AidData's Media-Based Data Collection Methodology

Version 1.0

April 2013

Austin M. Strange, Brian O'Donnell, Daniel Gamboa, and Bradley Parks

AidData is a joint initiative of Development Gateway, the College of William and Mary, and Brigham Young University

DEVELOPMENT GATEWAY WILLIAM & MARY BYU

www.aiddata.org info@aiddata.org

Table of Contents

Introduction to Media-Based Data Collection (MBDC) Methodology 3

MBDC Codebook

Stage One 7

Stage Two 11

Appendix A: Best Practices for Maximizing MBDC Workflow Efficiency 30

Appendix B: Notes on Chinese Language Searches 32

Appendix C: Personnel Outreach and Crowdsourcing 33

Appendix D: Flow Type and Flow Class 34

Appendix E: Pledges and Commitments 43

Appendix F: Donor Intent 45

Appendix G: Screenshots to Demonstrate Search and Data Collection Processes 49

Appendix H: MBDC Stage Two Process 54

Appendix I: Database Variable Definitions 55

Introduction[1]

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 assistance, the "market" for development finance 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, Russia, 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.[2] This makes it difficult to gauge the nature, scope and impact of the development finance activities of these actors, fueling uncertainty and speculation about the intentions of these non-Western donors.

In an attempt to help address this issue, in January 2012 AidData conducted a pilot initiative using media-based data collection (MBDC) methods to track project-level Chinese development finance to Africa from 2000-2011.[3] Political scientists, economists, sociologists, geographers, and computer scientists have already leveraged MBDC methodologies for a variety of purposes—for example, tracking violent and non-violent conflict incidents; documenting the scale, scope, and impact of natural and man-made disasters; and studying patterns of political interaction and sentiment (Schrodt and Gerner 1994; King and Lowe 2003; Shellman 2008; Leetaru 2010; Raleigh et al. 2010; Yonamine and Schrodt 2011; EM-DAT 2012;

[1] The authors are indebted to Alex Miller, Dylan Kolhoff, Jaclyn Goldschmidt, James Juchau, Sarah Christophe, Alexandra Foster, Kevin McCrory, Kyle Titlow, Wen Chen, Yaqing Wen, Patrick Leisure, Charles Perla, Henrique Passos Neto, Grace Perkins, and Amber Will for their outstanding research assistance. Andreas Fuchs was instrumental in working with AidData staff to define schemes to categorize flow types and flow classes. We are also very grateful to Robert Mosolgo, a Project Manager at AidData who created and continues to manage the online MBDC coding interface. The authors are solely responsible for any errors in this document.

[2] This is not entirely true, as Russia provides aggregate aid statistics to the OECD. However, it does not provide project-level data on outgoing Russian aid flows.

[3] During the pilot phase of the MBDC project, the methodology was designed to capture only official development finance flows. Researchers did not systematically search for unofficial finance, military aid, or official investment. Therefore, only data on official development finance (official projects with development intent) can be considered comprehensive. More extensive Stage One searches, for such terms as "invest*" or "joint venture", would provide a fuller picture of Chinese investment flows in Africa.

Hendrix et al. forthcoming). However, the study of development finance allocation and impact has not yet benefited from the systematic application of MBDC methods. Several ad-hoc efforts had been undertaken, but none had resulted in systematic, transparent, and replicable data collection procedures (Foster et al. 2008; Lum et al. 2009; Gallagher et al. 2012).

The goal of AidData researchers was to to test whether MBDC methods can be a viable way to accurately gather and standardize project-level development finance information, specifically regarding bilateral and multilateral agencies that are unwilling or unable to disclose their data. We concluded—after sixteen months of developing, testing, and fine-tuning a pilot methodology—that MBDC methods can indeed be a valuable tool for gathering and standardizing project-level development finance information. However, as we detail below, MBDC is inherently imperfect for several reasons, and is most effective as a research tool when a) the data are made widely accessible to users with the necessary knowledge to who can identify errors and b) leverage complementary qualitative data collection methods (such as in-country fieldwork and outreach with personnel involved with specific projects) to prevent overreliance on media reports.

This codebook outlines the set of procedures that were developed, tested, refined, and implemented by AidData staff and affiliated faculty at the College of William and Mary and Brigham Young University during the MBDC China pilot.[4] We initially employed these methods to achieve a specific objective: documenting the "known universe" of Chinese development finance projects in Africa from 2000 to 2011 (Strange et al. 2013). The procedures in this codebook constitute a flexible methodology that can be applied to broader research questions, and to track other development finance flows from donors outside the OECD's Development Assistance Committee (better known as "Non-DAC donors").

AidData coders follow standardized procedures at each stage of the MBDC process. By identifying each step and procedure in this codebook, we have made our methods transparent enough for others to scrutinize and test themselves. This methodology will undergo several rounds of review and revision; therefore, we strongly encourage input and constructive criticism from readers.[5]

[4] AidData's media-based data collection methodology is based in part on the methodology developed by the Public-Private Infrastructure Advisory Facility (PPIAF) (Foster et al. 2008).
[5] Questions, comments, and feedback can be directed to the authors or info@aiddata.org.

Notwithstanding the rigor and replicability of our approach, the nature of MBDC presents unique challenges for data completeness, accuracy, quality, and credibility (Woolley 2000; Schrodt et al. 2001; Reeves et al. 2006). First, as with any social scientific inquiry, there is potential for human error by coders throughout the data collection process. Second, information extracted from public media outlets throughout the world cannot substitute complete and accurate statistical data from official sources. Media-based data collection can only be as good as the imperfect data sources upon which it relies. In some instances, separate media reports provide conflicting or incomplete financial information for a single project. Did the Namibian presidential palace (ID 1255) cost N$60 million (as reported by the Chinese government) or N$30 million (as reported by an AllAfrica source)? If Wikileaks and BBC Media report two different commitment years and amounts for a Djiboutian fiber optics cable project (ID 421), which source is more reliable? In the absence of official project-level data, there is no foolproof method for adjudicating between conflicting media reports.[6] Similarly, several sources may report on a Zambian hydropower station (ID #977), but none will report the concessionality of the Chinese loan used to pay for the station, nor what percentage of these committed funds were actually disbursed. The challenge of conflicting or insufficient media reports may be exacerbated in less developed regions with fewer resources or political will to invest in journalism. The quality of many mainstream media reports is likely strained by local resource constraints in Africa and other regions that receive significant development assistance but lack developed journalism industries (Musakwa 2013). Similarly, the motives behind media reporting can be economic or political, a reality that can jeopardize the impartiality of reporting.

Third, relying on media reports faces the risk of "detection bias," or the risk that countries with lower levels of press freedom are less likely to report on official finance activities from various donors. Among sociologists and those who study conflict and terrorism, there is an acceptance that the use of media reports to identify inherently political events (e.g. political protests,

[6] However, it is also not the case that official sources are invariably more credible (and valuable) than media-based information. Media-based data collection that relies on information regarding the implementation and/or the completion of projects can provide more useful and accurate project-level information than official reports, depending on how official project information is collected, updated and presented. Indeed, the reliability and usefulness of "official" data often declines sharply as projects move from the planning stage to the implementation stage. As projects are carried out, donors and recipients often encounter formidable coordination and accountability challenges (Kharas 2007). Consequently, donor agency officials and their government counterparts often improvise and make course corrections, which results in activities that are often substantially different from the original project blueprint that is reflected in the "official" data.

terrorist attacks) introduces a risk of selection bias (McCarthy et al. 1996; Earl et al. 2004; Drakos and Gofas 2006; Drakos 2007).[7] MBDC is thus an imperfect way to fill data gaps that impede research and evidence-based policymaking. Its value as a tool for understanding development finance increases substantially as more people scrutinize and improve the data, and when it is used in parallel with other data collection techniques. Fourth, there is no way to confirm that MBDC methods cover the entire spectrum of available media reports. We highly doubt that AidData's MBDC approach, which uses Factiva, Google and donor search engines (e.g. Baidu), has captured all available reporting on Chinese development finance to Africa, particularly in regions where local language press is not translated into English.

Despite these imperfections, MBDC can be an effective method for capturing unreported or underreported sources of official development finance when used appropriately.[8] Our initial application of the MBDC methodology to track Chinese official development finance flows to Africa has resulted in a rich dataset with 1,673 officially financed projects, totaling more than $75 billion in official commitments.[9] For those eager to quantify China's aid to Africa, it may be instinctual to interpret the previous sentence as a claim that China has provided over $75 billion in aid to Africa between 2000-2011. Here we issue a firm warning to all data users: the available reporting on Chinese finance in Africa is far too nuanced and imperfect for sweeping generalizations. First, as explained above, our dataset relies primarily on imperfect media

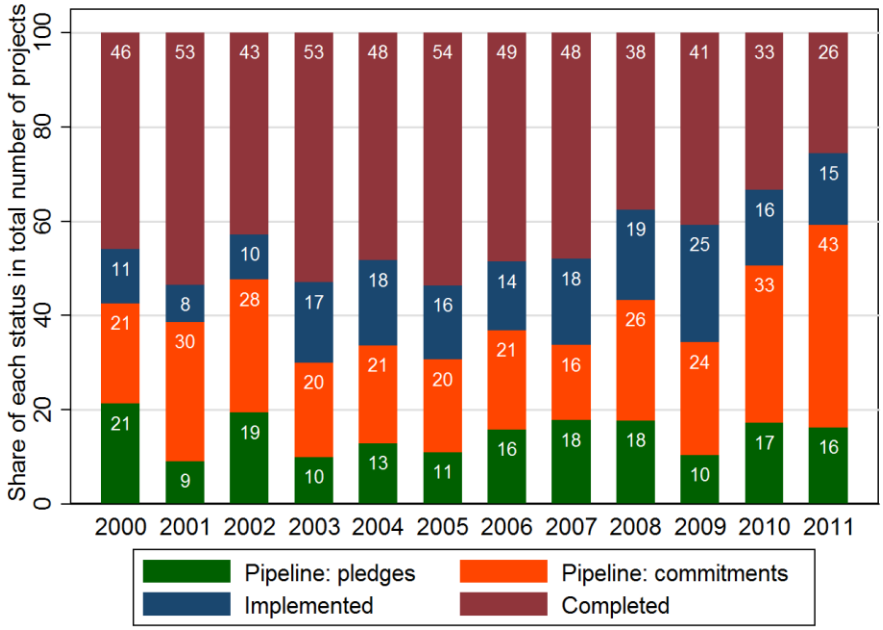
[7] Unfortunately, given that research on aid allocation and aid effectiveness has not benefited significantly from the use of media-based data collection methods, the existing literature does not offer much insight regarding whether, to what degree, and how detection bias might influence media-based aid and development finance data, and the inferences we draw based on such data.

[8] We have also attempted to improve upon previous media-based data collection efforts by requiring that all projects undergo a careful triangulation process that leverages English-, donor- and recipient-language sources and media reporting over a period of several years. On average, each official project record in our pilot dataset of official Chinese finance to sub-Saharan Africa relies on more than two (2.18) independent sources. AidData coders were instructed to "follow the money" from the project announcement stage to project completion. They did this by leveraging the information contained in initial media reports at the project announcement stage to conduct more targeted searches for project implementation information. When information from multiple media reports conflicted, the record was flagged as "suspicious". For these projects, an AidData staff member performed more thorough Google and Factiva searches for any additional corroborating sources. If no verifying evidence was found, they remained "suspicious" in the dataset. While the collection of multiple, independent sources does not completely eliminate issues of bias, our experience to date suggest that source triangulation substantially increases the volume and reliability of the resulting data.

[9] This figure includes all active projects in the database for which a formal commitment was reported in the media. It does not include projects that have been designated as "Pledged" only, "Suspended" and/or "Cancelled". Financial amounts are reported in 2009 U.S. dollars. All figures above exclude project records marked "suspicious". Official finance includes all forms of assistance from donor government agencies.

reporting. The result of adding imperfect figures reported in the media does not equate to perfect measures of Chinese development finance to Africa. Second, as displayed in Exhibit 1 below, a significant proportion of the projects in the database have reportedly been pledged or committed, but they have not necessarily been identified as implemented or completed.

