AidData Methodology: Tracking Underreported Financial Flows (TUFF)

Version 1.3

Austin M. Strange, Mengfan Cheng, Brooke Russell, Siddhartha Ghose, and Bradley Parks

October 2017

Special Thanks

The authors are indebted to all the researchers at the College of William and Mary and the National University of Singapore who have contributed to this project and earlier versions.

College of William and Mary

Melanie Aguilar-Rojas, Omar Alkhoja, Katherine Armstrong, Isabelle Baucum, Zach Baxter, Ellie Bentley, Liliana Besosa, Abegail Bilenkin, Allison Bowers, Ariel Cadby-Spicer, Emma Cahoon, Bree Cattelino, Alex Chadwick, Ava Chafin, Anissa Chams-Eddine, Tina Chang, Harrison Chapman, Yining Chen, Yuning Chen, Zihao Chen, Michelle Cheng, Tiffanie Choi, Miranda Clarke, Kate Connors, Graeme Cranston-Cuebas, Catherine Crowley, Hali Czosnek, Jenna Davis, Alex DeGala, Hannah Dempsey, Weiwei Du, Ashton Ebert, Caleb Ebert, Aili Espigh, Claire Etheridge, Ze Fu, Wesley Garner, Melanie Gilbert, Elizabeth Goldemen, Zijie Gong, Grace Grains, Liz Hall, Thompson Hangen, Sarah Harmon, Ethan Harrison, Collin Henson, Jasmine Herndon, Elizabeth Herrity, Gabrielle Hibbert, Carlos Holden-Villars, Keith Holleran, Weijue Huang, Daniel Hughes, Torey Beth Jackson, Jiaorui Jiang, Emmaleah Jones, Amar Kakirde, Naixin Kang, Ciera Killen, Ian Kirkwood, Emily Koerner, Dylan Kolhoff, Lidia Kovacevic, Martyna Kowalczyk, Mirian Kreykes, Dinu Krishnamoorthi, Isabella Kron, Karthik Kumarappan, Daniel Lantz, Caroline Lebegue, Jade Li, Yuwei Li, Yaseen Lofti, Adriane Lopez, Flynn Madden, Sarah Martin, George Moss, Marie Mullins, Qiuyan Ni, Jack Nicol, Brendan O'Connor, Alexandra Pancake, Carol Peng, Grace Perkins, Sophia Perrotti, Andrea Powers, Han Qiao, Kamran Rahman, Sarah Reso, David Rice, Natalie Santos, Faith Savaiano, Dominic Scerbo, Leigh Seitz, William Shangraw, Kaitlan Shaub, Samuel Siewers, Kyra Solomon, Yifan Su, Elizabeth Sutterlin, Mahathi Tadikonda, Joanna Tan, Rebecca Thorpe, Jessica Usjanauskas, Emily Walker, Yale Waller, Katherine Walsh, Xinyi Wang, Jason (Jiacheng) Xi, Hanyang Xu, Darice Xue, Erya Yang, Gaohang Yao, Antonio Tianze Ye, Lincoln Zaleski, Jack Zhang, Yue Zhang, Echo Zhong, Joana Zhu, and Junrong Zhu.

National University of Singapore

Silvia Tieri, Clara Lee Yen Yin, Hanissa Bte Sa'Ad, Nasuha Selamat, Isis Bte Iskandar, Marissa Ng, Marissa Foo, Ling Yu Huang, H. Barath, Wong Zhen Jie, Isaac Christian, Mah Jia Wei, and Shameer Yaqin B Muhammad Noor.

Others

Andreas Fuchs and Axel Dreher were also instrumental in helping us define flow type and flow class categorization methods. We would like to thank all of the participants in the September 2017 "Tracking International Aid and Investment from Developing and Emerging Economies" workshop at Heidelberg University for their constructive feedback and Steve Lewis-Workman for their careful scrutiny of AidData's Global Chinese Official Finance Dataset, Version 1.01. We are also very grateful to Robert Mosolgo, a former Project Manager at AidData who created the online TUFF coding interface as well as Brook Lautenslager and Francis Bartkowiak, who currently manage the interface. We would also like to thank Brian O'Donnell, Charles Perla and Harsh Desai -- three former project managers of the TUFF team. The authors are solely responsible for any errors in this document.

Citation

Strange, Austin, Mengfan Cheng, Brooke Russell, Siddhartha Ghose, and Bradley Parks. 2017. Tracking Underreported Financial Flows (TUFF) Methodology, Version 1.3. Williamsburg, VA: AidData.

For questions on this methodology or any related datasets, please email us at china@aiddata.org.

Contents

Acronyms

TUFF   Tracking Underreported Financial Flows
OECD   Organisation for Economic Co-operation and Development
DAC   Development Assistance Committee
ODA   Official Development Assistance
OOF   Other Official Flows
IATI   International Aid Transparency Index
QC   Quality Control

1. Introduction

The global development finance landscape is changing rapidly. Whereas several large multilateral development banks and Western governments once acted as the primary sources of development finance, the "market" for external grants and loans is now characterized by a wide variety of actors with diverse interests and capabilities (Klein and Harford 2005; Brainard and Chollet 2007; Manning 2006; IDA 2008; Woods 2008; The Economist 2009; Fengler and Kharas 2010; Severino and Ray 2010). Global reporting systems have not kept pace with these changes. China, Venezuela, Brazil, India, and Iran together reportedly provide tens of billions of dollars of development finance each year (Walz and Ramachandran 2011). However, none of these sovereign governments have opted to participate in existing reporting systems, such as the OECD's Creditor Reporting System (OECD CRS), the International Aid Transparency Initiative (IATI), or AidData.[1] This makes it difficult to understand the nature, scale, distribution, and impact of the development finance supplied by these actors, fueling uncertainty and speculation about the intentions of non-Western donors.

In an attempt to address this issue, AidData created an open source data collection methodology -- called Tracking Underreported Financial Flows (TUFF) -- to collect project-level data from suppliers of official finance who do not participate in global reporting systems.

This codebook outlines the set of TUFF procedures that have been developed, tested, refined, and implemented by AidData staff and affiliated faculty at the College of William & Mary. We initially employed these methods to achieve a specific objective: documenting the known universe of officially financed Chinese projects in Africa (Strange et al. 2013, 2017). We have since then employed these methods to track Chinese official finance to five major world regions: Africa, the Middle East, Asia and the Pacific, Latin America and the Caribbean, and Central and Eastern Europe (Dreher et al. 2017). Additionally, other social scientists have adapted and applied the TUFF methodology to identify grants and loans from Gulf Cooperation Council (GCC) members (Minor et al. 2014), under-reported humanitarian assistance flows from traditional and non-traditional sources (Ghose 2017), foreign direct investment from Western and non-Western sources (Bunte et al. 2017), and pre-2000 foreign aid flows from China (Morgan and Zheng 2017). However, this codebook focuses specifically on TUFF data collection and quality assurance procedures to track Chinese official finance between 2000 and 2014.

AidData coders follow a standardized set of procedures at each stage of the data collection process. By documenting each procedure in this codebook, we seek to expose our methods to external scrutiny and thereby collect constructive feedback that can be used to inform improvements to future versions of the data collection process. Since publishing our initial project-level dataset on Chinese official finance to Africa on china.aiddata.org in April 2013, we have received substantial feedback from users, which has in turn helped us to fine-tune our methods and enhance the accuracy, precision, and usefulness of the data we generate. We intend to continue improving the methodology based on constructive criticism and input from users of the data and other interested parties.[2]

In the remainder of this methodology document, we introduce the latest iteration of the TUFF methodology in the following order. We first highlight the primary improvements made to the methodology since its last iteration (version 1.2). Second, we provide a general outline of the three stages of the TUFF methodology and invite readers to view detailed, step-by-step coder instructions for each step in the data generation process. Third, we focus on several important dimensions related to classifying underreported financial flows into meaningful categories. Fourth, we point readers to a research release version of the dataset that allows researchers to download aggregate data files based on recommended queries. Fifth and finally, we include several appendices with supplementary information about additional TUFF-based data collection and quality assurance activities.

---

[1] Russia does provide aggregate aid statistics to the OECD. However, it does not provide project-level data on outgoing Russian aid flows.

[2] Questions, comments, and feedback can be directed to the authors or china@aiddata.org.

