

BRI-Tagged Chinese Official Finance Dataset, Version 1.0

AidData | Global Research Institute | William & Mary

Citation

Burgess, B. and S. Custer. 2022. *BRI-Tagged Chinese Official Finance Dataset, Version 1.0*. Williamsburg, VA: AidData at William & Mary.

1. Overview

As the volume of its overseas investments have quantifiably grown over the last decade, the Belt and Road Initiative (BRI) has become synonymous in both media coverage and policy discourse with the expanding global footprint of the People's Republic of China (PRC). Yet, it is unclear whether BRI did indeed represent a fundamental shift in the PRC's investment portfolio in terms of what and whom it funded, as opposed to serving as more of a way to package or (re)package what it was already doing.

For all the public scrutiny and policy discussion swirling around the BRI, there is a surprising lack of hard data to accurately assess whether and how it has altered the PRC's overseas development finance program in substantive ways—until now. This dataset is the result of an initial application of a new systematic project tagging methodology that allows us to tag projects—irrespective of time period, geography or membership status—that are “BRI-like” in that they correspond to clearly identified thematic foci laid out in key speeches and strategic frameworks introducing the BRI to the world. It is an important step forward to being able to better understand whether and to what extent there has been a change in the PRC's revealed priorities—not only what Beijing says it wants to fund, but how this translates into actual funded projects—before and after the announcement of BRI in 2013.

This tagging schema builds upon AidData's Global Chinese Development Finance Dataset, Version 2.0, which encompasses over 13,000 overseas development projects bankrolled by the Chinese government between 2000 and 2017, and is best considered as a value addition to that foundational dataset (Custer et al., 2021; Dreher

et al., forthcoming).¹ This is the first comprehensive approach to systematically define a taxonomy BRI-like activities, apply this tagging schema at the project level to AidData’s historical database, and use the results to monitor the presence of activities pertaining to BRI themes.² For more information on how we developed the BRI-tagging schema, please refer to our full methodology document “BRI-Project Tagging Methodology to Monitor the Historical Focus of Chinese Development Projects (Version 1.0)” which outlines the taxonomy development and coding process in depth.³

This project was made possible with funding from a buy-in to a USAID/DDI/ITR Higher Education Solutions Network (HESN) cooperative agreement (AID-A-12-00096). This ReadMe, along with the supporting methodology and companion dataset, are available for download via aiddata.org. For an early application of this data, see AidData’s “Delivering the Belt and Road” (Horigoshi et al., 2022) publication released in October 2022. The report incorporates the BRI-tagged data, along with other information, to assess whether and how Beijing’s overseas development program has changed with the advent of BRI in terms of revealed funding priorities, along with the expectations and experiences of partner countries in the Global South.

¹ For more information see:

<https://www.aiddata.org/data/aiddatas-global-chinese-development-finance-dataset-version-2-0>.

² Not all projects from the original Global Chinese Development Finance Dataset have been BRI-coded. The AidData team only applied the codes to projects recommended for aggregates. Projects that were not suitable for aggregates were primarily umbrella projects or other descriptions that would have skewed any higher-level quantitative analysis of PRC finance, so this exclusion should not negatively impact most researchers. For more information on this distinction between umbrella and non-umbrella projects, please refer to the supporting ReadMe and methodology for the original dataset available at: <https://www.aiddata.org/data/aiddatas-global-chinese-development-finance-dataset-version-2-0>.

³ We also want to acknowledge helpful input received from Thai Binh-Elston, Ana Horigoshi, Brook Lautenslager, Ammar Malik, Justin Schon, Kyra Solomon, Katherine Walsh, and Sheng Zhang along the way to refine the BRI-tagging methodology approach.

Table 1. Files Included in the BRI-Tagged Chinese Official Finance (Version 1.0) Data Release

File Name	Description
BRI_Tagged_Chinese_OF_FullDataset_v1.xlsx	Excel spreadsheet containing the full set of 10,849 BRI-tagged projects from AidData's Global Chinese Development Finance Dataset, Version 2.0.
methodology_BRI_tagged_projects_version_1.0.pdf	Methodology document that provides full documentation of the BRI-tagging approach.
readme_BRI_tagged_projects_version_1.0.pdf	The accompanying readme file (or codebook) describing the parameters of the dataset, caveats, and suggested uses.