Exhibit 1: Chinese development finance to Africa by project status, 2000-2011
Source: Strange et al. 2013



This dataset, which represents a major improvement in the granularity and comprehensiveness of existing Chinese development finance information, provides a resource that we hope will be used by scholars, policymakers, and journalists to track the distribution and impact of Chinese ODA, OOF, and commercial flows to Africa.[10] The procedures outlined in this methodology

[10] The DAC defines ODA as "[g]rants or loans to [developing] countries and territories … and to multilateral agencies which are: (a) undertaken by the official sector; (b) with promotion of economic development and welfare as the main objective; (c) at concessional financial terms (if a loan, having a grant element of at least 25 per cent). In addition to financial flows, technical co-operation is included in aid" (OECD DAC glossary). Members of the DAC have agreed that assistance to refugees, scholarships for developing country students, peaceful use of nuclear energy, and funding relevant research are included in ODA as well as specific types of peacekeeping, civil police work, and social and cultural programs. Military aid, anti-terrorism activities, peacekeeping enforcement, joint venture, and cooperative projects are excluded (OECD 2008). OOF is categorized as "[t]ransactions by the official sector with [developing] countries … which do not meet the conditions for eligibility as Official Development

also can be applied to research on other non-DAC donors, such as Venezuela, Iran, Russia, Cuba, and Saudi Arabia.

As mentioned above, we plan to continuously improve the database as users identify errors, help fill in missing information, and suggest the creation of new records. We also plan to complement MBDC by corresponding with Chinese and African government officials, project executives and managers, and local stakeholders. See Appendix C for more details on this approach, and visit the interactive website at www.china.aiddata.org to identify and flag shortcomings within our records. Instructions and tutorials for using the database are available on the website

Understanding Official Finance Flows from Non-DAC Sources

Solely using media reports to track development finance creates the risk of misinterpreting information, particularly for non-DAC donors who often administer development finance through modalities poorly understood by Western observers. For example, media outlets often report on Chinese "Megadeals" with various African governments. These multi-billion-dollar deals are usually announced with much fanfare at press conferences, but careful qualitative research has shown that a significant number of these deals are eventually cancelled, mothballed, or scaled back (Brautigam 2011a). Without making a concerted effort to "follow the money" from a project's initial announcement to its implementation, one runs a serious risk of over-counting (Brautigam 2011a). Similarly, strategic lines of credit as large as $10 billion dollars are cited in countless media reports as evidence of China's growing influence throughout Africa (Lum et al. 2009; Vines et al. 2009). These claims are not entirely baseless, but journalists and public intellectuals often commit a critical mistake: conflating the *availability* of a credit line and the *use* of a credit line (Brautigam 2010). Many lines of credit from the Chinese government are used sparingly or not at all. And this financing mechanism is generally not comparable to a lump sum donation or interest-free loan that would qualify as official development assistance (ODA) according to OECD standards (Brautigam 2011b). As a result, media-based data collection methods require more nuanced flow type designations. (Appendix D details the set of flow type categories used in AidData's MBDC China pilot.)

Assistance, either because they are not primarily aimed at development, or because they have a grant element of less than 25 per cent" (OECD DAC glossary).

Such challenges are not unique to Chinese aid. Non-DAC bilateral and multilateral institutions often support development finance and cooperation activities that defy conventional categorization and rarely adhere to traditional OECD reporting norms (Walz and Ramachandran 2011). For example, the Venezuelan government sometimes supplies oil at a discount to Cuba, and in exchange Cuba sends thousands of Cuban doctors and nurses to Venezuela to provide medical services (Jácome 2011). Other Caribbean nations have used agricultural commodities such as rice and beans to repay Venezuela for subsidized oil (James 2011). These examples of South-South cooperation activities challenge traditional definitions of "aid" and make it difficult to mechanically apply the traditional ODA and OOF definitions developed by the OECD.

The Two Stages of Media-Based Data Collection

AidData's MBDC methodology is divided into two stages.[11] 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, nongovernmental organization, or private foundation—are identified through Factiva, a Dow Jones-owned media database. Factiva draws on approximately 28,000 media sources worldwide in 23 languages. Most of these sources are newspapers, radio and television transcripts. In Stage Two, more specialized searches are conducted for projects initially identified during Stage One. Put simply, if Stage One is the equivalent of casting a wide net into the sea of media reports, Stage Two is closely examining the contents of the catch for more detailed information.

In what follows, we provide a step-by-step guide documenting how to conduct both stages of MBDC. Also, review Appendix A for a list of best practices on MBDC workflow, Appendix Hfor a visualization of how detailed project information is collected, and Appendix Gfor screenshots of each step in the search process. Finally, visit china.aiddata.org for AidData video tutorials and a list of frequently asked questions on conducting MBDC research.

Stage One: Conducting Media Searches to Identify Specific Projects

The first step in Stage One is to access Factiva.[12] Stage One searches are conducted for a single donor-recipient pairing at a time. Select both donor and recipient entities and establish a consistent set of "root search terms," separated by "or" for each country. For example, root terms for Libya include "(Libya or Libyan or Liby* or Tripoli)."[13] After the donor and recipient entities have been selected, connect both groupings of root search terms using "near5" in order to optimize results. "Near5" is usually preferable to other connecting terms, such as "AND", because the search yield will only include articles in which the donor and recipient root terms are located within five words of each other. Insert these terms into the Factiva "Free Text Search" box.

[12] A tutorial on the Factiva database is available online.
[13] Given that many development finance projects are announced in the capitals of recipient countries, it is often useful to include the recipient's capital city in the recipient root terms.

After the donor and recipient terms are inserted, enter a set of "aid keywords" into the search box as well as a specified date range. Unlike donor and recipient root terms unique to each donor and recipient, aid keywords are pre-assigned and universal for all searches in Stage One regardless of donor and recipient pairings. The full list of aid keywords are given in the example search below. To select a search year in Factiva, select the *Enter Date Range* option in the search box and enter January 1st and December 31st of the desired year as the start and end dates (e.g. "01/01/2010-12/31/2010").

Donor, recipient, aid keywords, and year are the four base search term groupings for Stage One searching. Other than the "near5" between donor and recipient root terms, groupings of base search terms are separated by "AND" in the search box. This means that within a media source, the base search terms can appear anywhere in the text and are not restricted to appear within five words around the donor and recipient root terms. Here is an example of base search terms:

(China or Chinese or Chin) near5 (Angola or Angolan or Angol* or Luanda) AND (assistance or grant or loan or concession* or donat* or donor or interest-free or interest or preferential or joint fund or invest* or finance or package or aid)*[14]

See Appendix G (Screenshot #1) for an example Stage One search input.

After clicking "Search," Factiva will reveal the magnitude of media reporting on the donor-recipient pair in question for a given time period captured by Factiva. Naturally, search yields vary in size depending on the scope and range of activities between the donor and recipient pair. Search yields merely represent the universe of news reports within Factiva, and not the true of amount of potential valuable media sources. Once generated, search yields should be sorted from "oldest to newest" to allow the coder to more easily identify and ignore duplicate reports.

The next step is to identify individual projects by extracting project-level information through examination of relevant media articles in the Factiva search yield. Throughout this initial process of turning search results into unique projects within a database, one should seek to "cast the

[14] Searches are conducted for media reports one year at a time, and thus "Year" is another base term. This variable needs to be set by the coder in Factiva, as stated above, but does not appear in the Factiva search box.

largest net possible." In order to ensure this "net" maximizes project-level data collection, it is important to:

1. Examine the title and passages where the base keywords are located within every article in the search yield to identify reports with potentially valuable and salient project-level information. Specifically, Factiva's search page automatically prepares a 2-3 sentence preview of each media report, showing the relevant section of the media report where base search terms are found. Scan source previews that display a cluster of search terms, which potentially reference a specific project (and not, for example, a list of countries present at a multilateral meeting). The resource requirements for this process depend on the search yields for various donor-recipient pairings. Some pairings may generate under 100 articles for one year while others (such as China and South Africa) may generate over 50,000. It is critical that coders do not alter their data extraction approach—they must go through the same motions for the entire search yield, regardless of the quantity of articles.

See Appendix G (Screenshot #2) for an example of a Stage One search yield.

2. Read carefully through reports that have been identified as potentially containing valuable data. Typically one will need to open several articles simultaneously in new tabs.[15]

3. Extract data and create a new project entry on AidData's online database[16] and pasting the entire contents of the news article into the box labeled "Project Description," including in-text citation of the article title in parentheses. Enter values for all variables in the database where quantifiable data is reported for a project.[17] All recipients should be listed for any article that mentions a project between one donor and multiple recipients.

4. Create a project entry for any and every potential project identified which *might be perceived as* a possible official finance project. This includes every project that is:

a. announced but not necessarily implemented

[15] Depending on the customized settings, Factiva automatically logs users out after five minutes of inactivity.
[16] AidData's web-based coding interface is available at china.aiddata.org. Researchers who are interested in replicating AidData's process need not use this particular interface to create and centralize project-level data. However, if one is working collaboratively with a team of researchers to build a single dataset, we strongly recommend use of a single, secure, customized, web-based database. Researchers who would like to explore the possibility of using AidData's web-based coding interface should send an email to info@aiddata.org.
[17] See Appendix I for a detailed list of database variables.

b. vaguely identified and unable to be confirmed by only reading one report

c. ambiguous as to whether it is an aid project or simply a non-concessional bilateral flow

See Appendix G, Screenshot #3 for an example of Stage One project data input.

Stage Two: Enhancing Data Quality Through Targeted Searches

After "casting the largest net possible" in Stage One, Stage Two entails searching for additional information about each project identified through Factiva, and then refining every project record based on all available information. This is done through public search engines including Google and the primary search engine used in the donor country.[18] Detailed searching considerably narrows search yields compared to Factiva, helping to pinpoint detailed sources related to specific projects and extract valuable data needed to fill remaining information gaps. In many but not all cases, Stage Two searching facilitates the accurate categorization of projects according to their concessionality, intent, and type of financing agent. It also helps reveal whether certain projects actually exist and have been implemented and completed, allowing coders to remove records that are not actually projects and avoid mislabeling a project that has merely been announced as one that has already occurred. During Stage Two coders must also eliminate duplicate records and add new projects that are identified while searching for finer-grained data on projects from Stage One.

The general strategy in Stage Two is to isolate a specific project by searching for *idiosyncratic terms* related to a project. Idiosyncratic terms may relate to the nature of the project, its capacity or financial value, sub-national location, sector, donor and/or recipient organizations involved, and/or precise announcement/start/end dates related to the status of the project (see Appendix G, Screenshot #4). For example, if a media report states that the China Overseas Engineering Corporation (COVEC) is implementing the construction of a hospital in Angola (ID #4), then "China Overseas Engineering Corporation," "COVEC," "construction," and "hospital" are all potentially useful search terms.[19] Update individual project records in the database as new information is found during Stage Two. See Appendix G (Screenshot #5) for an example Stage Two project record.

[18] Stage Two Google searches should be complemented by similar web searches through the primary search engine used in the donor country (see Appendix B). For the China pilot we used Baidu. Regional subsidiaries of Google, such as Google China (google.com.hk) and Google South Africa (google.co.za) also yielded results more targeted to some projects, but were not systematically searched.

[19] Ultimately, the process of filling data gaps in Stage Two should concentrate on "high priority" variables, as defined by the coder(s). For example, a research project focused on tracking Venezuelan-financed infrastructure in Latin America might care more about identifying the sector of every project entry. On the other hand, a research project estimating the total amount of Saudi Arabian foreign assistance given to Yemen may prioritize financial values over sector.

Follow the below guide for detailed step-by-step Stage Two instructions.

i. Stage Two Internet Searches

1. Locate Donor and Recipient Sources for Projects Identified through International Media Reports: Isolate all projects that have been identified in Stage One using an international media report as opposed to a donor or recipient media source. Attempt to identify additional donor-based or recipient-based sources confirming the existence of these projects. This can help increase credibility by verifying projects identified through sources outside the donor-recipient bilateral relationship.

