AidData Coder Instructions: Tracking Underreported Financial Flows (TUFF)

Version 1.3

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

October 2017

Special Thanks

The authors are indebted to all the researchers at the College of William and Mary and the National University of Singapore who have contributed to the TUFF methodology and this coder instruction document.

College of William and Mary

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

National University of Singapore

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

Others

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

Citation

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

Contents

References

Stage One: Project Identification

*The objectives of Stage One are to identify potential projects within a defined research scope and standardize the initial information into discrete project entries. This entails performing general searches across several different information repositories. When creating these new project records prioritize the "Title" "Sector" and "Intent" variables, as these fields are important for subsequent rounds of searches, duplicates checks, and flow class coding. For the "Chinese Official Finance" project, use project templates contained in **Appendix J**.*

There are four distinct information repositories our researchers have searched during Stage One. Depending on the donor of interest, these information sources may vary.

Searching Factiva

The first step in Stage One is to access Factiva.[1] 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)."[2] 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.

After the donor and recipient terms are inserted, enter a set of "official finance keywords" into the search box as well as a specified date range. These keywords are donor-specific and created by AidData staff after performing preliminary research on a donor's official finance activities. Before implementing a search phrase, a program manager will test it to ensure that it delivers relevant Factiva output.. 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").[3]

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 are examples of base search terms that the team has for different projects

(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 finance or package or aid)*[4]

(Venezuela or venez or Caracas) near5 (Cuba or cuba* or La Habana) AND (cooperación or cooperación técnica or coope* or colaboración or asistencia or asesoría or asesoramiento or asesor* or proyecto or programa or fondo or intercambio de expertos or intercambio de funcionarios or intercambio de experiencias or intercambio de conocimientos or pasantía* or*

---

[1] A tutorial on the Factiva database is available *online*.

[2] 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.

[3] When performing Factiva searches, be sure that your search filters are standardized. In the "sources" field you should select "all sources." In the "language" field you should select "English." Sort search results chronologically to better organize workflow when scraping Factiva output.

[4] 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.

seminario or capacitación or proyecto* or programa* or socio* or ayuda mutua or becas or transparencia de conocimientos or financiación or financiamiento)*

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

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

(مؤسسة الوليد طلال or الهلال الأحمر السعودي or الصندوق السعودي للتنمية or مساعدات نفطية or قرض or مساعدات إنسانية or مساعدات خارجية or معونة) AND (اليمن) near5 (السعودية)

See Appendix F (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. 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 net as widely as 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 sources 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 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.[5]

See Appendix F (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.[6]

3. 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

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 F, Screenshot #3 for an example of Stage One project data input.
4. Create a proper title when creating project entries, use the following titling formula (this is only a title guide, if information is missing, adapt the formula):

[Donor Country or Organization] [Status] [Activity/Flow type] [Dollar Amount] [Recipient/Location]

Example: China EXIM Bank Pledges Infrastructure Loan of $100 Million USD to Kampala Housing Project

[5] To improve workflow for donor-recipient pairings with high search results, research assistants can break down a year into months or quarters and perform searches individually for those sections.

[6] Depending on customized personal settings, Factiva automatically logs users out after five minutes of inactivity.

It is also essential to code the status of newly identified projects correctly (see Appendix D for complete status coding guide). A project's status has implications for aggregate financial totals AidData reports and can influence Stage two searches.

5. Create a resource entry for each source used to identify a project, filling each field (title, source URL, resource type, author, publisher, publish date, and publisher location) when the information is available. When saving a resource from Factiva, or any other site behind a pay wall, click the "Don't fetch this resource" marker. The "fetch" feature will save the HTML of the resource URL to preserve its information, in case of site crash or removal of the resource. Resources are stored independent from project entries, allowing multiple projects to be linked to a single resource.

6. Write a description based on the following template:

- **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?[7] **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.)[8]
- **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? If so, use the labeling convention (possible linked to project ID#....) in the title and description. For easily confused projects, often ones which are linked to each other, also use the description paragraph to explain why the project is distinct from similar records. For instance, linked projects could be different financing modalities for various phases of the same construction project, or "child" records of a more general "parent" project. In addition, if a project is likely to be a recurring project or is a part of a series of projects, write this in the description. Examples may include medical teams, peacekeeping taskforces, annual scholarships, technical training, youth volunteers, and friendship construction projects such as bridges and hospitals.

