators provide an assessment of the quality of governance in a country across various dimensions and are widely used in economic and political studies. The six indicators are the following:
  - Control of Corruption: Evaluates the perception of the level of corruption in the public sector, including both petty corruption and state capture by elites and private interests.
  - Government Effectiveness: Measures the quality of public services, the government's ability to formulate and implement policies, and the credibility of the government's commitment to those policies.

○ Political Stability and Absence of Violence/Terrorism: Assesses the likelihood that the government will be destabilized or overthrown by unconstitutional or violent means, including terrorism and political violence.
○ Rule of Law: Reflects trust in the rules of society, including the quality of contract enforcement, property rights, police, and courts, as well as the likelihood of crime and violence.
○ Regulatory Quality: Measures the government's ability to formulate and apply policies and regulations that allow and promote the development of the private sector.
○ Voice and Accountability: Evaluates the degree to which a country's citizens participate in selecting their government, as well as freedom of expression, freedom of association, and freedom of the press.

Estimates of these indicators are presented in a range of approximately -2.5 to +2.5, where higher values indicate better performance in that governance dimension. The database provides six variables for each indicator ($6 \times 6 = 36$ variables): the value of the estimate, standard error, number of sources used for its construction, a percentile rank (0-100), a percentile rank of the upper bound of 90% confidence intervale, and a percentile rank of the lower bound of 90% confidence intervale.

- The Global Terrorism Database (GTD) is a database that compiles information on terrorist incidents around the world since 1970. It is managed by the National Consortium for the Study of Terrorism and Responses to Terrorism (START) at the University of Maryland. The GTD is one of the most comprehensive and accessible sources for studying global terrorism and is widely used in academic research, public policy, and security analysis [17]. The base records the attacks committed, so it has been necessary to previously group the attacks perpetrated in each country-year. Three terrorism variables are added:
  ○ Number of terrorist attacks by country and year.
  ○ Number of deaths from terrorist attacks per country-year.
  ○ Number of people injured by terrorist acts per country-year.
- Finally, the battle-related deaths (number of people) by country and year are added, being lagged one year regarding the beginning of the project, from the World Bank [18].

The joint database is found in [19] and the descriptive statistics of each variable added to the original one are shown in Table 1, with its corresponding label that appears in the Codebook: Tables 2 and 3.

All these variables appear in the Codebook in Word format (Codebook.doc). The document provides explanations and details regarding all variables included in the Stata data file. It includes coding information for each variable and further clarifies the origin of variables. When used in conjunction with the raw data, this Codebook serves as a valuable guide for understanding the dataset. Especially because the ISO-3 and ISO-2 codes are added to each observation, allowing data to be added and merged massively with Stata "merge" type orders or similar.

## 4. Experimental Design, Materials and Methods

The 10,958 projects accumulated in 130 economies can be plotted in Fig. 1, through a choropleth map using the FLEMUSA method with open source in Python [20]. The World Bank database records the characteristics of the projects that are in the Codebook that we attach with the article.

From these codes, the World Bank database (GDP, GDP per capita, GDP growth, Population, Unemployment, Credit Private, Exports, Imports, FDI Inflows, Battle-Deaths) and the Governance Indicators were used, and one year was added to the years of registration. Thus, they are subsequently merged by matching the country with its ISO3 code and its project year between both bases, using the Stata merge command. In this way, the country-year data for each project is already directly imputed with a delay of one year.

Table 1

Macroeconomic variables imputed to projects by country-year of the World Bank*

Variable	Obs.	Mean	Std. Dev.	Min.	Max.
GDP	10,931	1.73e+12	3.08e+12	3.36E+07	1.47E+13
GDPgrowth	10,923	4.666	4.822	-44.9	19.675
GDPcapita	10,931	4588.372	3642.555	88.087	22046.69
Population	10,946	4.44e+08	5.45e+08	52834	1.41E+09
Unemployment	8,603	6.998	4.540	0.14	57
CreditPrivate	9,761	60.461	42.811	0.007	182.432
Exports	10,584	3.73e+11	6.65e+11	6923497	2.73e+12
Imports	10,584	3.51e+11	6.01e+11	2.25e+07	2.56e+12
FDI_Inflows	10,897	4.26e+10	6.70e+10	-1.02e+10	2.91e+11