2. Check for Official Recipient Reports:[20] Prior to beginning Stage Two Google searches for individual projects within a recipient country, check whether the recipient government publicly reports expenditures online. Search to see if the central government[21] publishes annual reports that show outstanding foreign debt or records of foreign aid receipt. Try searching for official records with terms such as "outstanding loans treasury site: [Official Government URL]" or "annual report finance site: [Official Government URL]." An example of this type of resource is Cambodia's Rehabilitation and Development Board Council for the Development of Cambodia, which publishes detailed project-level data on incoming Chinese assistance available at http://cdc.khmer.biz/index.asp).

3. Turn off Custom Results and Personalized Searches:
A. Clear and pause your Google web search history to stop "Customized Searches"
http://www.google.com/history
B. Turn off personalized results
http://support.google.com/accounts/bin/answer.py?hl=en&answer=54048

Employ the search tactics detailed below in steps 4 and 5 to determine the order of search keywords.

[20] This tactic was discovered after the MBDC China pilot project was already underway and thus was not applied comprehensively to the creation of the initial dataset (1.0).
[21] Of course, even if a country does upload its public expenditure records, the relevant ministry or agency that houses these data vary from country to country. For example, the Ministry of Commerce is the most central ministry with respect to foreign aid, while in India it is the Ministry of External Affairs.

4. Follow the general order of search keywords: Coders should follow this general order of search keywords below when performing Stage Two on Google and other search engines:

Donor Country *(Can be left out if have donor organization or official)* + **Recipient Country** *(Can be substituted for capital or left out if have subnational location)* + **Project Name** + **Organizations Involved in Project** + **Subnational Location(s)** + **Contact Name(s)** + **Number/Amount(s)** + **Year**

Criteria for Term Inclusion in Project-Specific Search

In order to minimize the number of irrelevant sources in a search yield, follow these criteria to determine which idiosyncratic terms should be employed to search for a given project ID:

- If one has information about the development finance institution and/or specific officials that can be used in the search, then one should include that information and leave the donor country name out of the search. However, if the database indicates that the development finance institution identified has multiple projects in one recipient country, then it may not be helpful to add the organization name as a search term. This means the organization name is no longer an "idiosyncratic term"). In such instances, the organization name may be excluded and the "donor" name included.

- If one has information about the specific project name, subnational location(s), recipient organizations and/or officials that can be used in the search, then one should include that information and leave the "donor" country name (e.g. China, Venezuela) out of the search. However, if the subnational location identified is the capital or a city with a high level of donor and/or development finance activity, then it is generally not helpful to use the location as a search term, and this variable may be excluded. The "donor" country name should be retained .

After submitting your search terms, scan the search yield for potential sources of project-level information, just as you did in Factiva during Stage One. Once Stage Two searches no longer reveal useful information for the project, begin to input data into the project record. These project data include: project description, flow type and flow class (see Appendix D), donor intent (Appendix F), year, transaction amount, CRS sector (Appendix I), and links to all sources which contained valuable information.

Additional Complementary Steps

Steps 5-7 below were discovered to help streamline and focus Stage Two searches, but they were not applied systematically from the start of the MBDC China pilot.

5. Connect search terms inputted into Google: The structural design of Stage Two Google searching should parallel Stage One Factiva searching. The "AROUND(#)" function of Google operates similarly to the "near" feature in Factiva (see page 9), and as such should connect Stage Two search terms.[22] For example:

"china OR chinese AROUND(3) niger OR nigerien" AND "Niamey" AND "power" OR "electricity"

This returns results with "china" or "chinese" within three words of "niger" or "nigerien" AND the exact word "Niamey" AND any of the "OR terms" exactly as they are typed.

"china OR chinese AROUND(3) niger OR nigerien" AND "Niamey" AND ~power OR ~electricity

These search criteria return results with the country terms AND the exact word "Niamey" AND any of the "OR terms" (as well as synonyms of "OR" terms).[23]

*Note: you *cannot* include spaces in the donor or recipient "OR terms" in Google searches.

6. Use date inputs in Google: If more than 20 unique sources are found in the search yield, the following methods can be applied to target certain time periods:

A. If you know that a project occurs within a certain timeline, or your search is being 'overshadowed' by a similar project within the same timeframe:[24]
a. On the left-hand side of your search page, click "Show search tools"
b. Click "Custom range..."
c. Enter the dates you want to search within
d. Choose whether to sort by relevance or by date
* *Note: This search will exclude results without date information*

[22] This feature of our methodology was added during the pilot project, and many Stage Two searches did not use the "AROUND(#)" function in Google. Early searches were performed using "AND".
[23] See http://googlesystem.blogspot.com/2010/12/googles-around-operator.html for details.
[24] Of course, a media report from one year often provides data on projects started and completed in other years, so it is important not to confine results too much. This technique should mainly be used to get around obstacles such as "overshadowing."

B. If you are looking for the most recent information on a project to determine its current status:

a. On the left-hand side of your search page, click "Show search tools"
b. Choose whether you want to search within the past year, month, or week
c. Click "Sorted by date" to get the newest results at the top

**Note: This search will exclude results without date information*

7. Search within websites: If you discover a website which requires special examination for project information:

a. After you have entered your search term into Google, type "site:" (without the quotes) followed immediately by the URL of the website you want to search for information. Example search term: "Hu Jintao Seychelles loan site: www.gov.mu."

ii. Writing Project Descriptions in Stage Two

Below is a basic template for writing project description paragraphs used by AidData coders:

- **Sentence 1:** Date of agreement and nature, scope, location, and overall cost of project
- **Sentence 2:** Who (organizations and individuals) is doing what? What are the financial details?
- **Sentence 2b:** *(If applicable)* Unique piece of information about project that does not fit nicely into the data collection interface (e.g. China is not the only funder.)[25]
- **Sentence 3:** What is the current state of the project according to the most recent sources? What key details still need to be obtained?
- **Sentence 4:** *(Separate from main paragraph, if necessary)* Is this project connected to any other projects?

Note: Even for vague projects it is still necessary to write the project description. For especially vague projects, a 1-2 sentence summary will suffice. Similarly, complex project IDs, such as an agreement for a multi-billion dollar loan encompassing many unique projects, often require multiple paragraphs to adequately document. Paragraphs should be written once the project variables are established using all available data sources.[26]

Example Project Description: ID #17832

Sentence 1: In 2011, Kenya secured a grant worth 1.5 billion KES from the China Development Bank to increase housing units in Kenya.
Sentence 2: The funds were planned to be used to put up 800 residential units in Kisumu and hostels for female students at Masinde Muliro University in Kakamega.
Sentence 3: The Ministry of Housing said they were seeking contractors to complete construction within a year, but the current status and start and end dates are unknown.
Sentence 4: May be linked to project ID #810.

[25] At the start the MBDC China pilot initiative, coders were instructed to collect information on whether projects were "tied" to the purchase of goods and services in the donor country, and input this information into the project description. As it became clear, media sources generally did not report whether a Chinese project was "tied" or not, this variable was dropped and no longer collected. However, this information remains in project descriptions for many records.
[26] If an ID already contains more than one source, it is generally most efficient to combine existing sources for a project ID and write the project description paragraph before actually performing Stage Two. This ordering allows the coder to first identify the information gaps and then proceed with Stage Two searching accordingly.

iii. Data Entry FAQs:

How should I handle "omnibus" agreements that comprise several smaller subprojects?
When dealing with a large loan or grant that will cover multiple projects, you may only break up the "omnibus project" record into smaller records if you have amounts for all of the smaller projects.[27] Otherwise, one incurs the risk of double-counting certain components of the financial commitment. There must be explicit evidence that each smaller project is indeed a component of the larger "omnibus" project. These subcomponent amounts must sum up to the value of the omnibus project. If there is not enough information to split the project, retain the same omnibus project and write a longer project description which details each known subcomponent.

- ***Split the package into its constituent subcomponents when…***
    - You find amounts for each subcomponent project, and these parts add up to the total package amount

- ***The package deal ("parent project") subsumes a subcomponent ("child project"), creating a merged record when...***
    - There is **explicit evidence** in a source document suggesting the subcomponent project falls within the larger package agreement
        - In this case, the child project record is deactivated

- ***The package deal and subcomponent records are kept separate when…***
    - No evidence can be found linking the subcomponent to the larger package (same sector and same funder does not count as evidence, as they may be separate commitments)
      --OR--

[27] An omnibus project may also be known as a "package agreement," or a "parent project" containing multiple "child projects." In these deals the total sum of all project amounts is usually committed by the donor, thus receiving a status of "Pipeline: Commitment." However, "framework agreements" were treated as non-legally binding pledges, in that the sum of all project amounts was an agreed guideline, but not a firm commitment by the donor.

- There is evidence linking all possible subcomponent projects to the larger package, but the combined values of the subcomponents do not match the reported value of the total package agreement
    - In this case, possible subcomponent/package projects are linked in the description, with a warning on possible double-counting

What should I do with projects that have multiple donors?
When you identify a project with multiple donors, do not remove it from the database. Indicate in the description that it is not bilateral; these projects can be isolated into a separate category and used for other research purposes.[28] If the project has multiple donors and the individual donor commitments cannot be isolated, mark "is cofinanced."

What should I do with projects that have multiple recipients?
When you identify a project with multiple recipients during Stage One or Stage Two searching, (a) take the financial amount out of the "Amount" box if this has not already been done keeping the financial information in the "Description" field and (b) in addition to individual countries, add "Africa, regional" as a recipient. Depending on the focus of the data collection project, additional "regional variables" may be possible, such as "South America, Regional" or "Eastern Europe, regional."

Once you have performed these tasks, conduct Stage Two searching for each recipient country mentioned in the project record, adding any additional sources to the record. After you have completed Stage Two searching, indicate in the "Description" field if you have been able to attach a specific financial value to the portion of the project going to the selected recipient, making sure to also include the original total amount for the multi-recipient project in the project description. At the conclusion of Stage Two, a program manager will reassess multi-recipient project entries on a case-by-case basis and disaggregate such projects as information allows.

How can I track project status when the completion of some projects is unlikely to be publicly announced?

[28] This is an issue of preference and thus only a guideline for AidData coders using this codebook; some researchers may simply choose to discard projects with multiple donors.

For some forms of official finance, the status of the "project" is difficult to ascertain as media sources typically only report the announcement, not the completion, of financial transactions. For "one-off" financial transactions, such as debt relief and debt rescheduling, the "project" in question should be coded as "Completed" even if a completion announcement cannot be found. This is the default status for these "one-off" project types unless coders identify explicit evidence that the transaction has been cancelled or not honored for some other reason. Such transactions are relatively instantaneous in nature compared to projects that involve the provision of goods and services that take a finite period of time to complete, such as the construction of a hospital or the deployment of a medical team. We acknowledge this practice introduces the possibility of over-counting the amount of completed projects and development finance. For all projects that require multiple financial transactions or the provision of goods and services, the project status should be noted as "pipeline" if only the announcement can be found. See Appendix E for guidance on how to distinguish between "Pipeline: Pledge" and "Pipeline: Commitment".

What projects should be "linked" in "Project description?"
Do not only link multiple projects that are announced at the same time—e.g. a single announcement that introduces three distinct cooperative agreements. Also link inter-related projects that have a connection which could improve our understanding of each of the linked projects. For example, one project may be the direct or indirect result of another project (e.g. success in of official financing for constructing a Nigerian power plant in 2002 [ID # 27948] led to upgrades of the facilities through foreign direct investment in 2010 [ID #28087]).), or one project ID may be for a single commitment within an "omnibus" agreement that has been split into multiple records. In cases such as these, the records should be "linked."

iv. Finding New Projects through Stage Two Searches

Undertaking Stage Two searches for a previously identified project often results in information about a new project that was not identified during Stage One. When this occurs, we recommend returning to Factiva and searching for additional information regarding these "snowballed" projects. New projects should receive Stage Two search treatment in both Factiva and Google (as well as the relevant donor search engine).

v. Data Quality Assurance

As previously discussed, MBDC has several flaws related to the imprecision of media reporting and human error. Data quality assurance exercises can help identify and correct errors. Below is a discussion of the methods employed by AidData to help correct deficiencies during the MBDC China pilot.