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 (see project ID #2034). Paragraphs should be written once the project variables are established using all available data sources.[9]

STAFF_NOTEs: When writing a project description, it is also vital to capture *any* information gaps associated with project records in a staff note. These meta-data are vital to present outside users with an accurate picture of the available information on a project. For example, if there is conflicting information about a project's financial value make this explicit in a staff note. These notes have an ancillary benefit of conveying information gaps to future researchers to address.

When writing a staff note, use the following labelling convention at the bottom of a project description: STAFF_NOTE:

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.

---

[7] Financial amounts are to be denominated in the currency used by the most "reliable" resource (see Appendix B of methodology document for our ranking of resource reliability).

Appropriate labeling convention for financial amounts is as follows: Numeric value, Units (thousand/million/billion), Currency.

[8] At the start of the Chinese Official Finance to Africa data collection initiative, research assistants 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.

[9] 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.

Sentence 4: May be linked to project ID #810.

Sentence 5: STAFF_NOTE: Current project status unknown.

Searching Donor Government Information Systems

This purpose of this step is to gather project-level official finance data made available by donor. This may seem counterintuitive since the TUFF methodology targets suppliers of official finance who generally do not publish much information about their overseas activities; however, many donors do still provide some information that can enhance the quality of existing project records and help identify other potential projects. While this information is rarely comprehensive, it is still useful to collect.

Solely using media reports to track official finance creates the risk of misinterpreting information, particularly for non-DAC donors who often administer official finance through modalities poorly understood by Western observers. For example, media outlets often report on "megadeals" emerging donors offer to 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).

To overcome the problems associated with overreliance on media reports, the TUFF methodology also employs targeted searches of donor and recipient government websites for project information. While such ".gov" sources, which include embassy cables, press releases, annual reports, and ministry websites, usually do not contain *comprehensive* project-level information, they enhance data quality by providing technical details lacking in media reports.

To effectively collect relevant information from official sources, RAs must first become familiar with the development agencies active in a recipient country and the recipient government ministries that manage development finance inflows. To identify the ministries or agencies of interest, examine the administrative structures of the donor and recipient of interest. For the "Chinese Official Finance to Africa" data collection exercise, we identified several rich sources of project-level information published by both donor and recipient governments. China publishes detailed information about development activities on country-specific embassy (operated by MFA) and Economic and Commercial Counselor websites (operated by MOFCOM). For another TUFF data collection project tracking Saudi aid, we found that the Government of Saudi Arabia issued press releases about external financing on its US embassy webpage. When scraping government information systems, it is important to perform extensive desk research in advance of data collection to identify all possible information repositories.

Searching Government Websites

To web-scrape an embassy website, review each page for development project announcements or updates in both the donor language and English. The pages for news/press releases, "Economic and Trade Relations," and "Other Exchanges and Cooperation" are the most valuable pages for Research Assistant (RA)'s to review. For Economic and Counselor websites, the pages "Bilateral Cooperation" "Commercial News" and "Bilateral Visits" contain the most valuable project information.

To web-scrape these pages, repeat steps 1-6 from Factiva searches.

Important Note: Do not assume that the English language versions of the website will contain all the information as the native language version. This means that RAs must perform web-scraping in both languages.

Searching Recipient Information Systems

Many developing countries track and publish information about incoming aid flows through aid information management systems (AIMS). Many of these systems are publicly available and offer detailed information about individual projects. Given the diversity of reporting systems, there are several strategies to extract project information (see Appendix K for screenshots of the different processes).

Export full AIMS dataset as a spreadsheet to data mine

AIMS will typically allow users to export project records as a csv. file. Once a project manager or senior RA has generated this export, RAs can input project information into new records or update existing entries. RAs should adapt steps 2-6 from Factiva data entry to create these project records.

Generate records directly from aid information management system

A few AIMS do not enable public users to export tailored spreadsheets. For these systems, RAs must pull information directly from AIMS project pages. To do this, RAs must first filter projects for the donor of interest. Next, RAs must access the individual project entries for each aid activity. This "long-form" entry contains the same type of information as a database export, but is not formatted as a spreadsheet. RAs should adapt steps 2-6 from Factiva data entry to create these project records.