* GDP: Gross Domestic Product in millions of current US dollars in the country hosting the project. GDPgrowth: Annual growth of the Gross Domestic Product in percentage in the country hosting the project. GDPcapita: Gross Domestic Product per capita in current US dollars in the country hosting the project. Population: The Total population expressed in thousands of people in the country hosting the project. Unemployment: Unemployment, total (% of total labor force) in the country hosting the investment project. CreditPrivate: Domestic credit to the private sector (% of GDP) refers to financing provided by financial corporations through loans, purchases of non-equity securities, and other credits. It may include credit to public enterprises and comes from banks, insurers, pension funds, and other entities. Exports: Exports of goods and services include all transactions involving a change of ownership from residents to non-residents, covering merchandise, net exports under merchanting, nonmonetary gold, and services in current U.S. dollars. Imports: Imports of goods and services include all transactions involving a change of ownership from non-residents to residents, covering merchandise, nonmonetary gold, and services in current U.S. dollars. FDI_Inflows: refers to direct investment equity flows into the reporting economy, including equity capital, reinvested earnings, and other capital. It involves cross-border investments where a resident holds at least 10% of voting stock in a foreign enterprise. Data are in current U.S. dollars.
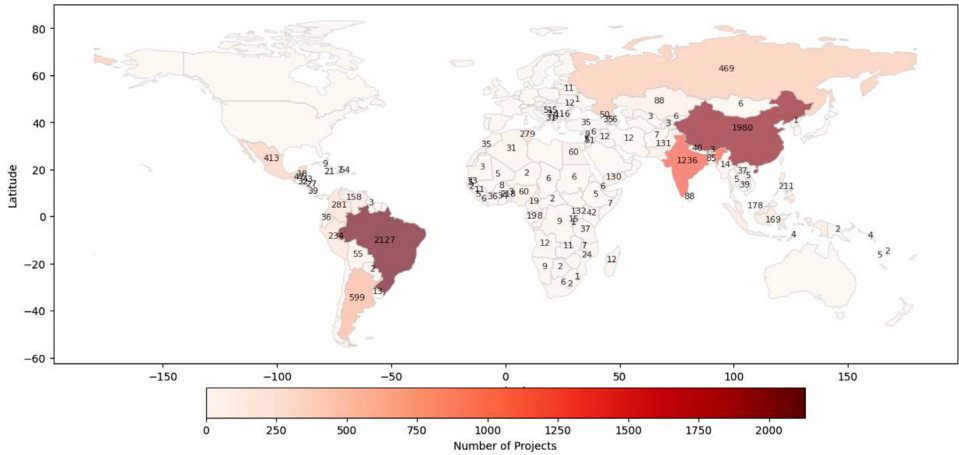
Fig. 1. Number of projects 1900-2021 (counts within each territory)

Regarding the terrorism variables, the treatment was more complex. The Global Terrorism Database records each conflict, so it is necessary to process the information, to add the number of conflicts in each country and year, the number of deaths, and the number of accumulated injuries. To do this, first, we had to sum by country using the Stata "collapse" command. Subsequently, a year was added to each variable regarding its original year, and, subsequently, a matching was carried out between both bases using the Stata merge command to link the corresponding country and year.

The descriptives of all the variables show a high number of imputed observations, being scarcer in the projects carried out around 1900, due to the lack of statistical data at that time for some countries.

Table 2
Governance Indicators imputed to the projects by country-year*.