AidData staff performed rigorous data quality assurance exercises following the conclusion of Stage Two searching. For the purposes of our pilot study of Chinese official finance to Africa from 2000 to 2011, all official and unofficial projects with reported financial commitments greater than or equal to $1 billion USD were reviewed for accuracy and missing information, either by a program manager or a research assistant who did not perform the Stage One or Stage Two searches for a given record. In this sense, billion-dollar project records were examined by three individuals: one at each stage of data collection, and a third-party reviewer. All records greater than $5 billion USD were reviewed for accuracy and missing information by a program manager.[29] Project records initially marked as "suspicious" during the Stage Two searches were re-examined by a program manager or a third research assistant and updated as needed.

For these "high-value" projects, staff reviewed and augmented existing records to ensure that:

- All fields were filled as completely and accurately as possible.
- Media sources provided corroborating evidence and did not conflict.
- The description and financial amount of the project accurately represented the information contained in the media source related to that project.
- The project description was detailed and correctly formatted.
- There was evidence that a formal agreement to execute the project was signed.
- There was no evidence the project was suspended or cancelled.
- The official finance transaction reported in the record did not contain other project records, nor was the official finance transaction contained in another record.[30]

[29] Billion-dollar projects were split into smaller records for each sub-component when exact amounts for each of sub-component could be confirmed through targeted searches.
[30] If the project had multiple components, the various components were not broken out into distinct projects. If we encountered clear evidence that two records, record A and record B, represented the same project, unique

After all variables were inputted, a second coder performed another round of coding for the "flow class" variable. All projects with two conflicting code-rounds for the "flow class" variable were arbitrated by a third coder or program manager.

After all coding exercises were completed, coders performed a final round of verification exercises for the largest 120 active official project IDs in the database, ranked by committed financial value, to ensure these projects were coded accurately on all variables. A team of two coders independently reviewed the accuracy of each project's description, source documentation, intent, flow_class and flow_type codes, and currency conversion. During the discussion of each project, they were required to collectively agree that all variables were accurately recorded before moving to the next project.

For complex project records, staff has inserted metadata to provide further details on creation of the project record itself. These metadata begin with "STAFF_NOTES:" in the project description.

Coders and staff re-examined information on the origin and types of organizations listed in the database, and applied them consistently across the database. The origin and type of organization is necessary to distinguish official and unofficial finance. All organization details were determined through Google (US) searches, and relied on either a primary source (e.g. organization's website) or secondary source (e.g. academic or new article stating the nature of the company). We admit that this is an imprecise approach, however the true origin and stakeholder composition of many entities involved in development assistance, such as Chinese enterprises, is often murky at best. We acknowledge that due to the inexact process by which organizational data were collected, some of our organization categorizations are likely flawed. Organization details that could not be determined were left blank.

Specific to our China pilot initiative, Deborah Brautigam's well-known blog *China in Africa: The Real Story* (www.chinaafricarealstory.com/) provided a critical source of project-level information. As a widely read, publicly available resource on Chinese flows to Africa, the blog regularly collects and publishes technical project details (such as loan terms and implementation

information in record A was merged into record B, and record A was marked "inactive." The goal of this exercise is to avoid double counting while maintaining a valid, comprehensive scope.

status) from reports, interviews, and online commenters. AidData staff cross-checked media-sourced data with every Brautigam blog post from January 2010 onward, and if necessary, modified record details for any projects described in the blog. In many cases her blog contains valuable information from interviews with officials in Africa as well as other field work, and is therefore not found in media reports during the MBDC process. We generally treated the blog's project information, much of which, as authoritative, unless our team uncovered a media update that was published after one of her blog posts and provided new updates about a project record. From this experience, we recommend that future iterations of MBDC approach these extensive repositories of qualitative research on development finance projects in a similar fashion.

In addition to these data quality exercises, AidData staff completed a series of "cross-checks" on the final database to discover potential strengths and weaknesses of our data collection methods.[31] Official aggregate and project-level data on Chinese official finance is scarce. AidData compared the media-sourced dataset with available official figures, public information from multilateral and recipient agencies, and select studies, including:

-China's MOFCOM Yearbooks from 2000-2005
-Humanitarian aid data recorded in the Financial Tracking Service (FTS)
-Food Aid Information System (FAIS)
-Malawi's Aid Management Platform (AMP)
-Various academic studies on Chinese aid

[31] For further discussion of these cross-checks, see Strange et al..

References

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).

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.

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.

Hendrix, Cullen, Idean Salehyan, Jesse Hamner, Christina Case, Christopher Linebarger, Emily Stull, and Jennifer Williams. forthcoming, Social Conflict in Africa: A New Database. *International Interactions*.

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

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.

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.

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.

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

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.

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: Best Practices for Maximizing MBDC Workflow Efficiency

The following activities were undertaken in order to make efficient use of scarce resources during the MBDC pilot. We consider these activities to be "best practices" for implementation of the MBDC methodology, but they are not necessary to ensure accurate or comprehensive data collection.

1) Time Limits for Project Searches

After initiating the process of Stage Two, a Principal Investigator or program manager should evaluate the overall resource envelope at his or her disposal and identify the point of diminishing returns for collecting information related to individual project IDs. While most projects can be thoroughly searched in less than 15 minutes, some relatively complex projects will require 30 minutes or more before valuable information is uncovered. After reaching the established time limit, the search for additional information related to a given project ID should be abandoned and flagged for review by a program manager.

2) Select Projects in the Database for Stage Two Searches

Stage Two searches should be performed according to the relative priority of each project's Stage One classification. For the MBDC-China Pilot, searches were broken into three discrete "waves", based on the status and size of projects identified in Stage One. That is, projects with a reported financial commitment above a threshold of $1 million were identified as priority projects that should be searched on before smaller projects. Similarly, projects with no evidence of reported implementation or completion were given priority over projects already reported as having been started and completed. The following are instructions received by AidData coders for prioritization of Stage Two searches.

After clicking "Search and Filter" on the database main page, add a filter for the "recipient" country of your consideration. Then identify priority projects by wave, beginning with first wave. As new information is found, it should be inserted into the appropriate spaces within the database.

- **First Wave:** Add the additional filter "Pipeline: Pledge" and "Pipeline: Commitment" under "Status." Within search results, sort pipeline projects by highest USD 2009 financial amount. Prioritize pipeline projects with the highest reported financial amount, except for projects

where commitments occurred in 2010 or 2011.[32] Also, do not search on projects with reported values of less than $1 million or with multiple recipients during initial Stage Two searches. These prioritized projects constitute the "First Wave" of Stage Two searching.

- **Second Wave:** Add the additional filter "Pipeline: Pledge" and "Pipeline: Commitment" under "Status." Within search results, prioritize all pipeline projects that are listed as "Raw" (meaning they have not received Stage Two treatment) except for projects where commitments occurred in 2010/2011.
- **Third Wave:** Within search results, prioritize all projects with a status other than "Pipeline." This should include all projects with a status designation of Cancelled, Suspended, Implementation, Completion, and Null (No entry). Also, perform Stage Two for all remaining "Raw" projects.

3) Supplemental Workflow Efficiency Activities

- **Upon completion of Stage Two for a project, mark the ID as either "checked" or "suspicious":** The "suspicious" button should be selected when the coder has information about a project but has reason to doubt the accuracy of the data source and/or content. Projects identified through "suspicious" sources are flagged for review during the final stages of data cleaning and review by the Principal Investigator or program managers.

- **Contribute to the Organization Glossary and Data Pool:** Stage Two also involves identifying the nature of donor and recipient agencies involved in various projects. After researching an organization to determine its type and origin, record your results in a pooled data document so that other coders can benefit from these findings.

- **Tracking Stage Two Progress:** Once you have searched on all projects qualifying for a given "wave" of searches within a recipient country, mark how many projects you cleared from that wave in a team progress chart. Cleared projects are those in which you were able to update the project status from "Pipeline/Identification" to a more precise classification.

[32] This is because projects reported most recently are less likely to have additional "follow-up" reports.

Appendix B: Supplementary Notes on Chinese Language Searches

1) **Use Baidu instead of Google**
While conducting searches on Chinese search engine Baidu, AidData coders found that the "News" or "新闻" tab is rarely helpful for data enhancement purposes. Stick to searches conducted through the "Web" or "网页" tab.

2) **Be careful when inputting Chinese search terms for several variables**
a. **Village and city names in recipient country**: Some sources write these differently. Typically country names and capital cities are uniform. Using one Chinese name for a village may preclude valuable results with a different name for the village.
b. **Name of project**: The Chinese project name can vary across sources as well. Using one name can produce a similar negative result as above.
c. **Recipient official names**: Again, these can be written differently depending on the choice of Chinese characters.
d. **Company abbreviations**: Some websites/articles may use an abbreviation for a Chinese and/or recipient organization. Baidu may pick some but not all of these up.
e. **Subsidiaries of companies**: Sometimes subsidiary/parent companies are used in place of parent/subsidiary companies, respectively, so it is crucial to have knowledge of what entities a given firm is connected to.

3) **Finding exact Chinese names through the web**
When specific people, bank, firms, corporations, places are mentioned in project, type their English names in *baidu.com,* search them, and find out the exact Chinese names they have. Don't translate the English names yourself. For example, China Development Bank can be easily translated to 中国发展银行, but its real official name in Chinese is 国家开发银行. If the incorrect Chinese name is used in searching for projects, finding the right project will prove very difficult.

4) **Exclude names of well-known figures**
If a well-known figure is connected to a project (e.g. Chinese Foreign Minister Yang Jiechi attended a ceremony to mark the construction of a new highway), it is better to not use their names when searching for Chinese projects, since many other news articles will result and make the searching process difficult by flooding search yields with irrelevant information.

Appendix C: Personnel Outreach and Crowdsourcing

Contingent on available resources, 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. Coders 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).

In addition to the static dataset, AidData is releasing a dynamic project data platform, 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.[33]

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.

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

Appendix D: Flow Type and Flow Class

All projects in the database should be classified according to their level of concessionality and type of financing provider. MBDC is designed for tracking official development finance flows, which include grants, technical assistance, concessional and non-concessional loans, debt relief, export credits, and other financial instruments. Therefore, each coder must confront the issue of whether a project is official or unofficial[34], and whether it is concessional or not. Projects are separated into flow_type and flow_class categories, as follows.[35]

ODA-Like: Flows of official financing administered with the promotion of the economic development and welfare of developing countries as the main objective, and which are perceived to be concessional in character with a grant element of at least 25 percent (using a fixed 10 percent rate of discount). The following flow types should be categorized as "ODA-like" if they 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

> Example: Project ID# 549: The Chinese government provided an official grant to Egypt to establish a school.

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

[34] The flow class coding scheme attempts to distinguish between official and unofficial sources of finance. However, individuals present at the signing ceremony may hold multiple positions within the government, Chinese Communist Party, state-owned enterprises, or foundations. 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.

[35] Projects with any of these flow categorization labels are subject to alteration upon identification of new data that changes the nature of the project's resource flow.

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 Chinese 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 Chinese government institution to a recipient institution (Brautigam 2011: 206). OOF activities also include "short-term credits to Chinese exporters (export sellers' credits) to help them finance foreign sales, and ... longer-term credits to foreign buyers to assist in the export of Chinese goods and services" (Brautigam 2011a: 206). OOF also includes lines of credit that the Chinese government provides to a Chinese enterprise (state-owned or not-state-owned) to do business overseas.[36] These projects may include any type of donor intent: development, commercial, representational, or mixed.