1.1 Methodological Updates Since 1.2

Since the publication of the 1.2 version of the TUFF methodology in 2015, AidData has continuously amended and refined the data collection and management through automated processes and systematic curation. The following updates are reflected in the 1.3 version of the methodology:

1) **"Umbrella" binary marker for umbrella agreements.** One distinct feature of China's overseas development program is the common practice of signing umbrella agreements before specifying the individual projects will fall underneath these broad agreements. As a part of this practice, significant amounts of funding are usually not specified at the original date of the umbrella agreement, but rather at a later date after parties sign another, separate agreement. Since the TUFF methodology often captures the signing of both the umbrella agreement and subsequent funding activities as commitments, we have introduced an "umbrella" marker to help avoid duplication and overcounting. Users ultimately need to determine whether it makes sense to include or exclude umbrella projects based on their specific questions. However, we have found that for most research projects and research questions it makes most sense to exclude umbrella projects. There are 399 umbrella projects in the latest research release (AidData's Global Chinese Official Finance Dataset, Version 1.0) representing official commitments or pledges of around $353.0 billion from 2000 to 2014.

2) **New "recommended_for_research" field.** This field is marked as "true" based on whether the following criteria are met: (a) Umbrella = 0; and (b) Status = commitment, implementation, or completed. In most, though not all research applications, dataset users wish to aggregate project-level data to aggregate estimates of "official commitments."[3] To approximate official commitments, AidData has always advised users of its datasets to only include projects that have reached the official commitment stage, projects in implementation, and completed projects. The logic of this summation procedure is that if a project has reached the implementation or completion stage, then it must have previously reached the official commitment stage. AidData has also always recommended to users who wish to estimate official commitments that they exclude all pledges that did not meet the official commitment stage, canceled projects, suspended projects and "umbrella" projects. However, in some cases, users of previous dataset iterations inadvertently included or excluded the wrong projects in their calculation of aggregate financial amounts or project counts. To help address this problem, we have introduced a "recommended_for_research" variable that allows dataset users to easily identify the subset of projects in the dataset that can be safely aggregated in estimations of "official commitments". We hope that the inclusion of this pre-set filter will significantly reduce the likelihood of inadvertent dataset misuse.

3) **Clarified flow class categorization by funding agency.** Concessional loans from Export-Import Bank of China with development intent are now systematically coded as ODA. According to the Chinese government, the concessional loans from the Export-Import Bank of China have fixed interest rates of 2-3%, maturity of 15-20 years with grace period and 2 payments per year.[4] All concessional loan projects with development intent provided by Export-Import Bank of China have thus been coded as "ODA-Like".

4) **New machine learning software.** The TUFF methodology requires searching through millions of potential sources. To simplify the data collection process and increase efficiency and accuracy, we have introduced a machine learning application (called the "TUFF Robot") during the first stage of data collection. The TUFF Robot first ingests a "training set" of past sources/search results that actually contained information about officially-financed Chinese projects (as identified by human researchers during previous rounds of TUFF data collection), which in turn allows machine learning software to "learn" what types of search results are most likely to have useful project information in them. The TUFF robot then combs through millions of search results -- at a rate of approximately 15,000 results per hour or 475,000 results per week — and categorizes search results as either "relevant" or "irrelevant" (in order to identify those projects

---

[3] Some researchers may be specifically interested in studying financial pledges, umbrella projects, or suspended and cancelled projects. For this reason, we include these records in the research release versions of our datasets.

[4] 一带一路内参: _http://caiec.mofcom.gov.cn/article/g/201610/20161001406979.shtml_. Also see 中国对外援助白皮书: _http://politics.people.com.cn/GB/1026/14450711.html_. The White Paper does not explicitly state that conditions refer to China Export-Import Bank but the conditions are the same.

that likely contain information about officially-financed Chinese projects).[5] Positive search results are then turned over to researchers to manually read through and code project information. This innovation has dramatically reduced the amount of research assistant time that is spent reviewing superfluous search results, and at the same time created less scope for human errors.

5) **Updated options for the variable "intent."** We eliminated "null", "mixed (some development)", and "mixed (no development)" categories in favor of only one "mixed" category. We did this in order to eliminate categories of limited use. Also, to simplify the coding process and minimize human discretion, we have treated "development" as the default intent and merged all the mixed intent categories into one "mixed" category.

6) **Credit line transaction amounts marked as umbrella projects.** Since many lines of credits are used to fund smaller projects, we have coded all the credit lines as umbrella projects and kept the transaction amounts so that users can choose to exclude these projects easily or look at all the credit line projects only.

7) **Expanded set of quality control (QC) procedures.** We expanded our QC procedures to include several additional rounds of data verification during a third stage of TUFF data verification and augmentation. In addition to previous QC procedures, the new methodology includes involving a wider set of researchers to help vet and stress-test the data before its official release. We also reviewed financial amounts subject to currency revaluations. In addition, the 1.3 methodology has better documented QC processes were implemented since the 1.2 version, but did not appear in the 1.2 methodology. QC (Stage Three) is executed after Stage One and Stage Two of data collection to eliminate human errors and ensure consistency across projects.

8) **Inclusion of IMF Staff Country Reports as an additional Stage One source.** Under Article IV of the IMF Articles of Agreement, the IMF conducts policy surveillance of its member states (Simmons 2000; IMF 2006: 7). More specifically, the "IMF's Executive Board conducts Article IV consultations with each member country once a year or every two years, based on staff reports that summarize recent developments and discussions with the national authorities" (IMF 2007: 8). These IMF staff reports sometimes contain detailed information about loans and other sources of official finance from China. As part of a broader effort to increase the number of official sources that are drawn upon to construct individual project records, AidData consistently extracted this project- or transaction-level information from the IMF documents during Stage One.

9) **Separation of financing and co-financing agencies from the funding_agency field.** To minimize confusion related to the use of the funding_agency field, we have separated the Chinese official financing agencies from other involved agencies (including co-financing agencies) into two separate fields. We have also eliminated the donor_agency field in the final version of the dataset.

---

[5] The machine learning software uses a linear SVM classifier (with balanced TRUE/FALSE classes and TFIDF transformation) from the scikit-learn python library classes was used to classify the documents. Classes were balanced by oversampling the TRUE class. For classifier package, please refer to: Scikit-learn: Machine Learning in Python, Pedregosa et al., JMLR 12, pp. 2825-2830, 2011.

2. TUFF Methodology - An Overview

AidData's TUFF methodology is divided into three stages including initial data collection, source triangulation, and many supplementary quality control steps that refine the data collection and organization process. Each stage is described in detail below.

2.1 TUFF Stage One: Project Identification

The objective of Stage One is to identify potential projects through extensive searching involving official sources and media reports. During Stage One, projects undertaken in a particular country and supported by a specific supplier of development finance—be it a sovereign government, multilateral institution, non-governmental organization, or private foundation—are identified through four major sources: (1) aid information management systems (AIMS) in recipient countries; (2) Chinese Embassy and Economic and Commercial Counselor websites; (3) IMF staff country reports; and (4) Factiva, a Dow Jones-owned media database that draws on approximately 33,000 media outlets worldwide in 23 languages, most of which are newspapers, radio and television transcripts.

2.2 TUFF Stage Two: Source Triangulation

The objective of Stage Two is to search for and synthesize additional sources for each project identified in Stage One. During this stage, RAs conduct tailored searches on individual project records with Google and other local search engines, such as Baidu, to confirm or disconfirm a project's existence and refine the accuracy or a record. These searches are geared to uncover critical details on the project, including financing agencies, the donor's financial contribution to the project, the status of the project, and so forth. During these searches RAs can uncover additional sources, including (a) media reports; (b) other donor or recipient government documents—loan agreements, budget documents, project documentation; (c) country briefings from the IMF, World Bank or African Development Bank; (d) information from implementing company websites; (e) multimedia evidence of project activities—photos and YouTube videos; (f) briefings or press releases for in-country NGOs; and (g) academic articles with country-level searches in Google Scholar. If any additional projects are identified from these new sources, those projects are created and thrown back to Stage One so they undergo their own Stage Two process. In addition, information gaps may also be abridged by corresponding with donor/recipient personnel.