2. Methods Overview, Data Dictionary, and Data Coverage

In designing and applying the BRI-like tagging schema, we used a semi-automated approach (blending human- and machine-coding) to balance precision and scalability. We used a 10-step process that can be broadly categorized within three stages: (i) conceptualization and identification of the key thematic foci of BRI, as articulated in official documents and speeches by senior Chinese government leaders; (ii) developing a standardized taxonomy to cross-walk the types of projects, by activity, captured in our historical database of Chinese state-financed overseas development project to the thematic foci of BRI; and (iii) assigning a code to each project to capture which areas of BRI activity it encompasses (if applicable). Please refer to our supporting methodology documentation for more information.

Data Coverage

- Scope: Chinese Official Development Finance
- Year Coverage: 2010-2017.
- Geographical Coverage: Global coverage.

Table 1. BRI-Tag Field Definitions

Field Name	Definition ⁴
BRI_1 Policy Communication	Binary indicator for projects identified as related to BRI thematic area #1: Strengthening Policy Communication. Xi Jinping's description of this policy area was: "Countries in the region can communicate with each other on economic development strategies, and make plans and measures for regional cooperation through consultations."
BRI_2 Infrastructure Connectivity	Binary indicator for projects identified as related to BRI thematic area #2: Improving Infrastructure Connectivity. Xi Jinping's original description of this policy area was: "To open up the transportation channel from the Pacific to the Baltic Sea and to gradually form a transportation network that connects East Asia, West Asia, and South Asia." This includes ports, roads, railways as well as power plants, telecommunications networks.
BRI_3 Trade	Binary indicator for projects identified as related to BRI thematic area #3: Promoting Trade Facilitation. Xi Jinping's original description of this policy area was: "All the parties should discuss the issues concerning trade and investment facilitation and make appropriate arrangements."
BRI_4 Monetary Circulation	Binary indicator for projects identified as related to BRI thematic area #4: Enhancing Monetary Circulation. Xi Jinping's original description of this policy area was: "All the parties should promote the realization of exchange and settlement of local currency, increase the ability to fend off financial risks and make the region more economically competitive in the world."
BRI_5 People to People	Binary indicator for projects identified as related to BRI thematic area #5: People-to-People Exchanges. Xi Jinping's original description of this policy area was: "All the parties should strengthen the friendly exchanges between their peoples to promote understanding and friendship with each other."
BRI-Like	Binary indicator for projects that included at least one of the five BRI thematic areas.
Non-BRI	Binary indicator for projects that were not within the scope of BRI activities.
Vague-BRI	Binary indicator for projects that include a blend of BRI and non-BRI activities and research assistants are unable to identify which activities are the primary focus of the project. This may also be applied if there is an

⁴ Note: This includes descriptions from Xi Jinping's 2013 speech at Kazakhstan's Nazarbayev University announcing the Silk Road Economic Belt.
https://web.archive.org/web/20201017031834/https://www.fmprc.gov.cn/mfa_eng/topics_665678/xjpfwzysiesgjtftshzzfh_665686/t1076334.shtml

	activity that appears connected to BRI, but the documentation does not provide concrete enough details to move forward with a BRI tag.
--	--

Note: This table only covers the new BRI-Tag related fields created in this process (Columns 71-78 / BS-BZ in the accompanying Excel dataset). The full list of field definitions for the dataset can be found in AidData's Global Chinese Development Finance Dataset, Version 2.0's 'Definitions' tab.

3. Data Use Recommendations and Limitations

With the addition of this BRI-like tagging schema to AidData's Global Chinese Development Finance Dataset, we hope that policymakers, practitioners, journalists, and scholars around the world will be able to use this data to answer a whole new set of critical questions. For member countries, what are the actual benefits of joining the BRI in differential levels of financing, types of projects, and terms? For non-member countries, what are the drawbacks or penalties for being a hold-out? For great powers that seek to compete with, or counter, the PRC's growing global influence, does BRI represent something fundamentally new to outmaneuver or more of what has come before, but with shinier packaging? How should Beijing assess its success in aligning its rhetoric with how the PRC manages its extensive development finance portfolio?

For projects that include a blend of BRI and non-BRI activities and research assistants were unable to identify which activities are the primary focus of the project, we include a "Vague-BRI like" tag. This may also be applied if there is an activity that appears connected to BRI, but the documentation does not provide concrete enough details to move forward with a BRI tag. This indicator allows researchers to "round up" to a more inclusive definition of BRI when computing aggregates, or "round down" to stay conservative and only work with the projects with a clear link to BRI in the project descriptions.