In addition to formal AIMS, some recipient ministry websites can also provide valuable project information. Annual reports from the Ministry of Finance or Treasury can contain valuable information on outstanding debt to donors or loan conditions for specific projects. Ministry of Transport or Energy websites will often contain press releases or documentation related to large infrastructure projects. While the availability and detail of these reports varies across recipients, this step is vital for uncovering more accurate and specific details on major projects.

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: "debt loan outstanding site: *http://www.minfin.gv.ao/*"

Keywords for targeted ministry searching can include, but are not limited to: Donor, Export-Import, debt, loan, credit, dam, road, highway, MoU. It is important to note that not all recipient ministry websites will be equipped for this type of web-scraping. If a government publishes PDFs which are not machine readable a research will have to download and search government documents manually.

Searching Google Scholar

Researchers can also search Google Scholar to extract project-level information from academic articles and high-end publications. Follow the general order of search keywords (see Appendix H for instructions on how to search for articles in Google Scholar):

[donor country] OR [donor demonym] OR [donor countr*] AND [recipient country] OR [recipient demonym] OR [recipient countr*] OR [capital of country] 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

Search IMF Article IV Staff Reports

Under Article IV, IMF conducts country consultations where economists visit the member countries to gather information and hold meeting with government and central bank officials as well as private investors and labor representatives, members of parliament and civil society organizations annually. After each consultation, the mission summarizes its findings in country reports, which is public on the IMF website. Sometimes these reports capture information about the loans that member countries currently take from other donors such as China, Saudi Arabia, etc. Researchers are supposed to scrape through these reports for additional project information. Once a project is identified, researchers are supposed to go through the same steps 1-6 as they treat projects identified through Factiva searches.

Stage Two: Source Triangulation

The objectives of Stage Two are to triangulate project information gathered in Stage One before final quality assurance checks in Stage Three. During Stage Two, RAs should prioritize writing a clear and concise project description, confirming "Transaction Amount" or finding one if it is blank, and populating any missing fields. RAs should remember that all fields coded during Stage One are subject to change if new information becomes available during subsequent rounds of searching.

After "casting the data collection net" as widely as possible in Stage One and entering search results into standardized project pages, Stage Two involves searching for additional information about each project identified during the initial rounds of data collection, and refining each project record based on the most reliable information that is accessible. This is achieved by drawing upon various information repositories and search engines including Google, the primary search engine used in the donor country, academic articles identified through Google Scholar, and recipient ministry websites.[10] Detailed searching considerably narrows search yields compared to Factiva, helping to pinpoint resources 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.

While performing Stage Two, it is critically important for RAs to bring a critical eye to *all* information about a project captured during the first round of searches, especially when that information is initially sourced from media reports. Oftentimes the media will conflate the face value of a project with a donor's financial contribution to that project. For large projects in particular ($1 billion+) this assumption is problematic because it creates the risk of unfairly inflating a donor's financial commitment to a recipient. For projects simply announced in media reports, RAs should prioritize finding updated information about that activity, not simply finding additional sources covering the initial announcement.

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 F, 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.[11] Update individual project records in the database as new information is found during Stage Two. See Appendix F (Screenshot #5) for an example Stage Two project record.

Follow the below guide for detailed step-by-step instructions for targeted Google Searches.

---

[10] 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 Chinese official finance 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.

[11] 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.

Google 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 any media sources. 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. 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.

3. 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 C), donor intent (Appendix E), year, transaction amount, CRS sector (Appendix G), and links to all sources which contained valuable information.

Additional Complementary Steps

Steps 4-6 below were discovered to help streamline and focus Stage Two searches, but they were not applied systematically from the start of the Chinese Official Finance to Africa pilot.

4. Connect search terms inputted into Google

The structural design of Stage Two Google searching should mirror 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.[12] 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).[13]

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

5. 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:[14]
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*

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*

6. 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."

Donor Search Engines

1. Like targeted Google Searches, RAs can use donor search engines uncover additional project information. Use the donor's native search engine (for example Baidu) when available. If the donor does not have a native search engine, use the donor's in-country Google server and perform searches in the donor's native language. Coders should follow this general order of search keywords when performing Stage Two donor search engine searches:

---

[12] 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".

[13] See http://googlesystem.blogspot.com/2010/12/googles-around-operator.html for details.

[14] 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."