2.3 TUFF Stage Three: Quality Control

The last and most critical stage in the TUFF methodology process is a series of rigorous, systematic quality assurance procedures. Using open source information has many potential pitfalls, and this last stage of the methodology seeks to identify and eliminate potential errors, biases, and data holes wherever possible. It includes the following procedures[6]:

1) **De-duplication:** Given the wealth of sources TUFF incorporates, often a specific project will be referenced in multiple sources and one project might be created as a record in the database several times as part of the first two stages of TUFF. In Stage Three, we use a web-based data platform with filtering and keyword search functions to identify and eliminate these duplicate records. Our researchers implement a rigorous process of filtering projects according to their recipient and sector values and identifying duplicate records along those criteria. During this process, they may encounter records that share many attributes, which may suggest that they are duplicate records. Any projects so deemed suspicious are flagged for a senior AidData arbitrator who reviews the records and deactivates any likely duplicate projects.

2) **Reducing and eliminating double counting:** In this iteration of the methodology, we have introduced the 'umbrella' field to help reduce the chance of double-counting flows from mega-deals or lines of credit. In Stage 3, senior AidData researchers review the project details and identify projects where some agreement was signed between two countries but the funds were not allocated at the original agreement date, only to be allocated through sub projects later on. These projects are marked as "umbrella" projects to mark their likely hierarchical relationship to other records in the dataset and allow users to exclude them from financial analysis. Umbrella projects are kept in the dataset because it is possible that some of the umbrella projects are not captured entirely in sub projects and some users may be interested in studying these projects specifically. As such, we have chosen to keep the information while being as transparent as possible.[7]

3) **Check Logical Inconsistency:** Many of the fields captured in the dataset require researchers to categorize flows according to several related schemas - including donor intent, flow, and flow class. Based on the definitions and schemas as well as automated decision rules described in this methodology, some combinations are not allowed. For example, a project cannot be coded as ODA-like (flow class) and Representational (intent). In Stage 3, projects that defy decision rules related to what combinations are acceptable (or other standing decision rules) are flagged for review by a senior AidData researcher to resolve the inconsistency.[8]

4) **Record-by-Record Review:** In Stage 3, *every* project in the dataset is vetted by a AidData program manager or a senior research assistant appointed by the program manager. This step is meant to ensure each project is reviewed by a trained eye to identify potential errors, missing data, or incorrect categorizations.

5) **Targeted Review:** After a record-by-record review, the dataset undergoes another review targeted towards at-risk projects (as indicated by low health of record scores)[9] and high-value projects (as indicated by high financial amounts). This review is meant to add an additional layer of scrutiny to (1) ensure no additional data can be identified, (2) financial values are accurate, and (3) variable codings are correct.

6) **Deflation & Financial Review:** To ensure the financial data is comparable across years, all financial values are calculated in US constant 2014 dollars using the AidData deflation methodology (see Appendix F). As part of this process, potential local currency changes and revaluations are identified and the currency exchange rates are adjusted accordingly.

7) **Extended Review:** Once the dataset has been vetted by the team producing the data (including the AidData program manager and his/her team of research assistants), the dataset is then reviewed by additional AidData staff and a cohort of external researchers. These reviewers vet the data using various methods, including but not limited to (1) running descriptive statistics with the dataset to identify anomalies or suspicious results; (2) comparing the dataset and the resulting financial amounts to other published estimates of Chinese Official Finance (or subsets thereof) to identify significant deviations from other estimates, including White Papers published by the Chinese Government and

---

[6] Detailed steps of quality control can be found in the Coder Instruction.

[7] We have considered linking such parent-child records as such in the dataset, but there is often not enough information to link individual projects to their parent mega-deals, credit lines, or other original commitment or pledge. As such, we have chosen to err on the side of caution and mark all umbrella projects (and their associated amounts) to allow users to exclude them from financial analysis.

[8] See Appendix E for all the automated decision rules.

[9] See Appendix D for calculation of Health of Record.

estimates published by third parties; (3) cross-referencing the finalized project list to project lists provided in various official sources (such as IMF staff reports); (4) reviewing individual project records for errors or missing data; and (5) identifying biases in the data and identify potential ways to address them.

In addition, in preparation for the release of the Global Chinese Official Finance dataset (Version 1.0), we added an additional layer of scrutiny by providing a pre-release version of the dataset to a diverse group of researchers involved in a workshop on China and other emerging donors/lenders. These researchers used the pre-release version of the dataset for the purposes of statistical analysis and provided detailed feedback on the dataset.

It is our hope that through the protocols and processes developed in Stages 1-3 of the TUFF methodology, we can minimize errors and biases and produce the most consistent, complete and replicable dataset possible. That said, our coverage of Chinese official finance, as is often the case in social science data collection, is not 100% exhaustive or without error. As such, the interactive platform available at china.aiddata.org is designed to facilitate the crowdsourcing of information and/or feedback from users. Users who have knowledge of any additional projects or information to update any existing projects have the option to "suggest a project" and "confirm", "challenge" or "comment" on any content of any project. Once program managers at AidData receive these notifications, they either verify and update the information accordingly or otherwise resolve the concern.More details on crowdsourcing protocols can be found in Appendix C.

The three stages outlined above each contain detailed steps to guide the data generation process. In the coder instruction document, we provide a step-by-step guide documenting how to track underreported financial flows along with multitude of appendices to introduced supplementary steps in the TUFF process. Please visit http://aiddata.org/pages/tracking-underreported-financial-flows-methodology for more information on our coder instructions and a list of frequently asked questions on how to apply our methodology.

3. Classifying Chinese Official Finance

The challenge of studying official finance from China and other non-OECD development financiers lies in not only uncovering the sheer volume of activity, but also classifying the flows accurately and in ways that make comparisons between different financiers valid. To make the study of financial flows from China more comparable to other traditional donors, the TUFF methodology is designed to classify Chinese official finance using the OECD's standards as set out in the OECD DAC Directives. We use these standard to classify all projects in the dataset according to their source of funding, project intent, level of concessionality and type of financing provided. The following sections will explain how these categorizations are coded.

3.1 Source of Funding (Official/Unofficial)

First, projects must first be classified as coming from an official source of funding or not. The OECD defines "Official Financing" as "transactions undertaken by the official sector (i.e. Government) at their own risk and responsibility, regardless of the source of funds (taxation of or borrowing from the private sector). Official agencies include federal, state and local departments and agencies." As such, government agencies must be extending the funding for a flow/project in order for it to be coded as official financing in our dataset.

In the case of China, it can be perplexing to define the scope of official finance because of the lackness of a centralized foreign aid management system. While the State Council or Ministry of Commerce makes instructional plans for foreign aid distribution in general, each agency/identity has its own budget and some autonomy in determining the recipients and the loan terms. It also makes it harder to draw the line between "official" and "unofficial" because of state-owned/state-backed enterprises and banks functioning on a commercial basis.[10] To make the definition of Chinese official finance more clear, the TUFF methodology tracks financial flows from the following institutions/entities:

A.   Government agencies, including ministries and Chinese embassies;
B.   Provincial governments;
C.   Policy banks, including China Development Bank and China Export-import Bank;
D.   State-owned commercial banks, including the Bank of China, the Industrial and Commercial Bank of China, the Agricultural Bank of China, and the Construction Bank of China.
E.   State-owned companies.[11]

3.2 Donor Intent

Each project must also be coded to capture the perceived intent of the finance provider. The broad categories covered by "donor intent" include: development in the recipient country (development), commercial interests in the donor country

---

[10] The official financing coding scheme attempts to distinguish between official and unofficial sources of finance, often looking at the agencies involved in a project's announcement or signing. However, individuals present at the signing ceremony may hold multiple positions within a government or positions in the government and private sector. These personal linkages and overlaps can obscure the true financial source of a project. For example, in one corporate aid project from a state-owned enterprise (http://china.aiddata.org/projects/489), the vice president of the company--who is also a secretary in the Chinese Communist Party--and the Chinese Minister of Commerce both attend the opening ceremony. That said, such conflicts of interest are certainly not a data issue exclusive to media-based data collection, and official records can encounter the same kind of challenge.

[11] Aid flows from state-owned enterprises are considered as official funding but coded as "Corporate Aid+Gov" instead of "ODA".

(commercial), the representational interests of the donor country or a donor-recipient relationship (representational), or a combination of two or more donor motivations (mixed).

The OECD advises that the decisive criterion for a project's eligibility as Official Development Assistance is a main objective to promote the recipient's "economic development and welfare", but also adds "in the final analysis it is a matter of intention."[12] Therefore, coding by donor intent provides basis to distinguish a donor's official development finance from its larger portfolio of official finance. For cases of ambiguous intent, the OECD also provides guidance on inclusion or exclusion from the ODA category, which has informed our criteria for a "development" category of donor intent.