Thematic area #4, Enhancing Monetary Circulation, was one of the most challenging to work with in our BRI-tagging process. This is due to the frequently opaque nature of activities like currency swaps, which were not recorded in AidData's Global Chinese Development Finance Dataset, Version 2.0. Therefore, this particular sub-indicator likely underreports the value of this pillar of BRI finance and its share of total BRI funding. AidData is currently piloting a methodology to capture official Chinese

currency swaps, which will be included in a future release of the Global Chinese Development Finance Dataset. Further research on BRI issues may choose to include these data when assessing the full impact of the BRI.

We made an intentional decision to base our BRI-like project taxonomy and methodology on the Chinese government's documentation related to the announcement of BRI in 2013 and the initial years of implementation through 2017 (i.e., BRI-1.0) that best corresponds to the time period covered by AidData's Global Chinese Development Finance database (version 2.0) which includes projects from 2000-2017. We recognize that in doing so, this methodology may not entirely address subsequent adjustments, such as 2018's "BRI Recalibration" or how BRI may evolve in future in light of President Xi's announcement of China's proposed Global Development Initiative at the U.N. General Assembly in 2021. In future years, as data becomes available, we may revisit this methodology to develop additional BRI-2.0 or BRI-3.0 tags, but that is presently beyond the scope of this first pilot effort.

Based upon our preliminary assessment, the 2018 recalibration and even the more recent 2021 announcement, do not appear to be a full restructuring of priorities; but we anticipate a greater emphasis in future on the "software" of development (e.g., human knowledge, institutional capacity) in response to scrutiny of the "hard" infrastructure investments,⁵ suggesting that these categories will still apply to projects moving forward.

IV. Licensing and Fair Use

Licensing Agreement

DATA LICENSE AND ATTRIBUTION

ODC Attribution License (ODC-By)

⁵ <https://www.nbr.org/publication/a-guide-to-the-belt-and-road-initiative/>

Preamble

The Open Data Commons Attribution License is a license agreement intended to allow users to freely share, modify, and use this Database subject only to the attribution requirements set out in Section 4.

Databases can contain a wide variety of types of content (images, audiovisual material, and sounds all in the same database, for example), and so this license only governs the rights over the Database, and not the contents of the Database individually. Licensors may therefore wish to use this license together with another license for the contents.

Sometimes the contents of a database, or the database itself, can be covered by other rights not addressed here (such as private contracts, trademark over the name, or privacy rights / data protection rights over information in the contents), and so you are advised that you may have to consult other documents or clear other rights before doing activities not covered by this License.

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

1.0 Definitions of Capitalized Words

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

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

"Contents" – The contents of this Database, which includes the information, independent works, or other material collected into the Database. For example, the

contents of the Database could be factual data or works such as images, audiovisual material, text, or sounds.

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

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

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

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

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

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

"Licensor" – Means the Person that offers the Database under the terms of this License. "Person" – Means a natural or legal person or a body of persons corporate or incorporate.

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

other query) from this Database, a Derivative Database, or this Database as part of a Collective Database.

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

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

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

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

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

2.0 What this License covers

2.1. Legal effect of this document. This License is: a. A license of applicable copyright and neighboring rights; b. A license of the Database Right; and c. An agreement in contract between You and the Licensor.

2.2 Legal rights covered. This License covers the legal rights in the Database, including: a. Copyright. Any copyright or neighboring rights in the Database. The

copyright licensed includes any individual elements of the Database, but does not cover the copyright over the Contents independent of this Database. See Section 2.4 for details. Copyright law varies between jurisdictions, but is likely to cover: the Database model or schema, which is the structure, arrangement, and organization of the Database, and can also include the Database tables and table indexes; the data entry and output sheets; and the Field names of Contents stored in the Database; b. Database Rights. Database Rights only extend to the Extraction and Re-utilization of the whole or a Substantial part of the Contents. Database Rights can apply even when there is no copyright over the Database. Database Rights can also apply when the Contents are removed from the Database and are selected and arranged in a way that would not infringe any applicable copyright; and c. Contract. This is an agreement between You and the Licensor for access to the Database. In return you agree to certain conditions of use on this access as outlined in this License.