> ***'Line of Credit' Marker Variable*:** In addition to coding lines of credit as 'OOF-like', coders should tick the 'line of credit' check box 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 should not be ticked. Once the line of credit has been completely used by the recipient, it becomes outstanding loans, rather than an outstanding line of credit.
>
> Example: Project ID# 434: Exim bank, an official entity of the Chinese government, provided sellers' credits to China National Machinery and Equipment Import and Export Company to complete a dam in the Republic of Congo.
>
> Example: Project ID# 21057: In April 2004, the Zimbabwean government came to an agreement with the state-owned China Development Bank (CDB) for the provision of a US$30 million line of credit to be directed towards the Zimbabwean agriculture industry. In order to identify the individual projects to be funded, the CDB is sending a team of experts during the second half of 2004 to make an assessment of the situation of Zimbabwean agriculture. Also included in the agreement was that the CDB would receive a stake in the Infrastructural Development Bank of Zimbabwe (IDZB), another state-owned organization. Once the funds are made available to Zimbabwean farmers,

[36] 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).

they will be able to access funds through the Agricultural Bank of Zimbabwe. The government of Zimbabwe was represented by the Minister of Finance, Tendai Biti.

Grant Element: The OECD characterizes a loan as ODA if the loan's grant element at least 25%. To ensure comparability with OECD data, AidData used the OECD's grant element calculator (available here: http://www.oecd.org/investment/stats/15344611.xls) to measure the concessionality of Chinese loans. These loans were 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.

However, 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%. This has clear consequences for researchers attempting to classify non-DAC financial flows as either ODA or OOF.

For the purposes of transparency and flexibility, AidData is currently working to provide both IMF and OECD grant element calculations on records of Chinese overseas loans.

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 subsumes ODA-like and OOF-like flows. These projects may have "development" or "mixed (some development)" intent.

Example: Project ID# 53: China Exim bank provided a "concessional" loan to state-run Sierratel to install a telecommunications system in Sierra Leone. It is uncertain whether the loan's degree of concessionality qualifies it as ODA or OOF.

Official Investment: An international investment 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.

> Example: Project ID #21504: "China Development Bank (CDB) acquired a majority stake in the Infrastructure Bank of Zimbabwe (IDBZ); the Chinese bank will inject long-term capital, notably lines of credit, into IDBZ and subsequently into Zimbabwe's energy, transport, and infrastructure sectors."

Joint Ventures with Chinese state involvement: An association between a state-owned enterprise and a (private or state-owned) firm in another country is formed to undertake a specific business project. Joint ventures are similar to a partnership, but they are limited to a specific project, such as producing a specific product or doing research in a specific area. Joint ventures are formed when two or more parties form and share equity in and ownership of an enterprise. These projects must have "commercial" intent. For the official OECD definition of a joint venture, see http://stats.oecd.org/glossary/detail.asp?ID=1028

> Example: Project ID# 209: "The state-owned China National Petroleum Corp (CNPC) acquired a 60% share in a project to build an oil refinery in collaboration with Chad's government."

Joint Ventures without Chinese state involvement: An association between a private Chinese enterprise and a (private or state-owned) firm in another country is formed to undertake a specific business project. Joint ventures are similar to a partnership, but they are limited to a specific project, such as producing a specific product or doing research in a specific area. Joint ventures are formed when two or more parties form and share equity in and ownership of an enterprise. These projects must have "commercial" intent. For the official OECD definition of a joint venture, see http://stats.oecd.org/glossary/detail.asp?ID=1028

> Example: Project ID# 2344: "A private Chinese company Zhongying Changiian International Investment Guarantee Ltd. and Zambian Kaidi Biomass Development PLC established Kaidi Biomass Zambia Ltd., and in conjunction with the Zambian government signed an agreement to produce various biofuels. There is no evidence of Chinese government entities or SOEs being involved in the project."

FDI with Chinese state involvement: An international investment that reflects the objective of a state-owned enterprise to obtain a lasting interest in private or state-owned enterprise resident in another country's economy. For the official OECD definition of foreign direct investment (FDI), see http://stats.oecd.org/glossary/detail.asp?ID=1028. FDI refers to process of an investor of a foreign country acquiring an asset in a host country and/or controlling the production and/or distribution of products or services in the host country. FDI

projects give the foreign investor some type of ownership stake or decision-making power (Djankov 2000). These projects must have "commercial" intent.

> Example: Project ID# 16439: "In 2007, China's state-owned Industrial and Commercial Bank of China Ltd. bought a 20 percent stake in South African Standard Bank Group Ltd. for $5.6 billion USD. It is the largest single foreign investment in Africa to date.")."

FDI without Chinese state involvement: An international investment that reflects the objective of a private Chinese enterprise to obtain a lasting interest in a private or state-owned enterprise resident in another country's economy. For the official OECD definition of foreign direct investment (FDI), see http://stats.oecd.org/glossary/detail.asp?ID=1028**.** FDI refers to process of an investor of a foreign country acquiring an asset in a host country and/or controlling the production and/or distribution of products or services in the host country. FDI projects give the foreign investor some type of ownership stake or decision-making power (Djankov 2000). These projects must have "commercial" intent.

> Example: Project ID# 1796: "In 2001, the Anshan Cotton and Dyeing Factory in northeastern Liaoning Province built a clothing factory in Morocco. The factory director was Xu Xu. The project was expected to be completed in June 2001, and required an investment of $1 million. All of the processing machinery for the factory was imported from Japan and the Republic of Korea. The status of the project is uncertain."

Vague (Residual Commercial Activities): Other commercial activities undertaken by Chinese government institutions, state-owned enterprises or private enterprises that do not fit within any of the following categories: "FDI with Chinese state involvement", "FDI without Chinese state involvement", "Joint Ventures with Chinese state involvement" and, "Joint Ventures without Chinese state involvement". These projects may have "commercial" or "mixed (no development)" intent.

> Example: Project ID# 2457: "On September 27, 2010, Sudan launched its first bioenergy project in Tabat area, White Nile State, some 360 km south of Khartoum. The project was implemented by Chinese Xing Ye Company and includes 25 units that could translate animal wastes into light. The ceremony was attended by the Chinese ambassador to Sudan Li Chengwen, representatives from the White Nile state government, including Mohamed Babikir Shinaibop, Minister of Agriculture of White Nile State, and the Sudanese Women General Union. The flow type, amount, and start date are unknown. Further searches could not confirm the source of funding for this project."

NGO aid: Flows of financing by non-governmental organizations administered with the promotion of the economic development and welfare of developing countries as the main objective, and which are concessional in character with a grant element of at least 25 percent (using a fixed 10 percent rate of discount). These projects may have "development", or "mixed (some development)" intent.

> Example: Project ID# 1016: The Red Cross Society of China donated emergency aid of $30,000 to Madagascar, which was hit by cyclones Fame and Ivan on January 28 and February 17, respectively.

Corporate Aid with Chinese State Involvement: Flows of financing by state-owned enterprise (SOEs) administered with the promotion of the economic development and welfare of developing countries as the main objective, and which are concessional in character with a grant element of at least 25 percent (using a fixed 10 percent rate of discount). These projects may have "development" or "mixed (some development)" intent.

> Example: Project ID #20955. "On May 28. 2008, China National Petroleum Corporation and South Sudanese government officials signed a support agreement in which CNPC donated 700,000 USD to set up an education and training fund in Juba University. In that fund, 200,000 USD was to be used for teaching and reference books for the University libraries and 500,000 USD for awards for excellent teachers. The current status and start and end dates are unknown." This is an aid flow (education and training fund) financed by a Chinese state-owned enterprise. It is not an official flow (ODA or OOF), and is not FDI (because there is no obvious objective to obtain a lasting interest in an enterprise resident in Sudan). As such, it should be classified as corporate aid (SOE).

Corporate Aid without Chinese State Involvement: Flows of financing by private enterprises administered with the promotion of the economic development and welfare of developing countries as the main objective, and which are concessional in character with a grant element of at least 25 percent (using a fixed 10 percent rate of discount). These projects may have "development" or "mixed (some development)".

> Example: Project ID #24866. "In 2010, Chinese firm Huawei donated $10,000 USD in cash and materials to a local NGO in Ghana to support cancer treatment and prevention." This is an aid flow financed by a private enterprise. It is not an official flow (no ODA, OOF), and is not FDI (because there is no obvious objective to obtain a lasting interest in an enterprise resident in Ghana). As such, it should be classified as corporate aid (non-SOE).

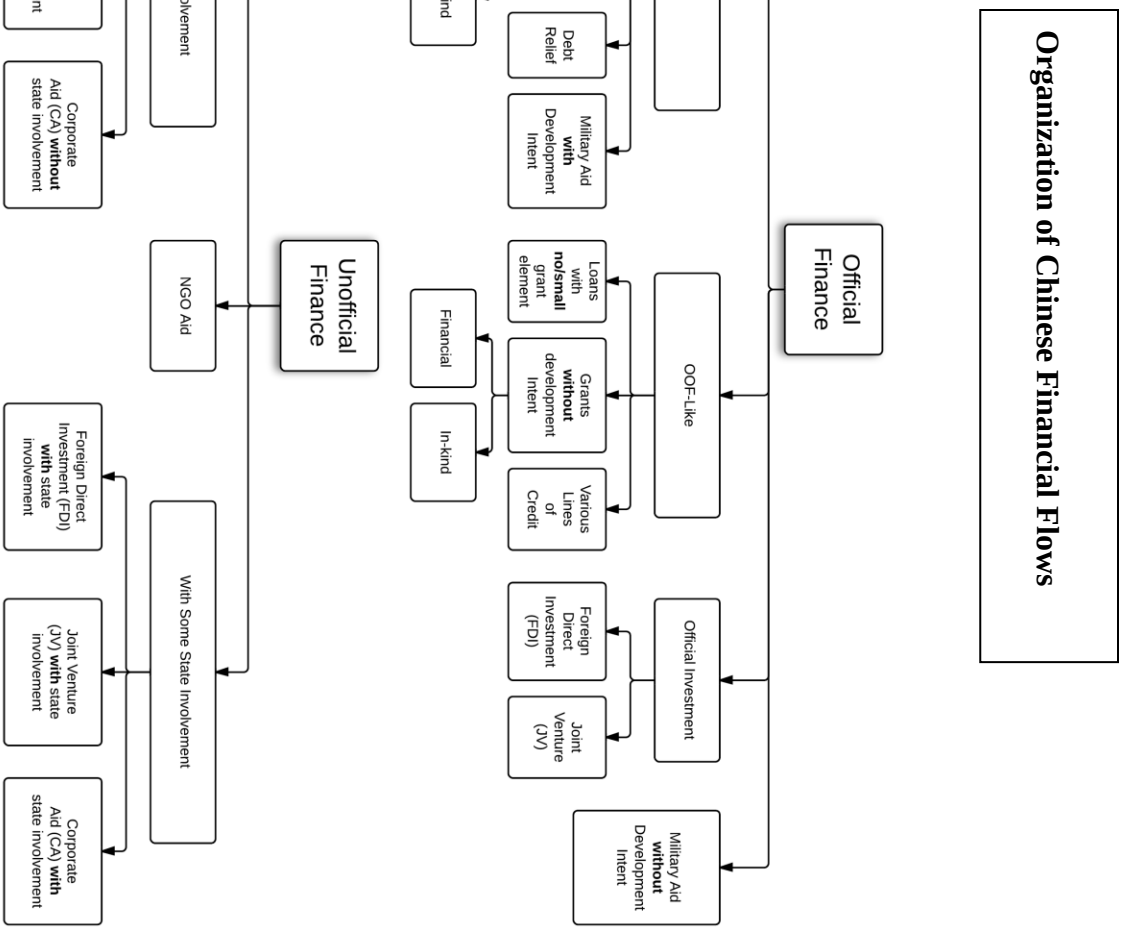
Military Aid: This category refers to the provision of military equipment or services, including peacekeeping services that do not have a developmental focus, from a Chinese government institution to a recipient institution. Military aid does *not* include activities that could receive the 152 OECD sector code ("'Conflict prevention and resolution, peace and security") designation. For example, weapons disposal, mine removal, military debt forgiveness, and repatriation and demobilisation of soldiers should receive the 152 OECD sector code, but these activities should not be classified as military aid. Unofficial aid for military purpose (e.g. donations of weapons from a private company) is Vague (Residual Commercial Activities). Military aid also excludes the cost of using Chinese armed forces to deliver humanitarian aid or perform development services. These projects may have "representational" or "mixed (no development)" intent.