Each coder must systematically assess the intention of a given project based on specific criteria outlined below. See Table 1 for descriptions of each intent category along with examples.

---

[12] "Is it ODA?" OECD Factsheet, November 2008. Accessed at: http://www.oecd.org/investment/aidstatistics/34086975.pdf

Table A: Donor Intent Examples

Intent Category	Description	Examples
Development	Promotes long-term economic development and welfare within the recipient country. The donor does not intend to receive a future monetary reward or profit from this assistance. Can include tied aid, as well as projects where the donor is both the funder and the implementer.	• Humanitarian assistance (including delivery by military) and emergency management • Capacity building within the recipient country to sustain social programs long-term • Institution building of recipient government through elections, training, or official government buildings
Commercial	Advances a donor's commercial, industrial, and economic interests, facilitates trade and resource transfers between the donor and recipient, or supports a capital investment with the expectation of commercial profit in the donor country.	• All foreign direct investment and joint venture activities (donor has "bought a stake" in the recipient enterprise) • Investments in extraction of natural resources (mining, oil drilling, logging, etc.) to be sent to donor country • Export credits and other commercial credits
Representational	Symbolic gesture of "good will" to advance an official relationship. The project is likely small enough it will not substantively advance recipient development, nor will it directly promote donor commercial interests. Includes diplomatic, military, and cultural promotion activities, as well as support to political parties.	• Promotion of donor culture (e.g. language training) • Military aid without a clear humanitarian/developmental purpose • One-off exchange of doctors, teachers, or other social service professionals (programs for recurring, long-term exchanges are considered *development*)
Mixed	Cannot be categorized into the development, commercial, or representational categories, because (a) the project has both commercial and development intent, or (b) the project has both representational and development intent.	• In-kind contribution in exchange for commercial benefits in the donor country, such as drilling licenses • Construction of an industrial park to contain a donor commercial operation, among other enterprises • Institutes of learning or research which include promotion of donor culture (e.g. Confucius Institutes)

3.3 Flow Type

Each project should be coded according to the type of financial transaction undertaken. Potential options for this flow class variable include grants, technical assistance, loans, debt relief, debt rescheduling, export credits, loan guarantees, scholarships, strategic/supplier credits, and other financial instruments. Usually the classification of flows is clear based on the description of the flow. Commodity donations/in-kind donations are coded as grants.

3.4 Grant Element

This variable estimates the concessionality level of a flow. It is easy to determine that all grants, free-standing technical assistance and scholarships/training are concessional. However, a loan can be both concessional and non-concessional. Concessionality of a loan is measured using "Grant Element" by OECD where any loan with a grant element of 25% or

higher is considered a concessional loan[13]. To ensure comparability with OECD data, AidData used the OECD's grant element calculator[14] to measure the concessionality of all loans we track in TUFF datasets.[15] Due to information gaps in the loan conditions of non-transparent donors, we assume the following for the loan conditions if the corresponding terms are not available: 1) Discount rate: 10% (as per OECD DAC Directives); 2) Repayments per annum: 2; 3) Type of repayment: equity principle. Loans are coded as ODA-like only if the record showed 1) sufficient information to measure grant element, 2) a calculated grant element of greater than 25%, and 3) development intent.

3.5 Flow Class

The most critical classification that enables cross-donor comparisons is the flow class field, which indicates whether a project likely qualifies as Official Development Assistance (ODA), Other Official Flows (OOF) or another class of flow. This classification pulls from the other classifications outlined above to determine the flow class of the project.

ODA-Like: To qualify as ODA-like, a flow must 1) be official financing, 2) administered with the promotion of the economic development and welfare of developing countries as the main objective (donor intent = development), and 3) be concessional in character with a grant element of at least 25 percent (using a fixed 10 percent rate of discount). As such, in our dataset, the following flow types should be categorized as "ODA-like" if they are "official financing" and have "development intent": grants, technical assistance, interest-free loans, in-kind contributions of goods and services, and debt relief. As a rule of thumb, loans with a fixed interest rate of 2 percent or lower will have a grant element of at least 25 percent. If the loan terms are known, we calculate the grant element to ensure any projects marked as ODA-like have a grant element of at least 25%.[16] All projects with a grant element of 25% or more and a development intent are automatically classified as ODA-like in our internal software system.

OOF-Like: The OECD defines other official flows (OOF) as "official sector transactions which do not meet the ODA criteria, e.g.: i.) Grants to developing countries for representational or essentially commercial purposes; ii.) Official bilateral transactions intended to promote development but having a grant element of less than 25 per cent; iii.) Official bilateral transactions, whatever their grant element, that are primarily export-facilitating in purpose. This category includes by definition export credits extended directly to an aid recipient by an official agency or institution ("official direct export credits"); iv.) The net acquisition by governments and central monetary institutions of securities issued by multilateral development banks at market terms; v.) Subsidies (grants) to the private sector to soften its credits to developing countries; vi.) Funds in support of private investment." The following activities and forms of official financing should be categorized as "OOF-like": grants with a representational or commercial purpose (i.e. grants that do not have a primary objective of promoting economic development or welfare in the recipient country), loans from a government institution that do not have any apparent grant element (commercial loans based on LIBOR or LIBOR plus a margin) or a grant element lower than 25%, and export credits from a government institution to a recipient institution (Brautigam 2011a: 206). OOF activities also include "short-term credits to exporters (export sellers' credits) to help them finance foreign sales, and ... longer-term credits to foreign buyers to assist in the export of goods and services" (Brautigam 2011a: 206). OOF also includes lines of credit that a government provides to a donor enterprise (state-owned or not-state-owned) to do business overseas.[17] OOF-like projects may include any type of donor intent: development, commercial, representational, or mixed.

---

[13] OECD amended the threshold to 35% in 2016. But since the amendment is not retrospective, we are sticking with the original definition.

[14] Available at https://www.oecd.org/dac/stats/31426795.pdf

[15] It should be noted that the World Bank (http://www.worldbank.org/ida/ida-grant-element-calculator.html) and IMF (http://www.imf.org/external/np/pdr/conc/calculator/default.aspx) use a different instrument to measure grant element. For some types of loans, the OECD may calculate a grant element of 25% or more, while the World Bank and IMF calculate it as less than 25% because the WB and IMF calculators use a discount rate of 5% instead of the OECD's 10%.

[16] The OECD DAC recently introduced a new way of calculating ODA based on a grant equivalency principle. Under the new rules, the concessionality threshold to qualify as ODA changes based on the income group of the recipient country. Also, only the grant-element of a project counts as ODA, rather than the full amount of the concessional loan. The current TUFF classification system outlined above is based on the previous DAC directives from 2013-2015. The new system underscores the importance of greater transparency around grant elements for all concessional financing.

[17] Example: "In Ethiopia. ZTE was able to offer finance for the Ethiopian Government's Millennium Telecoms project, securing a US$1.5 billion deal for which the interest rate was LIBOR plus 150 basis points (Personal communication, 2011). Huawei offered a Brazilian firm financing at LIBOR plus 200 basis points, with a two-year grace period (Bloomberg 2011). As with the other forms of non-concessional official finance, these strategic lines of credit are clearly not ODA" (Brautigam 2011a: 206).

Vague (OF): This category is reserved for flows of official financing that are either ODA or OOF, but for which there is insufficient information to assign to the flows to either the ODA-like or OOF-like category. Official Finance (OF) is an umbrella category that includes ODA-like, OOF-like, and Vague (OF) flows. These projects may have "development" or "mixed (some development)" intent.

Official Investment: This category captures international investments by a donor state agency in an enterprise resident in another country's economy. The donor agency must itself purchase a stake in the recipient enterprise, with the expectation of seeing a return on this investment for the donor government. Since the official donor agency is not simply providing equity or insurance, but is itself the investing agency "purchasing a stake" in the recipient enterprise, these Official Investments are distinct from Joint Ventures or Foreign Direct Investments with a lesser degree of "state involvement." In any unofficial Foreign Direct Investment or Joint Venture, the ultimate investing agent Direct Investment is NOT an official government agency. Official Investment projects must have "commercial" intent.

3.6 Umbrella

Projects are categorized as umbrella projects when the record refers to projects where some agreement was signed between two countries but the funds were not allocated at the original agreement date, only to be allocated through sub projects later on. These projects are marked as "umbrella" projects to mark their likely hierarchical relationship to other records in the dataset and allow users to exclude them from financial analysis. Lines of credit that are not allocated for a specific activity when established are also marked as umbrella projects since often the individual projects funded by the credit line are included elsewhere in the dataset.