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**

Edit Titles and Project Descriptions in Stage Two

After filling out additional information in the project entry, researchers are supposed to edit the titles and/or project description with additional information found in Stage Two. While keeping the structure of titles and project description (see Stage One title and description template for details) intact, researchers need to resolve the concerns mentioned in staff notes and correct the content if additional information is found. Researchers may leave further notes in the staff note if questions are not fully addressed.

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).

Stage Three: Quality Control

Data quality assurance exercises are essential to identify and correct errors made during data collection. This section provides a discussion of the methods employed by AidData to help correct deficiencies in project records. This stage is only conducted by program manager or senior RA appointed by Program Manager.
AidData staff perform rigorous data quality assurance exercises following the conclusion of Stage Two searching. These include megadeal checks, flow class arbitration, least verified project checks, illogical combination checks and duplicate checks. 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.

"Megadeal" Checks

A program manager or senior RA reviews all official finance projects with reported financial commitments greater than or equal to $100 million USD for accuracy and missing information. A program manager also reviews all records greater than $5

billion USD for accuracy and missing information.[15] A senior RA or program manager will also review all project records initially marked as "suspicious" during the Stage Two searches and update records 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.
- Resources 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.[16]

As an additional quality check, coders perform a final round of verification exercises for the largest 10% of projects in a particular dataset—ranked by committed financial value—to ensure these projects were coded accurately for all variables. A team of two coders should independently review 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, coders are required to collectively agree that all variables were accurately recorded before moving to the next project.

Record-by Record Review and Flow Class Arbitration

Program manager or Senior RA need to review and vet each project before it is published and arbitrate flow class when there is disagreement among RAs who have coded the projects. The flow class variable is essential to correctly calculate total official financing from a donor (see Appendix C for full flow class coding methodology). As such, it receives two rounds of coding from different researchers. When RAs disagree on a flow class, the program manager or senior RA will review *all* project resources and make a final determination on flow class.

Least Verified Checks

Before publishing a final dataset, researches make a final attempt to find verification for single-sourced projects. Because single source projects are generally small projects unlikely to be featured in media or government sources, there are strong diminishing returns to additional searches for these projects. To optimize resource allocation during this step of the process, the program manager will rank single-sourced projects by financial amount, prioritizing projects with larger financial amounts. RAs will also follow a strict time limit or ten minutes per project when looking for additional sources.

Logical Inconsistency Checks

Flow class coding is contingent on how the flow type, intent, and grant element variables are defined; Program Manager must check a dataset for values assigned to different variables for an given project that are logically inconsistent (See Appendix C for flow class definitions and illogical combinations). For example, export credits by definition cannot be

[15] 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.

[16] 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 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.

counted as an ODA-like flow regardless of its level of concessionality. To uncover logically inconsistent values assigned to different variables for an individual project, first conduct a targeted query of the dataset for the following logical inconsistences (See Appendix J for full details).
Once the program manager has identified coding errors, a RA will review each project record to identify where the illogical code assignment occurred. RAs should review each resource and look for the underlying source of the mistake.

Duplicate Checks

In light of the fact that the TUFF methodology draws information from a range or sources and tracks development projects over time, there is a risk of double counting projects (whether from capturing announcements of commitments and implementation/completion or capturing sub-components of a larger deal). Duplicate records may exist in the database prior to publication for several reasons. First, at first glance coders may not recognize two identical projects that use different financial currencies. Another possibility is that certain project records are overly vague which makes duplicate identification very difficult. There may be more than one name for a single project. To reduce the likelihood of double-counting, the TUFF methodology includes a round of duplicate checks before the release of a new dataset.
Searching for Duplicates

To check for duplicates, first filter the dataset by recipient and sector. For all recipients the following sectors should be examined: emergency response, communication, education, energy generation and supply, government and civil society, health, other multisector, transport and storage, industry, mining, and construction. Projects with the sector label "Unallocated/unspecified" should also be examined in case they can be reallocated to toher sectors. In addition, RAs should check whether there are a high number of projects for specific recipients in other sectors, and proceed to check those sectors as well. Examine the Title, Description, and Transaction amount of each project entry to identify potential duplicate records. For any potential duplicate projects, examine each resource to determine whether the projects are really capturing the same flow. At this stage, the commitment date, start date, implementing agency, and contacts are particularly helpful to reveal duplicate records.