> Example: Project ID# 26375: "In 2005, China provided the armed forces of Zimbabwe with 39 military trucks, educational and medical equipment worth more than $3 million USD. According to a report from the University of Stellenbosch, the facilities are part of a $3 million USD grant." The direct transfer of military equipment does not qualify as ODA under the 152 sector code.

Flow Class Coding Scheme

	Official Finance (from a state entity)				
Flow_class	**ODA**	**OOF**	**OF-Vague**	**Official Investment**	**Military**
Intent	Development Only	Any	Devt, Mixed (Some Devt)	Commercial only	Rep, or Mixed (No Devt)
Funding Source Type	**Any Official** (Government Agency, Policy Banks, Embassies, Party (CCP) and party controlled orgs, Provincial Govts)	**Any Official**	**Any Official**	Government agencies or Official Policy Banks: CDB, CADF, ADBC, EXIM	Government Agency
Flow_Type	Grants, technical assistance, interest-free loans, in-kind contributions of goods and services, debt relief, debt rescheduling, military aid for development/humanitarian purposes, LOANS WITH <2% INTEREST OR >25% GRANT ELEMENT	Grants for Comm, Rep, or Mixed (No Devt) purpose; Loans with grant element LESS THAN 25%; All Lines of Credit including Export credits, Export Seller's credits, Strategic/Supplier Credit, and Credit to Chinese enterprises for overseas work	Cannot be placed with certainty into either ODA or OOF category. Includes all remaining Mixed (some development) projects, loans with unclear concessionality, and Vague-TBD flow type	FDI or JV. The government financier must be buying a stake and seeking a return on this investment.	Grants, loans, technical assistance, or in-kind
Special Rules	If not enough info to calculate an official loan's grant element, OF-vague	Mark "Line of Credit" box if needed. "Funds in support of private investment" may fall under Official investment			Military aid for development/ humanitarian purposes is ODA

	Unofficial Finance				
Flow_class	**JV (+/-Gov)**	**FDI (+/-Gov)**	**CA (+/-Gov)**	**NGO aid**	**Commercial-Vague**
Intent	Commercial	Commercial	Development, Mixed (Some Devt)	Development, Mixed (some Devt)	Commercial, Mixed (Some Devt) or (No Devt)
Funding Source Type	Private Sector, Public-Private Partnership, SOE	Private Sector, Public-Private Partnership, SOE	Private Sector, Public-Private Partnership, SOE	National NGO, Regional NGO	Might be unclear, but **NOT OFFICIAL**
Flow_Type	All Joint Ventures seeking a return on investment (including In-kind or technical assistance related to a joint venture)	All Foreign Direct Investment (including In-kind or technical assistance related to a joint venture)	*ODA-like rules apply for flow type*	*ODA-like rules apply for flow type*	Vague-TBD
Special Rules	Only make "+Gov" if deal involves SOE or official organization	Only make "+Gov" if deal involves SOE or official organization	Only make "+Gov" if deal involves SOE or official organization	Does not include official aid funneled through NGOs	Only when cannot be placed into other commercial category


Organization of Chinese Financial Flows
Official Finance
ODA-Like
Technical Assistance and Scholarships
Loans with large grant element
Grants with development intent
Financial
In-kind
Debt Relief
Military Aid with Development Intent
OOF-Like
Loans with no/small grant element
Grants without development Intent
Financial
In-kind
Various Lines of Credit
Official Investment
Foreign Direct Investment (FDI)
Joint Venture (JV)
Military Aid without Development Intent
Unofficial Finance
Without Any State Involvement
Foreign Direct Investment (FDI) without state involvement
Joint Venture (JV) without state involvement
Corporate Aid (CA) without state involvement
NGO Aid
With Some State Involvement
Foreign Direct Investment (FDI) with state involvement
Joint Venture (JV) with state involvement
Corporate Aid (CA) with state involvement

Appendix E: Pledges and Commitments

To identify commitments as defined by the OECD, all projects with a status of "pipeline" are split into pledges and commitments.[37]

The pipeline category is split according to the following criteria
Pipeline: Pledge (verbal, informal agreement)
Pipeline: Commitment (written, formal binding contract)
See http://china.aiddata.org/statuses

The instructions for coding pledges and commitments are as follows:

Within the context of the transaction described in the record…

1. If project description has the term **pledge*** then label as **Pledge**.

2. If project description has the term **commit*,** as in, to an agreement, then label as **Commitment**.

3. If the two sides sign a document ***other than a Memorandum***, such as a loan, grant, or framework agreement, then label as **Commitment**.

4. Label as **Pledge** if the project description has none of the above evidence of a pledge or commitment, but includes any of these terms**: talk*, discuss*, extended a pledge, expressed interest, expression of interest, establish* a [general] line of credit, "Memorandum of Understanding", "Memorandum of Investment", memo, or MOU.**

5. If any of the following terms are used, check the source documentation for additional details. If no evidence of a signed commitment is found, mark as pledge.
   a. **Agree***
   b. **Pact**
   c. **Accord**
   d. **Mutual Understanding**
   e. **Offer**
   f. **Granted**
   g. **"Made plans to"**
   h. **Extended**

[37] For more detailed information, see FTS definitions
https://docs.unocha.org/sites/dms/CAP/FTS%2002%20Definitions%20Pledge%20Commitment%20Contribution.pdf

i. **Approved**
j. **Gave [x]**
k. **"Donate" (in reference to a loan or grant)**
l. **Earmark**
m. **Secured**

6. If the above terms in 1-5 are NOT mentioned in the description, check the source documentation. If the nature of the agreement is still too unclear to place in the Commitment category, label the project as **Pledge**.

7. If the project description and/or source documentation show conflicting reports for pledge/commitment, label the project as **Commitment** (i.e., any sign of a formal, written commitment is enough to label as such).

8. If the record shows evidence that the donor has disbursed any portion of the proposed transfer, change the project's status to **Implementation**. Terms indicating partial disbursement include:

a. **deliver***
b. **sent**
c. **start***
d. **provided**
e. **received**

Appendix F: Donor Intent

All projects in the database should be classified according to 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 (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."[38] 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.

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

Category Name	Description
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.
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.
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.
Mixed (Some Development)	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.
Mixed (No Development)	Cannot be categorized into the development, commercial, or representational categories, because the project has both representational and commercial intent.
Null	Insufficient information available to code donor intent.

The following are examples of projects falling into each category:

Development Intent

- 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
- Domestic agricultural or industrial production
- Public infrastructure such as highways, dams, and stadiums
- Environmental protection
- Cancellation or rescheduling of debts
- Technical assistance and technical cooperation
- Support for recipient trade systems
- Extension of line of credit (excluding export or other commercial credits)
- Un-earmarked grants or loans to the recipient government
- One-off contribution of office materials, sports supplies, or musical instruments
- Scholarships for recipient students to study in donor country
- Projects promoting development in the recipient country being financed AND implemented by the donor. Includes vague reference to 'various development projects', "economic aid" and 'economic and technical cooperation'
- Medical missions or brigades/doctors if recurring
- Support to a particular political party or government building (e.g. presidential palace)

Commercial Intent

- 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
    - ***Note: If the offer is export credits or extraction of natural resources, it is commercial. Loans being repaid in export credits or natural resources can be development.***

Representational Intent

- 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*)
- Conferences, workshops, or seminars
- Diplomatic gifts clearly from (or to) individuals acting in official capacity

Mixed (Some Development) Intent

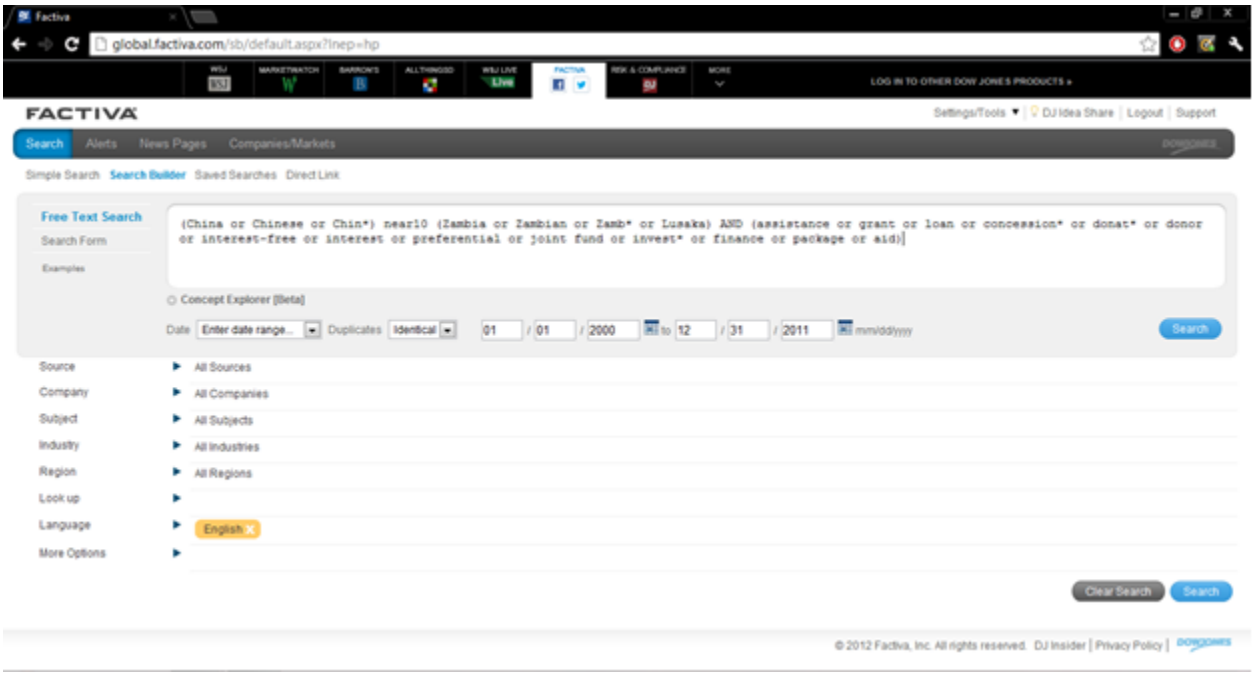
- In-kind contribution in exchange for commercial benefits in the donor country, such as drilling licenses
    - ***Note: Loans being repaid in export credits or natural resources can be development.***
- Housing for employees of a donor commercial operation
- Construction of an industrial park to contain a donor commercial operation, among other enterprises
- Institutes of learning or research which include promotion of donor culture (e.g. Confucius Institutes)