3.7 Lines of Credit

In addition to coding lines of credit as 'OOF-like', these records are also marked as a *'line of credit'* in the database interface. Given that lines of credit may or may not be used in their entirety, this marker variable will enable analysis of OOF with and without lines of credits. If there is evidence the line of credit was completely expended, then the 'line of credit' check box is not ticked and it becomes an outstanding loan, rather than an outstanding line of credit.[18] Also, the line of credit marker is not ticked for sub-projects financed by lines of credit.

3.8 Recommended for Research

This field is marked as true for all projects that meet two criteria: 1) they are not umbrella projects, and 2) they have moved through the project cycle to at least the committed, implementation, or completed stages. Users should only use records marked as "true" for any normal analysis of Chinese Official Finance.

---

[18] Journalists and public intellectuals often conflate the *availability* or a credit line and the *use* of a credit line (Brautigam 2010). Many credit lines from the Chinese government are used sparingly or not at all.

Figure A: Classification of Official Finance
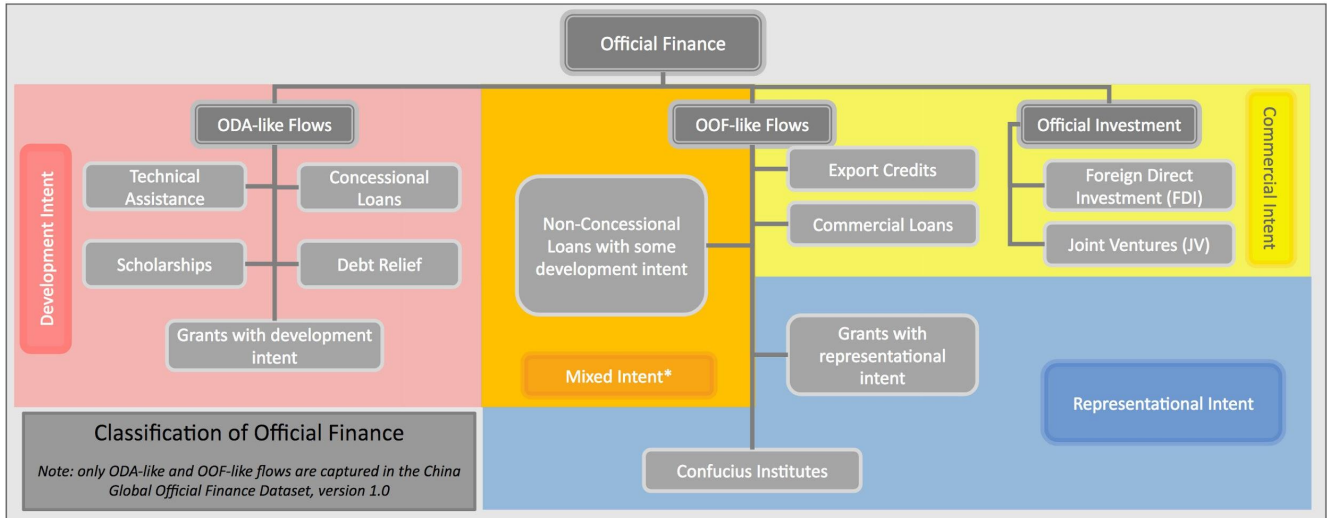
*Mixed Intent indicates the donor intent was some combination of development, commercial, or representational intents.

4. Research Release Publication

In addition to actively maintaining the data dashboard (http://china.aiddata.org/), AidData has been publishing research releases, which are static versions of the database, with different year and/or regional coverage. Research Releases include projects that meeting the following standards:
• Verified
• Active
• Time-stamped
• Flow class is ODA-Like, OOF-Like or Vague (Official Finance).

While we take a relatively inclusive approach when including projects in research releases, we recommend using only projects that meet all the following standards in any rigorous academic research or estimate:
• Not an umbrella project
• Status is "Pipeline: Commitment" or "Implementing" or "Completion"

An easy way to include only projects that are recommended for analysis is to use the "recommended_for_research" field in the dataset as a filter.[19] However, users can make their only decision in choosing what types of projects to focus on for their own purpose.

---

[19] While debt relief projects are included in the recommended projects, we would like to encourage users to make their own decision as to whether include or exclude these flows due to possibility of double counting.

Works Cited

Brainard, Lael and Derek Chollet. 2007. Development's Changing Face: New Players, Old Challenges, Fresh Opportunities. Scene-Setter for 2007 Brookings Blum Roundtable.

Brautigam, Deborah. 2010. China, Africa and the International Aid Architecture, Working Papers Series N° 107. Tunis, Tunisia: African Development Bank

Brautigam, Deborah. 2011a. "Chinese Development Aid in Africa: What, Where, Why and How Much?" in *China Update 2011*, edited by Jane Golley and Ligang Song. Canberra, Australia: National University.

Brautigam, Deborah. 2011b. Aid 'With Chinese Characteristics': Chinese Aid and Development Finance Meet the OECD-DAC Regime. *Journal of International Development* 23(5).

Bunte, Jonas B., Harsh Desai, Kanio Gbala, Bradley C. Parks, Daniel Miller Runfola. 2017. Natural Resource Sector FDI and Growth in Post-Conflict Settings: Subnational Evidence from Liberia. AidData Working Paper #34. Williamsburg, VA: AidData.

Corking, Lucy. 2011. Redefining Foreign Policy Impulses toward Africa: The Roles of the MFA, the MOFCOM and China Exim Bank, in: Journal of Current Chinese Affairs, 40, 4, 61-90.

Downs, Erica. 2011. China Development Bank's Oil Loans: Pursuing Policy – and Profit. *China Economic Quarterly*. 15.4 (2011): 43-47.

Drakos, Konstantinos. 2007. The Size of Under-Reporting Bias in Recorded Transnational Terrorist Activity. *Journal of the Royal Statistical Society: Series A (Statistics in Society)* 170 (4): 909-921.

Drakos, Konstantinos and Andreas Gofas. 2006. The Devil You Know but Are Afraid to Face: Underreporting Bias and its Distorting Effects on the Study of Terrorism. *Journal of Conflict Resolution* 50(5): 714-735.

Dreher, Axel, Andreas Fuchs, Bradley Parks, Austin M. Strange, and Michael J. Tierney. 2017. Aid, China, and Growth: Evidence from a New Global Development Finance Dataset. AidData Working Paper #46. Williamsburg, VA: AidData.

Djankov, Simeon and Bernard Hoekman. 2000. Foreign Investment and Productivity Growth in Czech Enterprises. *World Bank Economic Review* 14(1): 49-64

EM-DAT. 2012. EM-DAT: The OFDA/CRED International Disaster Database – www.emdat.be. Brussels, Belgium: Université Catholique de Louvain.

Fengler, Wolfgang and Homi Kharas. 2010. *Delivering Aid Differently: Lessons from the Field*. Washington D.C.: Brookings Institution Press.

Foster, Vivien, William Butterfield, Chuan Chen, and Nataliya Pushak. 2008. Building Bridges: China's Growing Role as Infrastructure Financier for Sub-Saharan Africa. Washington D.C.: World Bank/PPIAF.

Gallagher, Kevin P., Amos Irwin, and Katherine Koleski. 2012. The New Banks in Town: Chinese Finance in Latin America. Washington D.C.: Inter-American Dialogue.

Ghose, Siddhartha. 2017. Tracking Funding for the Ebola Epidemic in Sierra Leone. AidData and the US Global Development Lab.

Hendrix, Cullen, Idean Salehyan, Jesse Hamner, Christina Case, Christopher Linebarger, Emily Stull, and Jennifer Williams. 2012. Social Conflict in Africa: A New Database. *International Interactions*. 38 (4): 503-511.

International Development Association (IDA). 2008. Aid Architecture: An Overview of the Main Trends in Official Development Assistance Flows. Washington D.C.: World Bank.

International Monetary Fund (IMF). 2006. Evaluation Report: Financial Sector Assessment Program. Washington DC: Independent Evaluation Office (IEO) of the International Monetary Fund. Retrieved from http://www.ieo-imf.org/ieo/files/completedevaluations/FSAP_Main_Report.pdf

International Monetary Fund (IMF). 2007. Evaluation Report: IMF Exchange Rate Policy. Washington DC: Independent Evaluation Office (IEO) of the International Monetary Fund. Retrieved from http://www.ieoimf.org/ieo/files/completedevaluations/05172007exrate_full.pdf

Jácome, Francine. 2011. Petrocaribe: The Current Phase of Venezuela's Oil Diplomacy in the Caribbean. Policy Paper 40. Bogotá, Colombia: Friedrich-Ebert-Stiftung.