Deactivating Duplicates

Once a duplicate record has been identified, the researcher should select one of the two projects to serve as the active record (alpha) and deactivate the other project (beta). Before deactivating the project, transfer any relevant project information to the alpha record along with any resources from the beta record.

Additional Notes on Duplicate Records

Different emergency aid projects with same financial value are not necessarily duplicates, as one could be an in-kind grant and the other could be a cash grant. Projects that are recurring, such as scholarship programs or health teams, are not duplicates. Make sure the commitment year and the description is different for each individual project. Sometimes projects with different commitment years can still be duplicates of each other because the "commitment year uncertain" marker could be checked. Different financial commitments for a single physical project are not duplicate records, and should be separate but lined projects. Be sure to change the title of the deactivated project by writing "duplicate of #project number" so that in the future RAs can ascertain why the project has been deactivated.

Deflation & Financial Review

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

Extended Review

Once the dataset has gone through previous steps of vetting by staff who are directly involved with data collection, the dataset is then reviewed by additional AidData staff and a cohort of external researchers. These reviewers vet the data using various methods, including but not limited to (1) running descriptive statistics with the dataset to identify anomalies or suspicious results; (2) comparing the dataset and the resulting financial amounts to other published estimates of Chinese Official Finance (or subsets thereof) to identify significant deviations from other estimates, including White Papers published by the Chinese Government and estimates published by third parties; (3) comparing individual project records to official sources to ensure comprehensive and accurate coverage; (4) reviewing individual project records for errors or missing data; and (5) identifying biases in the data and identify potential ways to address them.

Data Entry FAQs

How should I handle "omnibus" or umbrella agreements that comprise several smaller subprojects?
When dealing with a large loan or grant that will cover multiple projects, you may check the marker for "umbrella" so that users would be able to exclude them from analysis to avoid double counting. A typical example of umbrella projects would be economic and technical agreements that China sign with recipients countries and would be distributed to projects that both countries agree on at a later stage. Credit lines are also considered as umbrella projeccts. If subcomponent projects are identified, please link them to the umbrella projects. When all the subcomponent projects are identified and the total transaction amounts add up to the amount of umbrella project, the umbrella project may be deactivated.

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.[17] If the project has multiple donors and the individual donor commitments cannot be isolated, mark "is cofinanced." Only enter the financial commitment of the donor of interest into the "Transaction Amount" field. Other commitments can be captured in the description. Add all financing agencies in the "Organizations" field.

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?

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

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

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 D 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, as this can 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.”

What do I do if I encounter a project that only has sources in a language that I cannot understand, e.g. Chinese, Spanish, Arabic, or Portuguese?

Copy the project ID number and send it to your senior RA or program manager. They will reassign the project to someone with the requisite language skills.

How much time should I spend per project on Stage 2?

About 20 minutes on average. 15 minutes for shorter projects and 30 minutes max for anything. Generally you will spend 30 minutes on projects with larger financial flows. **However, we want high quality project entries**, so don’t worry if you spend a bit more time.

If China is providing a grant for a president for the rehabilitation a presidential lodge (palace, residence), what should be the intent and CRS sector? (See #34598)

The methodology prescribes that such a project is development intent with Government and Civil Society as the CRS sector.

If refugee aid is being granted one country to assist refugees from another country, who should be the recipient?(See #34489)

The recipient should be listed as the host country initially receiving the aid inflows, not the country whose citizens hav sought refuge in the host country

Appendix A: Best Practices for Maximizing Workflow Efficiency

The following activities were undertaken in order to make efficient use of scarce resources during the TUFF pilot. We consider these activities to be "best practices" for implementation of the 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 One, 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 "Chinese Official Finance to Africa" data collection initiative, 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 the most two recent years in the time series.[18] Also, do not search for 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 the most recent two years in the time series.
- **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.

The above "Wave" strategy has been adapted and modified in subsequent data collection efforts. For instance, while generating data on Chinese financing commitments in South and Southeast Asia, RAs also utilized three waves to order projects by their perceived importance. The first wave included all potential "megadeals" with purported financial transaction values over $100 million (2009 USD). The second wave focused on all non-megadeals with a "Pipeline" (Vague, Pledge, or Commitment) status. The third wave included all remaining 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.

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

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.