Mixed (No Development) Intent

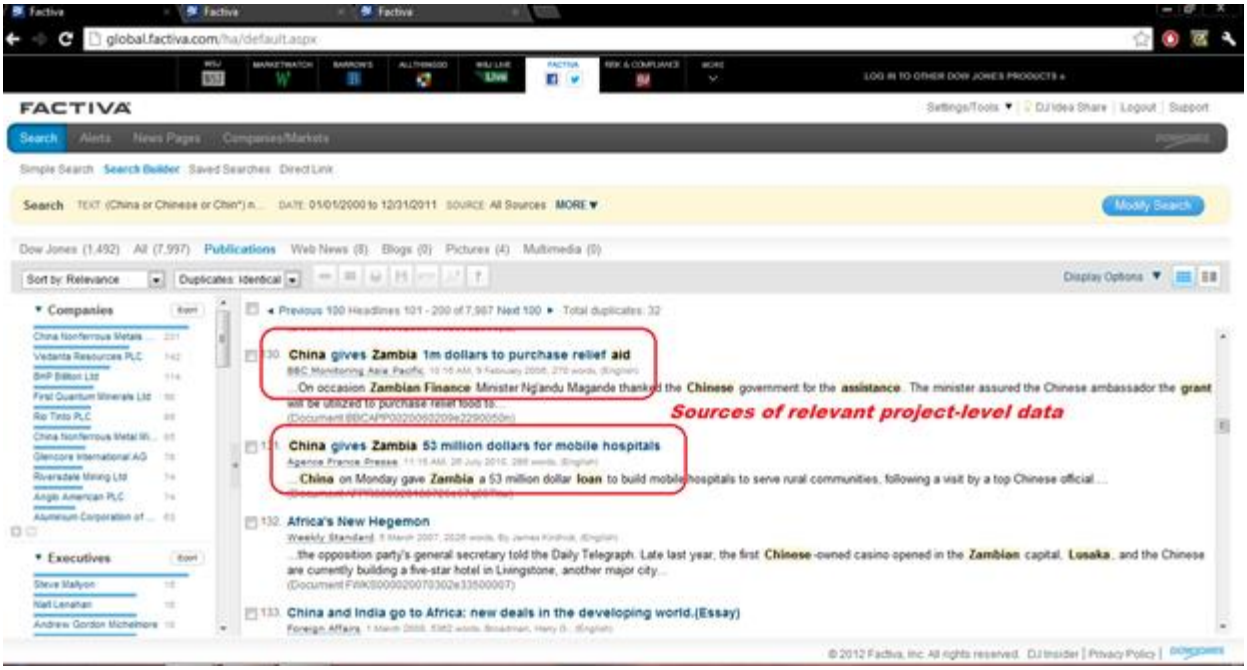
- Conferences primarily focused on business and commerce
- One-off training exercises to strengthen business and commerce ties

Appendix G: Screenshots to demonstrate search and data collection processes

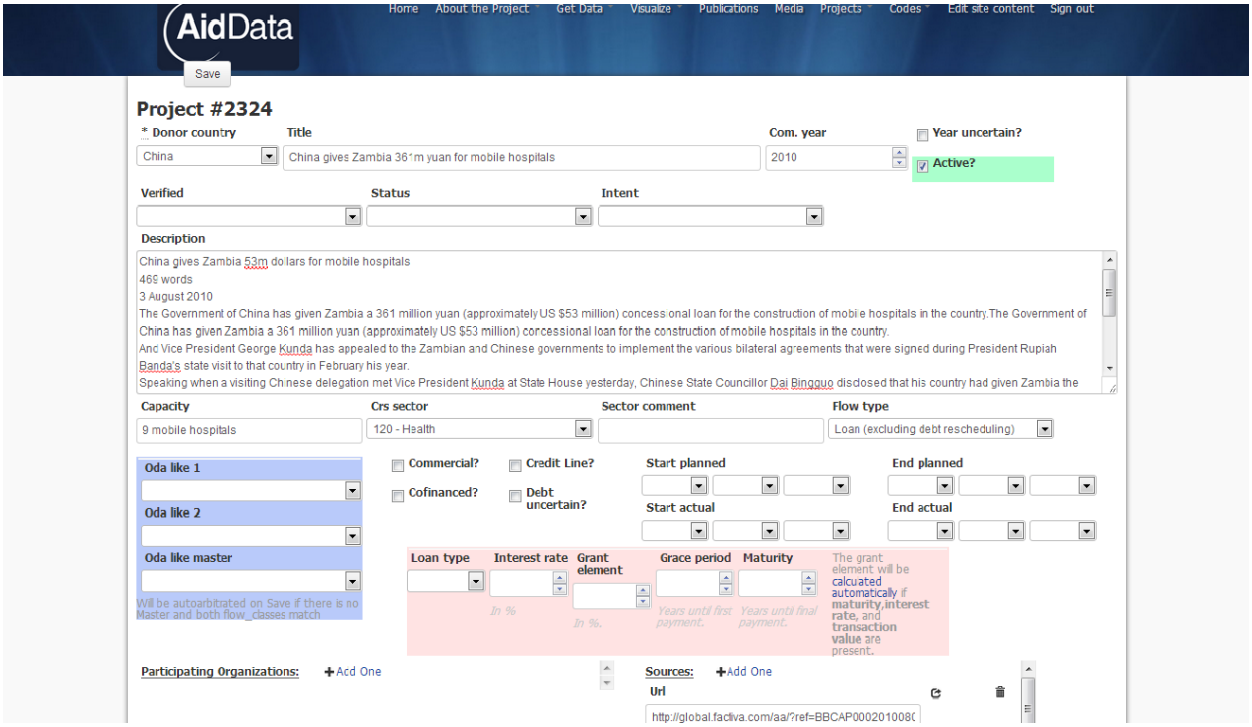
Screenshot #1: Stage One Factiva Search Terms



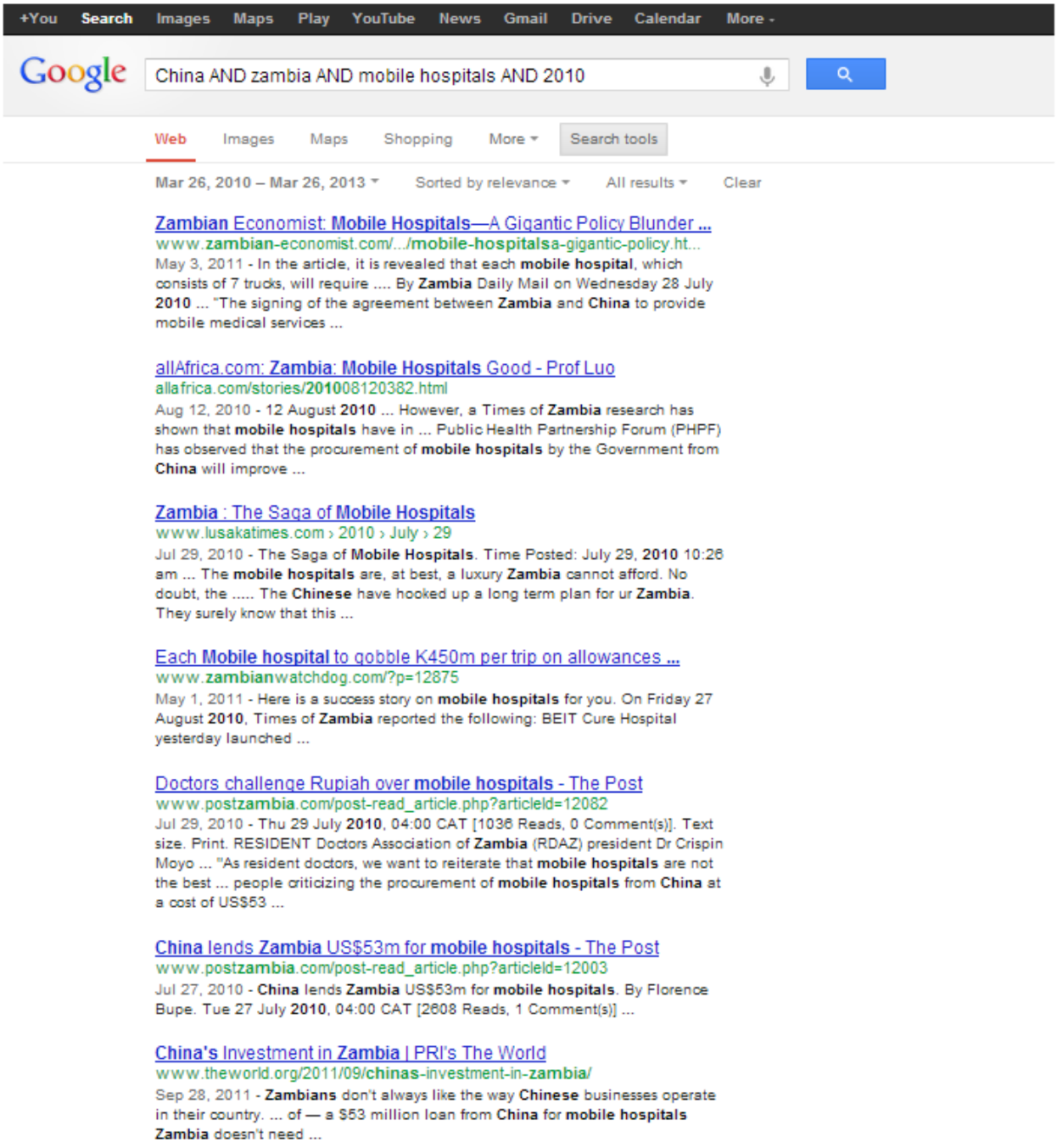
Screenshot #2: Stage One Factiva Search Yields



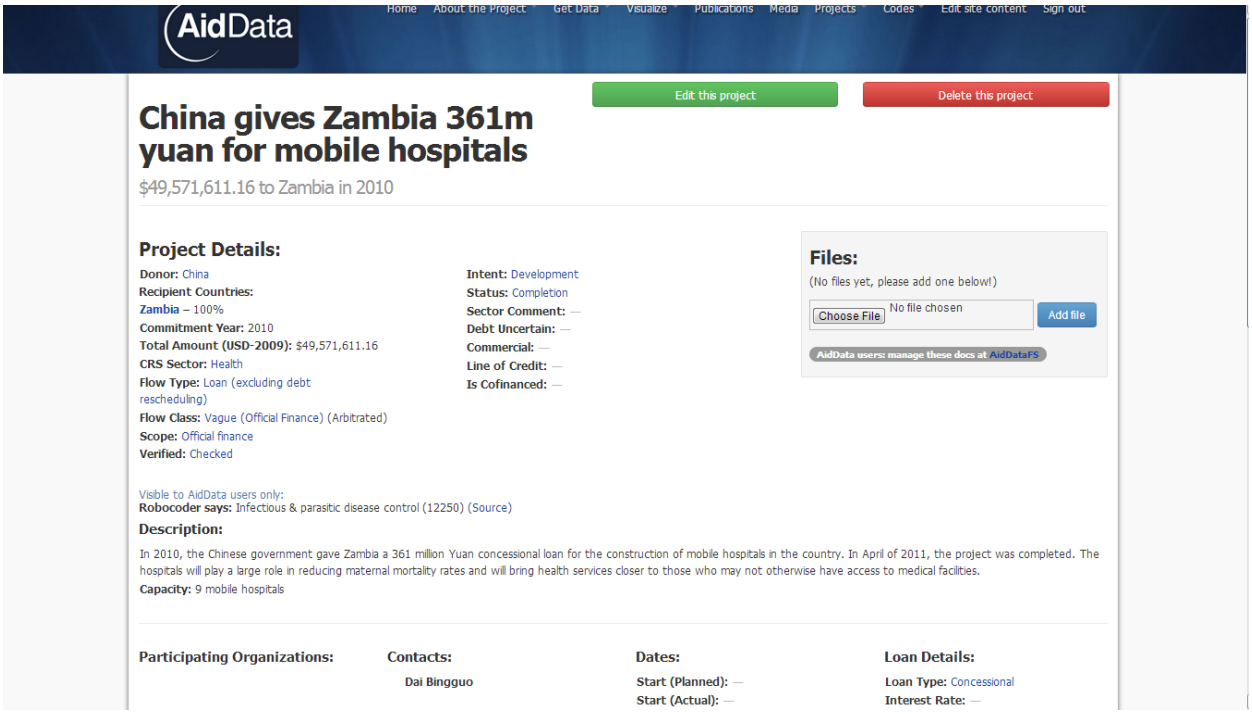
Screenshot #3: Create New Record to Input News Results from Factiva Stage One Searches