2.3 Rights not covered. a. This License does not apply to computer programs used in the making or operation of the Database; b. This License does not cover any patents over the Contents or the Database; and c. This License does not cover any trademarks associated with the Database.

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

3.0 Rights granted

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

exercised in all media and formats whether now known or created in the future. The rights granted cover, for example: a. Extraction and Re-utilisation of the whole or a Substantial part of the Contents; b. Creation of Derivative Databases; c. Creation of Collective Databases; d. Creation of temporary or permanent reproductions by any means and in any form, in whole or in part, including of any Derivative Databases or as a part of Collective Databases; and e. Distribution, communication, display, lending, making available, or performance to the public by any means and in any form, in whole or in part, including of any Derivative Database or as a part of Collective Databases.

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

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

4.0 Conditions of Use

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

4.2 Notices. If You Publicly Convey this Database, any Derivative Database, or the Database as part of a Collective Database, then You must: a. Do so only under the terms of this License; b. Include a copy of this License or its Uniform Resource Identifier (URI) with the Database or Derivative Database, including both in the Database or Derivative Database and in any relevant documentation; c. Keep intact any copyright or Database Right notices and notices that refer to this License; and d. If it is not possible to put the required notices in a particular file due to its structure, then You must include the notices in a location (such as a relevant directory) where users would be likely to look for it.

4.3 Notice for using output (Contents). Creating and Using a Produced Work does not require the notice in Section 4.2. However, if you Publicly Use a Produced Work, You must include a notice associated with the Produced Work reasonably calculated to make any Person that uses, views, accesses, interacts with, or is otherwise exposed to the Produced Work aware that Content was obtained from the Database, Derivative Database, or the Database as part of a Collective Database, and that it is available under this License. Please use this notice: AidData. 2022. China Global Public Diplomacy Indicators Dataset, Version 1.6.

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

5.0 Moral rights

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

jurisdictions allowing waiver of moral rights, Licensor waives all moral rights that Licensor may have in the Database to the fullest extent possible by the law of the relevant jurisdiction under Section 10.4; b. If waiver of moral rights under Section 5.1 a in the relevant jurisdiction is not possible, Licensor agrees not to assert any moral rights over the Database and waives all claims in moral rights to the fullest extent possible by the law of the relevant jurisdiction under Section 10.4; and c. For jurisdictions not allowing waiver or an agreement not to assert moral rights under Section 5.1 a and b, the author may retain their moral rights over certain aspects of the Database. Please note that some jurisdictions do not allow for the waiver of moral rights, and so moral rights may still subsist over the Database in some jurisdictions.

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

6.1 This License does not affect any rights that You or anyone else may independently have under any applicable law to make any use of this Database, including without limitation: a. Exceptions to the Database Right include: Extraction of Contents from non-electronic Databases for private purposes, Extraction for purposes of illustration for teaching or scientific research, and Extraction or Re-utilization for public security or an administrative or judicial procedure. b. Fair dealing, fair use, or any other legally recognized limitation or exception to infringement of copyright or other applicable laws.

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

7.0 Warranties and Disclaimer

7.1 The Database is licensed by the Licensor "as is" and without any warranty of any kind, either express, implied, or arising by statute, custom, course of dealing, or trade usage. Licensor specifically disclaims any and all implied warranties or conditions of

title, non-infringement, accuracy or completeness, the presence or absence of errors, fitness for a particular purpose, merchantability, or otherwise. Some jurisdictions do not allow the exclusion of implied warranties, so this exclusion may not apply to You. ###

8.0 Limitation of liability

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

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

9.0 Termination of Your rights under this License

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

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

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

9.4 Reinstatement of rights. If you cease any breach of the terms and conditions of this License, then your full rights under this License will be reinstated: a. Provisionally and subject to permanent termination until the 60th day after cessation of breach; b. Permanently on the 60th day after cessation of breach unless otherwise reasonably notified by the Licensor; or c. Permanently if reasonably notified by the Licensor of the violation, this is the first time You have received notice of violation of this License from the Licensor, and You cure the violation prior to 30 days after your receipt of the notice.

9.5 Notwithstanding the above, Licensor reserves the right to release the Database under different license terms or to stop distributing or making available the Database. Releasing the Database under different license terms or stopping the distribution of the Database will not withdraw this License (or any other license that has been, or is required to be, granted under the terms of this License), and this License will continue in full force and effect unless terminated as stated above.

10.0 General

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

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

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

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