Appendix B: Supplementary Notes on Using Donor Search Engines

4) ***Search Engines***
While conducting Chinese-language searches, RAs have found that Chinese search engine Baidu is less useful then Google. The former has many advertisement and irrelevant search results. Nonetheless, when using Baidu, AidData coders found that the "News" or "新闻" tab is rarely helpful for data enhancement purposes. Searches conducted through the "Web" or "网页〞tab have been more fruitful.
Similarly, for Malay-language ministry searching, it is more effective to search through Google instead of directly searching in the ministry sites. (For example, to search "China loans Indonesia $10 million" in the MFA website, search instead for "China pinjam $10 juta site: kemlu.go.id" in Google)
5) ***Search Keywords***
By default, start with only three to four terms in the search. As an example, if searching for information on project #32459, try searching "中国 柬埔寨 6亿 温家宝." While Google and other search engines have more powerful searching functions, in this case a simple query is sufficient. Other useful strategies include:
-Use the most identifiable term, e.g. 100MW power plant, then add 中国 and [recipient country name in Chinese] to the search box
-Look for the Chinese company involved in the project, use Google to find the Chinese name of the company (or if they have a website just go from English to Chinese)
-Search with abbreviated country, company, organization names, e.g. China Development Bank – 国家开发银行 – 国开行，China Metallurgical Group Corporation – 中国冶金科工集团 – 中冶，Cambodia – 柬埔寨 – 柬
Common terms to include: 救灾，重建，电站，进出口银行，大坝，物资，培训，纺织厂 (textile factory), 经济特区 (SEZ), 无偿援助、无息贷款，援外优惠贷款，优惠出口买方信贷/优买信贷，水电/火电/煤电/核电，可行性研究, 水利，灌溉，现汇，奖学金，青年志愿者，谅解备忘录(MoU)， 优惠贷款协议框架
6) ***Be careful when inputting donor-language search terms for several variables***
If the resource section already has a Chinese source, find the project name, 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. When information conflicts between multiple resources, prioritize translations given by Chinese official sources since the English-language resources may have been translated from the original source. Moreover, note the following:
a. **Village and city names in recipient country**: Some sources write these differently. Typically country names and capital cities are uniform. Using the wrong name for a village may preclude valuable results with a different name for the village. Use Wikipedia/Google Maps to get the Chinese name of the location
b. **Name of project**: The project name can vary across sources as well. Using one name can produce a similar negative result as above. Names of infrastructure projects may differ a lot. For instance, the Prek Kdam Bridge in Cambodia is referred to as "北克丹桥" (direct translation), "洞里萨河大桥" (the name of the river the bridge is on), and "第二座柬中友谊桥" (second Cambodia-China Friendship Bridge) across different sources. A bridge like this could also be named after the province/city it is in or the cities it is linking, so it is essential to know whether the different names point to the same project to prevent duplicates.
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 donor and/or recipient organization. Native language searches 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.
f. **Transaction amounts:** These are often not uniform across English and non-English sources, so it would be better to not include the transaction amount, but to include the sub-national locations and the names of the people involved (e.g. in signing of MoUs), whenever possible. This insight was discovered specifically during Malay-language sreahcing.
7) ***Finding exact 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.

8) ***Exclude names of well-known figures***

If a well-known figure is connected to a project, 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: Flow Type and Flow Class

All projects in the database should be classified according to their level of concessionality and type of financing provided. This methodology 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[19], and whether it is concessional or not. Projects are separated into flow_type and flow_class categories, as follows.

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 1: Project ID# 549: The Chinese government provided an official grant to Egypt to establish a school.
Example 2: Project ID# 32567: Saudi Arabia Sends Medical Personnel and Equipment to Setup 3 Field Hospitals in Egypt.