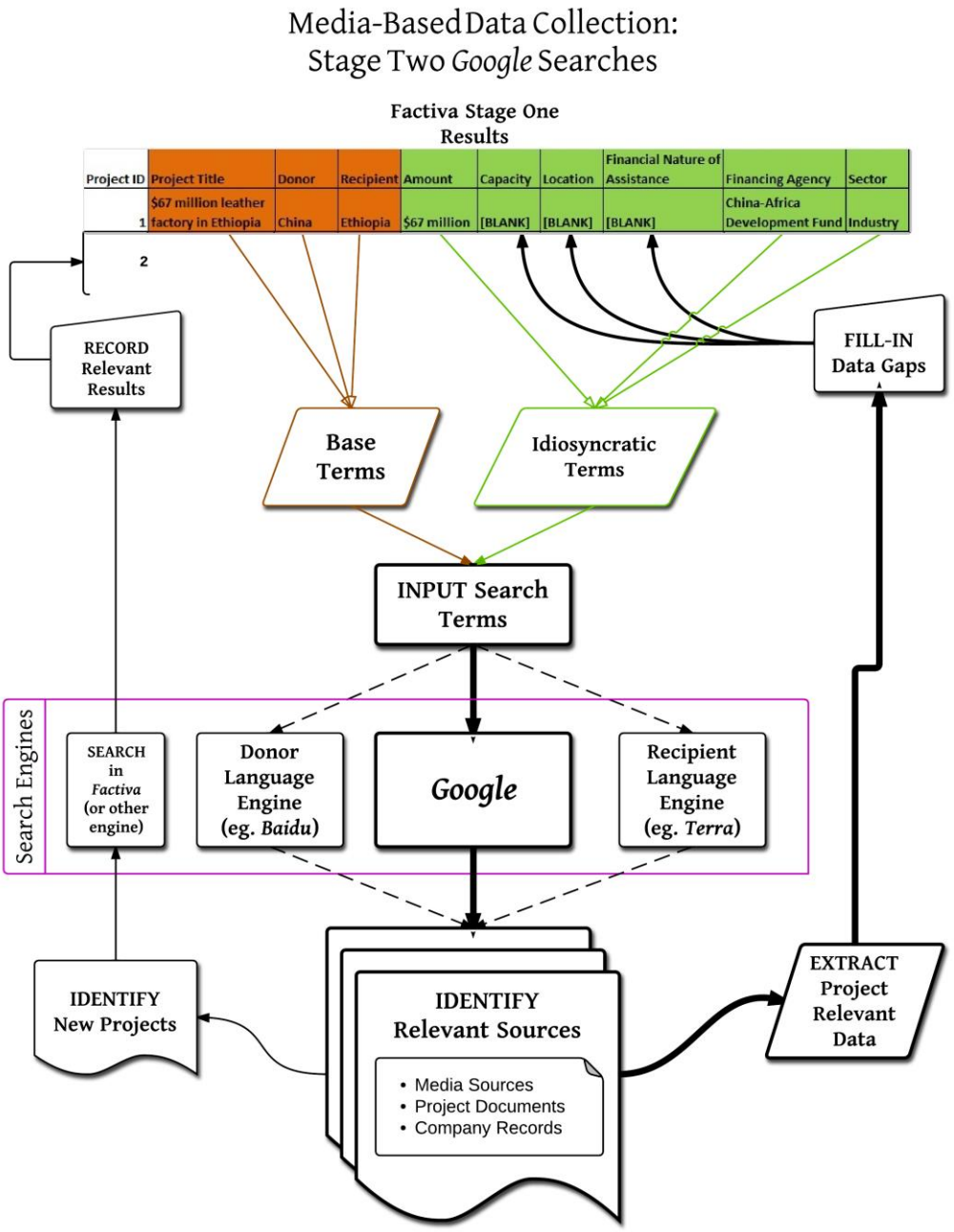
Screenshot #4: Google Search in Stage Two



Screenshot #5: Complete Results for a Project after Stage Two



Appendix H: MBDC Stage Two Process



Appendix I: Database Variable Definitions

Below is a glossary of the variables that AidData's MBDC team tracks while collecting official development finance data. While each project ID can only have one entry for each of the variables under "Project Description," it is often the case that there are multiple Organizations, Recipients, Sources, Amounts, and Contacts for a given project ID.

Project Description:

Variables under this heading provide quantitative and qualitative details about the nature and scope of project IDs in the database

Donor: Entity providing assistance to recipient country. This can be a sovereign state or multilateral organization providing assistance to a recipient country.

ID: Unique identification number assigned to every project that is created in the database. Within the MBDC database, projects can be accessed by ID numbers by using the "Access project by ID" function.

Title: Short phrase describing the nature of a given project.

Year: The year in which an agreement was reportedly made between a supplier of development finance and a recipient for a project. If available, the agreement year is the year of the formal signed commitment for a project; if commitment year is unavailable, or if a commitment has not been made, the year of the informal pledge is the agreement year for a project.

Year Uncertain: Marker for projects without any sources reporting a specific pledge or commitment year. In these cases, the year of the earliest media report serves is coded as "agreement year."

Capacity: Any non-monetary quantitative detail(s) about a given project that help define the scope of the project. This variable is designed primarily for in-kind contributions in which donors provide recipients with a set quantity of goods and/or services. Common examples include medical supplies, food aid, and technical specialists that perform training inside the recipient country. However, the "Capacity" variable can also be used for projects that do specify a monetary value. For example, a rural school that was built using a $1 million donation may have a "capacity" of 25 classrooms. This idiosyncratic data is particularly useful for targeted searches in Stage Two.

Description: 1-2 paragraph comprehensive summary of the activities supported by the project. See page 15 of the methodology for a detailed explanation of how to record this variable.

CRS Sector: 3-digit sector classification based on OECD purpose codes, as below.

Code	Name
110	Education
120	Health
130	Population Policies / Programmes and Reproductive Health
140	Water Supply and Sanitation
150	Government and Civil Society (including peace and security systems)
160	Other Social infrastructure and Services
210	Transport and Storage
220	Communications
230	Energy Generation and Supply
240	Banking and Financial Services
250	Business and Other Services
310	Agriculture, Forestry and Fishing
320	Industry, Mining, Construction
330	Trade and Tourism
410	General Environmental Protection
420	Women
430	Other Multisector
510	General Budget Support
520	Developmental Food Aid/Food Security Assistance
530	Non-food Commodity Assistance
600	Action Relating to Debt
720	Emergency Response
910	Administrative Costs of Donors
920	Support to Non-governmental Organizations (NGOs) and Government Organizations

Sector Comment: Short phrase for coders to document additional details regarding the specific subsector of the project. For projects coded as "Multisector," coders should list each of the known sectors in "Sector Comment."

Status: Tracks the progression of a project. Labels include: Pipeline: Pledge, Pipeline: Commitment, Implementation, Completed, Suspended, Cancelled.

Active: Every project in the database is either active or inactive. Projects are initially coded as "active" and remain such until labeled "inactive" by a coder. Project IDs found to be duplicates of preexisting project IDs, as well as those that are not found to constitute a project, are labeled as "inactive." Inactive projects are later reviewed for potential data grabs and are subsequently deleted from the database.

Donor Intent: Perceived intent of the finance provider. Categories covered by "donor intent" include:

- Development
- Commercial
- Representational
- Mixed (some development)
- Mixed (no development)

Flow Type: Details on how financial flows, goods or services are transferred from the donor to recipient for a project. Flow types include:

- Debt Forgiveness
- Debt Rescheduling
- Export Credits
- Foreign Direct Investment
- Freestanding Technical Assistance
- Monetary Grant
- In-Kind Grant
- Vague Grant (either monetary or in-kind, or both)
- Loan
- Joint Venture with Recipient
- Scholarships/Training in Donor Country
- Strategic/Supplier Credit

Flow Class: Coders are also instructed to assign all projects to one of the following flow class categories:

- ODA-like
- OOF-like
- Vague (Official Finance)
- Official Investment

- Joint Ventures with Chinese state involvement
- Joint Ventures without Chinese state involvement
- FDI with Chinese state involvement
- FDI without Chinese state involvement
- Vague (Residual Commercial Activities)
- Corporate Aid (with Chinese state involvement)
- Corporate Aid (without Chinese state involvement)
- NGO aid
- Military Aid

More details on these flow categorizations can be found in **Appendix D.**

Note: While the above flow categories were designed to accurately represent the complexity of Chinese Development Finance, they could easily be tailored to any donor under investigation.

Line of Credit: Denotes a project or financial arrangement where the donor extended a line of credit to a recipient entity. This credit may or may not be used in its entirety by the recipient.

Tied: Provides details on whether projects are connected (tied) to other agreements for the exchange of goods or services between the donor and recipient. Labels include: N/A and Tied. N/A refers to projects for which coders were unable to find explicit evidence that a project ID was connected to a specific agreement for the exchange of goods and/or services between the donor and recipient.

Start Planned: The announced start date for a project.

Start Actual: The actual date that the implementation of a project began on.

End Planned: The announced completion date for a project.

End Actual: The actual date that a project was completed on.

Verified: All project IDs are initially coded as "raw." After Stage Two searching is completed for a project, coders select either "S2: Checked" or "S2: Suspicious." The former selection is for project IDs that appear relatively straightforward, while the latter categorization is for IDs that contain potential errors. These errors may be due to conflicting data sources, seemingly hyperbolic data sources, and/or utter lack of data to provide coders with enough confidence to mark them as "S2: Checked."

Organizations:

Donor and recipient organizations, both in the public and private sector, which are involved in some way with the project

Name: Name of organization

Type: Nature of organization. Categories include: Academic, Training and Research; Foundation; Government Agency; International NGO; Multilateral; National NGO; Political Party; Private Sector; Public Private Partnership; Regional NGO; and State-Owned Company. Origin and type for each organization is determined through a primary source (e.g. organizational website) or secondary source (e.g. academic article or news article stating the nature of the company).

Role: Responsibilities of an organization with respect to the given project. Labels include: Accountable, Funding, and Implementing

Origin: Where the organization is located. Labels include: Donor, Other, and Recipient.

Is Cofinanced: Denotes projects with multiple countries acting as donor (including the donor of investigation), where the financial amount provided by each donor separately could not distinguished.

Recipients:

Variables under this section provide details on the recipient entities of projects in the database.

Recipient: Entity receiving development assistance from donor entity. This is a sovereign state unless otherwise specified within a research project. A project ID can have multiple recipients in cases where a donor entity is providing financing, goods or services to more than one country under the umbrella of one project.

Percent: What percentage of the flows used for the project are going to a given recipient. This will be 100% for all projects with one recipient.

Detail: This space is for sub-national geographic detail about the recipient(s) associated with a given project ID

Sources:

Variables in this section document information about the media sources used to create project IDs.

URL: HTML link to a data source used in the creation of a project ID.

Doc. Type: Nature of the document used to create and/or add information to a project ID. Labels include: Donor Media Report, International Media Report, Recipient Media Report, and Non-Media Source.

Source Type: Origin of the source documented. Labels include: Baidu, Factiva, Google HK, Google SA, Google US, Local Language Source, Other English Source

Date: Date that data source was published on the Internet.

Amounts:

Variables in this section describe the monetary and financial details associated with projects in the dataset.

Amount: Monetary amount pledged or committed by the donor entity for the completion of a project.

Currency: Currency associated with the monetary amount for a project.

USD-2009: Deflated monetary equivalent of reported monetary amount in reported currency to 2009 U.S. Dollars.

Debt Uncertain: Debt reduction agreement in which the amount of debt reduced in the final deal cannot be ascertained.

Contacts:

Variables in this section provide information on individuals associated with projects.

Name: Name of contact

Org: Organizational affiliation of contact

Info: Other miscellaneous details about the listed contact

Loan Details:

Variables in this section provide key information on the terms of loans within loan projects

Loan Type: represent the general nature of the loan. Categories covered include: interest-free; concessional; non-concessional; no information.

Interest Rate: Reported interest rate of the project loan, in percent

Maturity: Reported duration of the project loan, in years

Grace Period: Reported grace period of the project loan, in years

Grant Element: Grant element of the loan, in percent. Uses grant element reported by project sources, unless it can be calculated independently with the OECD grant element calculator (assumes 10 percent rate of discount, 2 payments per annum, and equity principal payment).