James, Ian. 2011. Haiti Leader Says Venezuela Aid is Key. *Associated Press*. December 4, 2011.

Kharas, Homi. 2007. Trends and Issues in Development Aid. Washington D.C.: Brookings Institution.

King, Gary and Will Lowe. 2003. An Automated Information Extraction Tool for International Conflict Data with Performance as Good as Human Coders: a Rare Events Evaluation Design. *International Organization* 57(3): 617–42.

Klein, Michael and Tim Harford. 2005. *The Market For Aid*. Washington D.C.: International Finance Corporation.

Leetaru, Kalev. 2010. The Scope of FBIS and BBC Open-Source Media Coverage, 1979–2008 (U). *Studies in Intelligence* 54 (1): 17-37.

Lum, Thomas, Hannah Fischer, Julissa Gomez-Granger, Anne Leland. 2009. China's Foreign Aid Activities in Africa, Latin America, and Southeast Asia. Congressional Research Service Report for Congress, February 25, 2009. Washington D.C.: Congressional Research Service.

Manning, Richard. 2006. Will 'Emerging Donors' Change the Face of International Cooperation? *Development Policy Review* 24: 371-385.

McCarthy, John D., Clark McPhail, and Jackie Smith. 1996. Images of protest: Dimensions of Selection Bias in Media Coverage of Washington Demonstrations, 1982 and 1991. *American Sociological Review* 6(3): 478-99.

Minor, Allison, Pranay Sinha, Brooke Russell, and Charles Perla. 2014. Defining & Measuring Development Assistance from GCC Countries: A Methodological Note. Williamsburg, VA: AidData.

Morgan, Pippa, and Yu Zheng. 2017. Tracing the Legacy: China's Historical Aid and Contemporary Investment in Africa. Working Paper.

Musakwa, Tendai. 2013. Deborah Brautigam Discusses Doing Research on China-Africa Relations. The China-Africa Project. Interview transcript available at *http://www.chinaafricaproject.com/deborah-brautigam-discusses-doing-research-on-china-africa-relations/*.

OECD. DAC Glossary of Key Terms and Concepts. Available at http://www.oecd.org/development/stats/dacglossaryofkeytermsandconcepts.htm.

OECD. 2008. "Is it ODA?" OECD Factsheet November 2008. *http://www.oecd.org/investment/aidstatistics/34086975.pdf*

Raleigh, Clionadh, Andrew Linke, Håvard Hegre, and Joachim Karlsen. 2010. Introducing ACLED: An Armed Conflict Location and Event Dataset. *Journal of Peace Research* 47(5): 651-660.

Reeves, Andrew, Stephen Shellman, and Brandon Stewart. 2006. Fair & Balanced or Fit to Print? The Effects of Media Sources on Statistical Inferences. Paper presented at the International Studies Association conference, 22– 25 March 2006, San Diego, CA.

Severino, Jean-Michel and Olivier Ray. 2010. The End of ODA (II): The Birth of Hypercollective Action. Center for Global Development Working Paper 218. Washington DC: Center for Global Development.

Shellman, Stephen. 2008. Coding Disaggregated Intrastate Conflict: Machine Processing the Behavior of Substate Actors Over Time and Space. *Political Analysis* 16:4 64–477.

Schrodt, Philip A. and Deborah J. Gerner. 1994. Validity Assessment of a Machine-Coded Event Data Set for The Middle East, 1982–1992. *American Journal of Political Science* 38(3):825–54.

Schrodt, Philip A., Erin M. Simpson, and Deborah J. Gerner. 2001. Monitoring Conflict Using Automated Coding of Newswire Sources: A Comparison of Five Geographical Regions. Paper presented at the PRIO/Uppsala University. Uppsala, Sweden 8–9 June 2001.

Scikit-learn: Machine Learning in Python, Pedregosa et al., JMLR 12, pp. 2825-2830, 2011

Simmons, Beth A. 2000. The Legalization of International Monetary Affairs. *International Organization* 54: 573-602

Strange, Austin M., Vijaya Ramachandran, Brad Parks, Michael J. Tierney, Axel Dreher, and Andreas Fuchs. 2013. China's Development Finance to Africa: A Media Based Approach to Data Collection. Working Paper. Washington D.C.: AidData/Center for Global Development.

Strange, Austin, Axel Dreher, Andreas Fuchs, Bradley Parks, and Michael Tierney. 2017. Tracking Underreported Financial Flows: China's Development Finance and the Aid-Conflict Nexus Revisited. *Journal of Conflict Resolution* 61 (5): 935-963.

The Economist. 2010. Brazil's Foreign-Aid Programme: Speak Softly and Carry a Blank Cheque. 15 July 2010.

United States. Cong. Senate. Committee on Foreign Affairs. *Hearing on China's Growing Role in Africa. Nov 1, 2011.* 112th Cong. 1st sess. Washington: GOP 2011 (statement of Deborah Brautigam, Professor, School of Advanced and International Studies).

Vines, Alex, Lillian Wong, Markus Weimer, and Indira Campos. 2009. Thirst for African Oil: Asian National Oil Companies in Nigeria and Angola, Chatham House, London, August.

Walz, Julie and Vijaya Ramachandran. 2011. Brave New World: A Literature Review of Emerging Donors and the Changing Nature of Foreign Assistance. CGD Working Paper 273. Washington D.C.: CGD.

White Paper on China's Foreign Aid. 2014. State Council of the People's Republic of China.

Woods, Ngaire. 2008. Whose Aid? Whose Influence? China, Emerging Donors and the Silent Revolution in Development Assistance. *International Affairs* 84: 1205-1221.

Woolley, John T. 2000. Using Media-Based Data in Studies of Politics. *American Journal of Political Science* 44: 156-173.

Yonamine, James E. and Philip A. Schrodt. 2011. A Guide to Event Data: Past, Present, and Future Version 1.0. Working Paper.

Appendix A: Personnel Outreach and Crowdsourcing

Personnel Outreach

Donor and/or recipient personnel outreach is a complementary sub-step in Stage Two searching that can help fill critical data gaps remaining after exhaustive Factiva and web searching. "Personnel Outreach" is the process of establishing a direct dialogue with donor and/or recipient personnel affiliated with a given project via email, telephone or Skype. RAs then engage these personnel with targeted questions that seek to fill outstanding information gaps for a given project record (e.g. the repayment terms of a loan).

Crowdsourcing Protocol

The TUFF methodology is designed to provide multiple layers of scrutiny for every project entry with an even more rigorous standard of evidence for large projects. We opened project records to user feedback to provide an additional avenue to refine project data. To maintain a fully transparent and replicable methodology, this section introduces our process for adjudicating between new sources of crowdsourced feedback and existing project information.

AidData has created a dynamic project data platform at china.aiddata.org, which allows users to investigate individual projects more thoroughly. Each project record has a page where its attributes are displayed. These pages are accessible by three paths:

- Visualization (accessible on the navigation bar > Visualize > Country-level Map. After clicking a country, a user may select a sector to see the projects in that country-sector)
- Search and filter (accessible on the navigation bar > Data > Search Projects)
- Direct access by project ID (accessible on the navigation bar > Data > Find by ID)

These project pages include all the data in the static dataset, including links to source documents. These source documents are often behind a firewall of Factiva, which can be accessed through many university library systems or a 30-day free trial.[20] The project page also includes a comment box. We understand that some users will have information to add to these records and we encourage them to do so via the comment box. AidData staff members will track and moderate comments, addressing any data issues and integrating new content provided by users. Users may add important facts to these records which are not always available in media reports: cancellation, changes in funding or implementation, local impact or project outcome. Users may also help maintain the project database by reporting errors in the database, reporting duplicated projects or providing links to additional documentation. AidData team members will use these comments to improve the data whenever possible. If a comment reports an issue, AidData team members will remove the comment only after that issue is addressed. Comments referencing new sources or providing new information will be preserved. Also, data users may suggest new projects not uncovered during Stage One or Stage Two searching.

We encourage users to provide evidence when commenting on a project record. However, to ensure data quality, AidData staff review and curate all comments prior to publication. To evaluate the validity of comments without a citation, AidData staff also repeat Stage Two Google search for the project looking for the most recent information.