Indicator	Variable	Obs.	Mean	Std. Dev.	Min.	Max.
Voice and Accountability (va)	vae	8,648	-0.293	0.819	-2.237	1.222
	vas	8,648	0.137	0.029	0.103	0.355
	van	8,648	13.437	3.292	2	20
	var	8,648	40.805	22.397	0	90.384
	val	8,648	34.761	21.269	0	74.876
	vau	8,648	46.924	23.619	1.877	100
Political Stability and Absence of Violence/ Terrorism (pv)	pve	8,646	-0.581	0.604	-2.860	1.320
	pvs	8,646	0.251	0.047	0.192	0.646
	pvn	8,646	7.426	1.700	1	10
	pvr	8,646	29.471	15.329	0	93.085
	pvl	8,646	19.206	11.826	0	72.857
	pvu	8,646	41.477	18.491	0.483	100
Government Effectiveness (ge)	gee	8,646	-0.085	0.373	-2.236	1.267
	ges	8,646	0.190	0.017	0.155	0.517
	gen	8,646	9.233	1.937	1	13
	ger	8,646	51.399	13.348	0	85.853
	gel	8,646	39.926	13.789	0	80
	geu	8,646	60.995	11.604	0.473	90.291
Regulatory Quality (rq)	rqe	8,646	-0.149	0.391	-2.445	1.240
	rqs	8,646	0.185	0.033	0.149	0.589
	rqn	8,646	9.659	1.755	1	14
	rqr	8,646	47.191	13.438	0	85.784
	rql	8,646	35.830	13.905	0	75.980
	rqu	8,646	57.097	12.363	1.895	96.195
Rule of Law (rl)	rle	8,648	-0.337	0.369	-2.423	1.077
	rls	8,648	0.151	0.023	0.118	0.669
	rln	8,648	14.093	2.894	1	19
	rlr	8,648	42.232	13.037	0	83.663
	rll	8,648	32.982	13.285	0	74.752
	rlu	8,648	50.258	11.989	0	93.564
Control of Corruption (cc)	cce	8,648	-0.375	0.349	-1.712	1.159
	ccs	8,648	0.147	0.028	0.111	0.498
	ccn	8,648	12.414	3.109	1	17
	ccr	8,648	42.296	14.122	0	83.495
	ccl	8,648	32.098	14.256	0	74.285
	ccu	8,648	51.199	12.948	4.040	91.747

* The last letter of each variable corresponds to: (e) estimate; (s) standard error; (n) number of sources; (r) percentile rank 0-100; (l) a percentile rank of the upper bound of 90% confidence interval; (u) a percentile rank of the lower bound of 90% confidence interval.

Table 3
Indicators of acts of terrorism and deaths due to armed conflicts imputed to the projects by country-year*.

Variable	Obs.	Mean	Std. Dev.	Min.	Max.
T1	8,559	119.671	247.463	1	3934
T2	8,559	164.163	447.958	0	14095
T3	8,559	256.216	623.153	0	16804
Battle-Deaths	3,382	808.280	2010.408	0	64657

* T1: Number of terrorist incidents by country and year where the project is hosted. T2: Sum of the number of deaths from terrorist acts by country and year where the project is hosted. T3: Sum of the number of people wounded by terrorist acts by country and year where the project is hosted. Battle-Deaths: Deaths in battle-related conflicts between warring parties, usually involving armed forces. This includes traditional battlefield fighting, guerrilla activities, and all kinds of bombardments of military units, cities, villages, etc.

Limitations

There are some limitations of the database that should be considered:

1. The World Bank is updating the database in a Stata dta file. Over time it is observed that they change the file, increasing the temporal range and therefore including more projects to the base in said added years. However, compared to older archives, the number of projects per year varies, generally decreasing. That is why it is advisable to review previous versions because they can allow robustness tests to be carried out.
2. Although macroeconomic variables are imputed by country in the places where the project is located, it must be considered that many projects are carried out in developing countries. This means that there are variables that are not available for some countries when they were run a long time ago. This is why there may be some missing observations, especially at the beginning of the time series. This can be seen in the descriptive tables of the article and precisely for this reason variables that are usually available in many countries historically are added (e.g. GDP, population, etc.).

Ethics Statement

The authors confirm that they have read and comply with the ethical requirements for publication in Data in Brief. Furthermore, they state that this work does not involve human subjects, animal experiments, or data collected from social media platforms.

Credit Author Statement

Marta Fernández-Olmos: Conceptualization, Methodology, Software, Validation, Formal Analysis, Data Curation, Writing-Original Draft, Writing Review & Editing, Visualization, Supervision, Project Administration; Funding Acquisition. **Jorge Fleta-Asín:** Conceptualization, Methodology, Software, Validation, Formal Analysis, Data Curation, Writing-Original Draft, Writing Review & Editing, Visualization, Supervision, Project Administration; Funding Acquisition. **Talía Gómez-Aguas:** Conceptualization, Methodology, Software, Validation, Formal Analysis, Data Curation, Writing-Original Draft, Writing Review & Editing, Visualization, Supervision, Project Administration; Funding Acquisition. **Fernando Muñoz:** Conceptualization, Methodology, Software, Validation, Formal Analysis, Data Curation, Writing-Original Draft, Writing Review & Editing, Visualization, Supervision, Project Administration; Funding Acquisition. **Carlos Sáenz-Royo:** Conceptualization, Methodology, Software, Validation, Formal Analysis, Data Curation, Writing-Original Draft, Writing Review & Editing, Visualization, Supervision, Project Administration; Funding Acquisition.