What user-generated content will AidData staff review first?

AidData staff and faculty have created a protocol to determine whether to edit a project record in response to a user comment. The arbitrator will first assess the reliability of the user's source relative to all other project resources (resource ranking in Appendix B). If two different project sources conflict on an objective point of fact (commitment date, status, project amount, etc.), AidData staff will pull from the most verifiable resource type to populate the disputed data field.

A data point is considered most verifiable if it has the following characteristics (in order):

1. Distributed by a reliable and accountable source (see rankings below)
2. Two or more independent sources report same data point
3. Distributed by a widely circulated source (e.g. AP or BBC)
4. Specific and detailed

---

[20] See http://www.factiva.com/integration/modules/microsoft/trial.asp

By priority:
1. Project variables (amount, status, etc). **Require source for verification.**
2. Alterations to project description. **Require source for verification.**
3. Opinions on project performance, quality, or necessity
4. Other non-substantive opinions

Appendix B: TUFF Source Prioritization Protocol

It is common for sources to have conflicting information on a certain project. In this case, it is necessary to have a hierarchical ranking of how much we weigh in each source.
Ranking of Resource Types based on Reliability of Project Data

1. Official government source, from a donor or recipient government agency
2. Implementing or intermediary agency report/website
3. Other official Source (e.g. World Bank, CIA, etc.)
4. Peer-reviewed scholarly article
5. Other scholarly output, including working papers and dissertations
6. NGO, civil society, or advocacy group report/website
7. Media reports, including Wikileaks
8. Social media, including blogs from any unofficial source

If two sources conflict that are in the same level on the hierarchy (e.g. two media sources report conflicting project amounts), then the AidData staff member must arbitrate by explicitly stating within the project description a) the source of this conflict and b) the reasoning for the proposed solution.

Appendix C: Health of Record Scoring Criteria

The purpose of the Health of Record score is to rate the completeness and verifiability of each record. The score can be used internally to prioritize subsequent rounds of research and quality checks or it can be used by external actors who only want to analyze data with a particular level of accuracy. Below are the criteria for the resource and completeness scores.

Source Triangulation Score: This score is designed to capture our level of confidence in the information in a project record. The most important factors influencing the accuracy of a project entry are the type and number of resources used to create the project record. Other factors that increase our level of confidence in a project record are successfully executing the ground-truthing methodology and finding multimedia evidence (videos or photos) proving the existence of a project.

1. Base Score: This portion of the score is determined by the number of *media reports* a project has. The base score is informed by the actual distribution of resources in the database.
    a. Projects will receive 1 point for each additional media report (2 and above)
    b. Points will be capped at 4 because the added value of media reports is diminishing due to repetition of the same information[21]
2. Value Added Score: This portion of the resource score awards extra points to project records which have other (more credible) types of resources informing them. A project will receive extra points for each *category* of resource it draws from.
    a. Official Government Sources (Donor/Recipient): 3
    b. Other Official Sources (non-Donor/non-Recipient): 3
    c. Implementing Agency Source: 2
    d. Academic Journal Articles/Other Academic Sources: 2
    e. NGO/Civil Society/Advocacy: 1
    f. Social Media, including unofficial Blogs: 1
3. Bonus Points: We also award additional points to project that have undergone additional types of evidence for their existence, such being ground-truthed or having video/photo evidence of existence.
    a. Successfully ground-truthed: 4 points

Data Field Completeness Score: In addition to scoring based on the accuracy of the resources informing it, a project's health of record score is also determined by its level of completeness (i.e. having all of its fields populated). The extent of penalty for a missing field in a project record will be in weighted based on how central the field is for analysis of the record. A project will also *earn* points if certain fields are populated. A central goal of the completeness score is to distinguish projects that would benefit additional rounds of searches and scrutiny from completed records.[22]

High value fields:

– Transaction Amount: Projects with missing financial amounts will receive a 1 point deduction
– Commitment Year: Project without a commitment year or tagged "year uncertain" will receive a 1 point deduction
– Flow Class: "vague" records will receive 1 point deduction
– Flow Type: Vague-TBD/Uset records will receive a 1 point deduction
– Sector: Unallocated/Unspecific projects will receive a 1 point deduction[23]

Status: To identify records that merit an additional round of searchers to seek additional information, the completeness score will take status into account. It is reasonable to assume that completed or cancelled projects will not receive additional media coverage whereas pipeline, implementing, or suspended projects could receive additional coverage.
– Projects that are marked as completed or cancelled will receive 1 point since we can be confident that additional information will not be forthcoming.
– Projects that are marked pipeline or implementation will receive 0 points.

---

[21] We chose to cap the score rather than set up an equation for diminishing marginal points because we want to keep the scoring criteria as simple as possible.

[22] Projects will receive a base score of 8 to prevent projects from receiving a negative data completeness score.

[23] Although it is the case that some grants and loans are given without being earmarked, this deduction is designed to show that a project record could benefit from additional follow-up i.e. how was the grant spent if at all.

Other fields:
Funding Agency: Projects without a specific funding agency will lose a point.

Implementing/Accountable Agency: Projects without an implementing or accountable agency will also lose a point.

Appendix D: Automated Data Collection Process and Logical Inconsistency Check

To standardize the coding and eliminate human errors, AidData has developed a set of automated

Supplier: Certain government agencies are responsible for providing certain modalities of financing or targeting their financing in certain sectors. These decision rules draw heavily from the _2014 White Paper on Chinese Foreign Aid_ and other high level academic research on China's Official Finance portfolio.

Export Import Bank of China

- If "funding agency" = Export-Import Bank, then "flow" = Loan OR Export Credit OR Strategic/Supplier Credit OR Debt Rescheduling[24] (United States Senate 2011)
- If "loan type" = concessional, then "funding agency" = Export-Import Bank (Corkin 2011; Davies et al. 2008; "China's Foreign Aid" 2014)
- If "funding agency" = Export-Import Bank, then "flow" ≠ any other flow type
- If "funding agency" = Export-Import Bank and sources indicate the loan is concessional, then "flow class" = "ODA-Like"

China Development Bank
- If "funding agency" = China Development Bank, then "flow" = Loan OR Export Credits[25]
- If "funding agency" = China Development Bank, then "flow class" = OOF-like (Downs 2011)
- If "funding agency" = China Development Bank, then "loan type" ≠ Concessional (Davies et al. 2008; cdb.com 2015)

China Africa Development Fund
- If "funding agency" = China Africa Development Fund, then "flow" = Official Investment [26] [27]

People's Bank of China
- If "funding agency" = People's Bank of China, then "flow class" is Official Finance.

Other Logically Inconsistent Variable Combinations:
- If "grant element" > 24% then "loan type" must ≠ "non-concessional"
- If "grant element" > 24% then and "intent" = Development, then "flow class" ≠ "OOF-like" OR "Vague OF" (check recipient oda eligibility status first)
- If "project type" = Medical Team, then "flow class" ≠ "OOF-like" OR "Vague OF"
- Vague (OF) and Grant element >25%: If a loan has a grant element above 25% it should not be coded as Vague (OF).
- If "flow type" = Export Credit, then "flow class" ≠ "ODA-like" or "Vague OF"
- If "intent" = "Commercial" or "Representational" then "flow class" ≠ "ODA-like"
- If "intent" = Commercial or "Representational" then "flow class" ≠ "Vague OF"

If "credit line = TRUE" then "umbrella = TRUE"

---

[24] In rare cases, China Export-Import Bank has also provided grants.

[25] In rare cases, China Development Bank has also provided grants and scholarships.

[26] _Brautigam 2009_, pg. 205.

[27] _Brautigam 2011_, pg. 4.

Appendix E: Tracking the Official Finance Activities of Gulf Cooperation Council Donors

In summer of 2014, AidData successfully extended the TUFF methodology to the official financing activities of Saudi Arabia and Qatar. While the methodology for tracking Arab financing is generally identical to tracking Chinese official financing, there are several idiosyncrasies in data collection and coding activities that must noted.
Stage One Factiva Search Phrase:

Saudi Arabia

(Saudi Arabia or Saudi or Saud* or Riyadh) near5 (Recipient) or (assistance or assist* or grant or loan or concession* or donat* or interest-free or joint fund or package or aid or humanitarian or oil aid or saudi development fund or saudi red crescent or waleed or foundation or ministry)

Qatar

(Qatar or Qatari or Qatar* or Doha) near5 (Bangladesh or Bangladeshi or Bangladesh* or Dhaka) AND (assistance or assist* or grant or loan or concession* or donat* or interest-free or joint fund or package or aid or humanitarian or oil aid or foundation or ministry or budget support or silatech or zakat)

1) In addition to performing Stage One Factiva searches in English, RAs use an identical Arabic search phrase (below). When performing Arabic searches change Factiva's language setting from "English" to "Arabic." Execute Stage One searches using the same steps described on pages 42 to 45.

Saudi Arabia

(مؤسسة الوليد طلال or الهلال الأحمر السعودي or الصندق السعودي للتنمية or نطفية مساعدات or قض or إنسانية مساعدات or
خارجية مساعدات or معونة) AND (اليمن) near5 (السعودية)

Qatar

(قطر) near5 (اليمن) AND (معونة or خارجية مساعدات or إنسانية مساعدات or قض or الصندق السعودي للتنمية or الهلال الأحمر
السعودي or مؤسسة الوليد طلال)

2) To further increase coverage of Gulf Cooperation Council projects, our team designed an additional *organization specific* Factiva search phrase for Saudi Arabia and Qatar. Execute searches using the same steps on described on pages 42 to 45. The Saudi and Qatari organization specific searches are listed below.

Saudi Arabia

(Saudi Fund for Development or Saudi Red Crescent Authority or Alwaleed Bin Talal Foundation or Prince Sultan bin Abdelaziz Foundation or King Abdullah International Foundation for Charity) AND (Recipient)

Qatar

(Sheikh Jassim Bin Jabor Al Thani Charitable Foundation or Education Above All or Reach Out to Asia) AND (Bangladesh or Bangladeshi or Bangladesh* or Dhaka)

3) Neither Saudi Arabia nor Qatar operate country-level Economic and Commercial Counselor websites in the same way as China. However, the team was able to replicate Stage One+ government information system searches using the *Saudi embassy website*. The website contains a page on Saudi Arabia's foreign aid over the last calendar year. To scrape information from previous years the team used an internet archiving tool called *Wayback Machine*.

4) Idiosyncrasy of the GCC project was the creation of a flow class variable called "Official Religious Aid." This modality captures official *Zakat* (see *this blogpost* for more details on this aid modality). The categories and definitions of Official Religious Aid are below.

Purpose of ORA	Description
Explicitly religious activities (ORA-1)	Aid where the exclusive purpose is to facilitate religious worship or other religious activities. This could include construction of mosques, exclusively religious education (including Qur'anic schools), etc.
Aid linked to religious activities (ORA-2)	Aid that addresses religious worship or other religious objectives, but also serves other development objectives. This could include food aid during religious holidays (fast-breaking meals during Ramadan, meat donations during Eid al Adha, etc.), funding of religious schools that also provide instruction in other subject areas, etc.
Aid with a special religious meaning (ORA-3)	Aid that promotes the development and welfare of the recipient country (i.e., fits into existing ODA purpose codes) [28] but also holds special religious importance. This could include activities that are common forms of zakat or sadaqa, like aid to orphans/orphan sponsorship. *Note that ORA-3 will be categorized as ODA. Because it is difficult to distinguish from ODA when it is not explicitly identified as religious aid, it will not be included in the cumulative ORA figures.*

5) By assumption, our team coded *all* loans from the Saudi Fund for Development (SFD) with "development" intent as "ODA-like" regardless of whether they could uncover information on a loan's conditions. This assumption is based on public reporting by SFD on its general loan conditions.

---

[28] See 'Code List 2014,' DAC and CRS Code Lists, Organization for Economic Cooperation and Development (OECD). http://www.oecd.org/dac/stats/dacandcrscodelists.htm

Appendix F: AidData's Deflation Methodology

1 Currency Conversion and Deflation Purpose

Financial values collected as part of AidData's data collection activities, including TUFF, must be converted and deflated so that they are comparable across currencies and years. AidData's methodology follows after the OECD's methodology.[29] The full methodology involves two steps: 1) Calculating nominal exchange rates and 2) calculating deflation rates detailed below. Where available for countries that belong to the OECD Development Assistance Committee, we use the OECD's deflators with a base year of 2014. For other countries, including China, we calculate the deflators based using the OECD's methodology using World Bank sources for exchange rates and inflation.

2 Exchange Rates

2.1 Exchange Rate Methodology

Before deflation, all values must first be expressed in nominal (current) U.S. dollars (USD). This is done with an LCU per USD exchange rate, applied by:
(original value) / (LCU per USD) = (new value)
For example:
100 EU / .7 = 142.57 USD

2.2 Exchange Rate Sources
World Bank -- Annual Exchange Rate for Local Currency Units (LCUs) per USD.
http://data.worldbank.org/indicator/PA.NUS.FCRF/countries

2.3 Currency Revaluations/Changes in Currencies
The standard data from the World Bank does not take into consideration currency revaluations and currency changes. So to reflect this nuance, we identified the complete list of countries that had undergone currency changes or revaluations that would affect the exchange rates used in TUFF datasets. In cases where the financial amount was quoted in old currencies, we used historical exchange rates (annual period average) from OANDA to calculate the exchange rate to USD.

3 Deflators

Deflation is necessary to take the USD nominal amount and deflate (or inflate) that amount into a constant year across the whole dataset so all the financial values are comparable despite year values. Deflators control for two changes over time: inflation in the donor country and change in buying power in the donor country relative to the United States. Both of these changes are calculated separately, and then multiplied together to get the final deflator used by AidData. The formula is the following:

*Inflation * Change in Buying Power = Deflator*

3.1 Inflation

The first part of the deflator formula is to calculate the inflation value from the base year to the constant year.

Inflation is measured as relative to a given base year. AidData 4.0 will use the base year of 2014. Percentages were then generated using the following formula:

$\text{Percentage}_{Year} = \text{Percentage}_{Previous\ Year} + (\text{Percentage}_{Previous\ Year} * \text{Inflation}_{Year})$

For example, in 2014, Colombia's GDP inflation was 4.2%. Taking 2014 as the base year, the percentage for 2014 is 100%. So, to calculate the percentage for 2010, using 2014 as the start year:

$100 = P_{previous} + (P_{previous} * .04)$

$P_{year}\ I_{year}$

This yields 96% as Colombia's percentage for 2012. (Decimals have been rounded for this example, but were not rounded for AidData's deflator table.) In 2012, Colombia's GDP inflation was 8%. Then, to calculate 2012, 2013 is the start year:

$$96 = P_{previous} + (P_{previous} * .08)$$

$$P_{year} \, I_{year}$$

This yields 89% as Colombia's percentage for 2012.

The following sources are used to compile the inflation values.

World Bank GDP Inflation -- http://data.worldbank.org/indicator/NY.GDP.DEFL.KD.ZG

3.2 Change in Buying Power

The second part of the deflators formula is to calculate the change in Buying Power for the donor country.

The change in buying power is taken from the LCU per USD rate and expressed as:

$$\text{Exchange Rate}_{\text{Base Year}} / \text{Exchange Rate}_{\text{Transaction Year}} = \text{Change in Buying Power}$$

For example, the Korean Won to USD rate was 1273.9 in 2014 and 804.4 in 1996. The subsequent change in buying power is:

$$1273.9 / 804.4 = 1.58$$

Note that this methodology yields a ratio of 1 for all currencies pegged to the USD.
The data used for the buying power formula are generated from the historical exchange rates described above. See the Source section 2.2.

3.3 Finalized Deflators

The GDP inflation and change in buying power numbers are combined to create annual deflators for donor countries:

$$\text{Inflation} * \text{Change in Buying Power} = \text{Deflator}$$

4 Examples on Using GDP Deflators

Amounts in LCU should be converted to nominal USD, using the LCU per USD exchange rates found in sheet A1. Then, the values should be *divided by* the percentages in sheet "E1". For example, in 1975, Kuwait funded an electrification project in Bangladesh worth 6,400,000 *KD1975* (AidData ID 2427051). To convert this amount to USD 2014, first, convert it to USD 1975:

6,400,000 *KD 1975* / (.29003 *KD/USD 1975*) = 22,066,505.30 *USD 1975*

Next, divide it by the AidData deflator:

22,066,505.30 *USD 1975* / 20.83% = 105,936,175.20 *USD 2014*

Note that amounts that are *already reported in USD* do not need to be converted. They only need to be deflated (divided by the appropriate deflator).