Data Availability

Improved database of public-private partnerships from World Bank with imputed economic, institutional and conflict data (Reference data) (ZENODO).

Acknowledgements

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] J. Fleta-Asín, F. Muñoz, C. Sáenz-Royo, Unravelling the influence of formal and informal institutions on the duration of public concessions, Socioecon. Plann. Sci. (2024) 101966.

[2] A. Jiménez, M. Russo, J.M. Kraak, G.F. Jiang, Corruption and private participation projects in Central and Eastern Europe, Manag. Int. Rev. 57 (5) (2017) 775–792.

[3] J. Fleta-Asín, F. Muñoz-Sánchez, J.M. Gimeno-Feliú, Does the past matter? Unravelling the temporal interdependencies of institutions in the success of public–private partnerships, Public Manag. Rev. (2022) 1–27.

[4] J. Fleta-Asín, F. Muñoz, "Sand" or "grease" effect? The impact of corruption on the investment volume of public–private partnerships, Int. J. Emerg. Markets 18 (5) (2023) 1168–1186.

[5] J. Fleta-Asín, F. Muñoz, Risk allocation schemes between public and private sectors in green energy projects, J. Environ. Manage 357 (2024) 120650.

[6] J. Fleta-Asín, F. Muñoz, When bigger is better: Investment volume drivers in infrastructure public-private partnership projects, Socioecon. Plann. Sci. 86 (2023) 101473.

[7] H. Wang, Y. Liu, W. Xiong, J. Song, The moderating role of governance environment on the relationship between risk allocation and private investment in PPP markets: evidence from developing countries, Int. J. Project Manag. 37 (1) (2019) 117–130.

[8] J. Fleta-Asín, F. Muñoz, How does risk transference to private partner impact on public-private partnerships' success? Empirical evidence from developing economies, Socio-Econ. Plan. Sci. 72 (2020) 100869.

[9] M. Basílio, The degree of private participation in PPPs: evidence from developing and emerging economies, in: The Emerald Handbook of Public–Private Partnerships in Developing and Emerging Economies: Perspectives on Public Policy, Entrepreneurship and Poverty, Emerald Publishing Limited, 2017, pp. 81–111.

[10] J. Fleta-Asín, F. Muñoz, Renewable energy public–private partnerships in developing countries: determinants of private investment, Sustainable Dev. 29 (4) (2021) 653–670.

[11] J. Fleta-Asín, F. Muñoz, J. Rosell-Martínez, Public-private partnerships: determinants of the type of governance structure, Public Manag. Rev. 22 (10) (2020) 1489–1514.

[12] A. Jiménez, N.C. Lupton, Terrorism hazard and infrastructure projects: the moderating role of home experience and institutions, J. Bus. Res. 135 (2021) 721–730.

[13] https://ppi.worldbank.org/en/methodology/glossary.

[14] https://www.worldbank.org/en/publication/worldwide-governance-indicators.

[15] S. Handoyo, Worldwide governance indicators: cross country data set 2012–2022, Data Brief 51 (2023) 109814.

[16] D. Kaufmann, A. Kraay, P. Zoido-Lobaton, Aggregating governance indicators, World Bank Policy Res. Working Paper 2195 (1999).

[17] START (National Consortium for the Study of Terrorism and Responses to Terrorism). (2022). Global Terrorism Database 1970 - 2020 [data file]. https://www.start.umd.edu/gtd.

[18] https://databank.worldbank.org/metadataglossary/world-development-indicators/series/VC.BTL.DETH.

[19] https://zenodo.org/records/14786273.

[20] J. Fleta-Asín, F. Muñoz, C. Sáenz-Royo, A methodological approach for enhancing visualization of country data representation in the presence of significant spatial disparity, MethodsX. 13 (2024) 102833.