OOF-Like: The OECD defines other official flows (OOF) as **"**official sector transactions which do not meet the ODA criteria, e.g.: i.) Grants to developing countries for representational or essentially commercial purposes; ii.) Official bilateral transactions intended to promote development but having a grant element of less than 25 per cent; iii.) Official bilateral transactions, whatever their grant element, that are primarily export-facilitating in purpose. This category includes by definition export credits extended directly to an aid recipient by an official agency or institution ("official direct export credits"); iv.) The net acquisition by governments and central monetary institutions of securities issued by multilateral development banks at market terms; v.) Subsidies (grants) to the private sector to soften its credits to developing countries; vi.) Funds in support of private investment." The following activities and forms of official financing should be categorized as "OOF-like": grants with a representational or commercial purpose (i.e. grants that do not have a primary objective of promoting economic development or welfare in the recipient country), loans from a government institution that do not have any apparent grant element (commercial loans based on LIBOR or LIBOR plus a margin) or a grant element lower than 25%, and export credits from a government institution to a recipient institution (Brautigam 2011a: 206). OOF activities also include "short-term credits to exporters (export sellers' credits) to help them finance foreign sales, and ... longer-term credits to foreign buyers to assist in the export of goods and services" (Brautigam 2011a: 206). OOF also includes lines of credit that a government provides to a donor enterprise (state-owned or not-state-owned) to do business overseas.[20] 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.[21] For all lines of credit, remove transaction amounts until there is evidence that the line has actually been used. Also, do not check the line of credit marker for subprojects financed by lines of credit.

Example 1: 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 2: 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

---

[19] 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 a government or positions in the government and private sector. These personal linkages and overlaps can obscure the true financial source of a project. For example, in one corporate aid project from a state-owned enterprise (*http://china.aiddata.org/projects/489*), the vice president of the company--who is also a secretary in the Chinese Communist Party--and the Chinese Minister of Commerce both attend the opening ceremony. That said, such conflicts of interest are certainly not a data issue exclusive to media-based data collection, and official records can encounter the same kind of challenge.

[20] 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).

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

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, 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.

Example 3: Project ID# 31429: On September 5, 2012, Saudi Finance Minister Ibrahim Abdel-Aziz al-Assaf signed an agreement that would grant Yemen a total of US $3.25 billion in foreign aid in an effort to help the country stabilize after its political crisis in 2011. The package comprised of three parts, *one of which was a US $500 million export subsidy to Yemen*; the other parts included a US $1 billion deposit in the Central Bank of Yemen as well as a US $1.75 billion towards development projects in Yemen. This project is linked to #31417 (CBY deposit) and #31525 (development projects).

Grant Element: The OECD characterizes a loan as ODA if the loan's grant element is 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 for all loans we track. 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 RAs attempting to classify non-DAC financial flows as either ODA or OOF. AidData provides figures using the OECD grant element calculator exclusively in TUFF dataset loans. Due to information gaps in the loan conditions of non-transparent donors, we assume the following for the following loan conditions.

Discount rate: 10%

Repayments per Annum: 2

Type of Repayment: Equity Principle

Automatic Flow Class Overrides: To minimize instances of human error in coding, we have added an automated flow class override the will change the "Flow Class Master" variable to ODA-like when the calculated grant element is above 25% **AND** the intent field is "development."

LIBOR: Many development banks rely on LIBOR—the London Interbank Offered Rate—to set base lending rate for concessional loans. In order to code flow class systematically, it is vital to standardize treatment of projects which use LIBOR for their base lending rate.

Project documentation for loans which use LIBOR as their base interest rate typically report interest rate in the following way: LIBOR + 2% interest. When calculating the grant element for these projects, find the average 6-month LIBOR rate for the year the agreement was signed and add it to the interest rate.

Example: Project ID# 151 details a loan China Exim Bank made to Ghana for a rural electrification project. To calculate grant element, we extracted the relevant variables from available project documentation and found the average 6-month LIBOR rate from the following URL: *http://www.global-rates.com/interest-rates/libor/american-dollar/2006.aspx*.

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."

Corporate Aid: It is common to see especially when collecting data on China that state-owned companies also provide grants or scholarships. While state-owned companies are semi-governmental agencies, we have a separate flow class for these grants as "Corporate Aid + Government" shortened as "CA+Gov".

Example: Project ID #34347: "In 2001, China Radio Television International and Radio Nepal signed a memorandum of understanding on bilateral cooperation. As part of the MoU, China will provide radio training for Nepali technicians and promote assistance for the construction of a relay station to expand Radio Nepal's programmes. China Radio Television International will also provide Radio Nepal with a model FM Transmitter and a 10 watt antenna."

Appendix D: Pledges and Commitments

To identify commitments as defined by the OECD, all projects with a status of "pipeline" are split into pledges and commitments.[22]

The pipeline category is split according to the following criteria

Pipeline: Pledge (verbal, informal agreement)

Pipeline: Commitment (written, formal binding contract)

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
    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

---

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

Appendix E: 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."[23] 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.

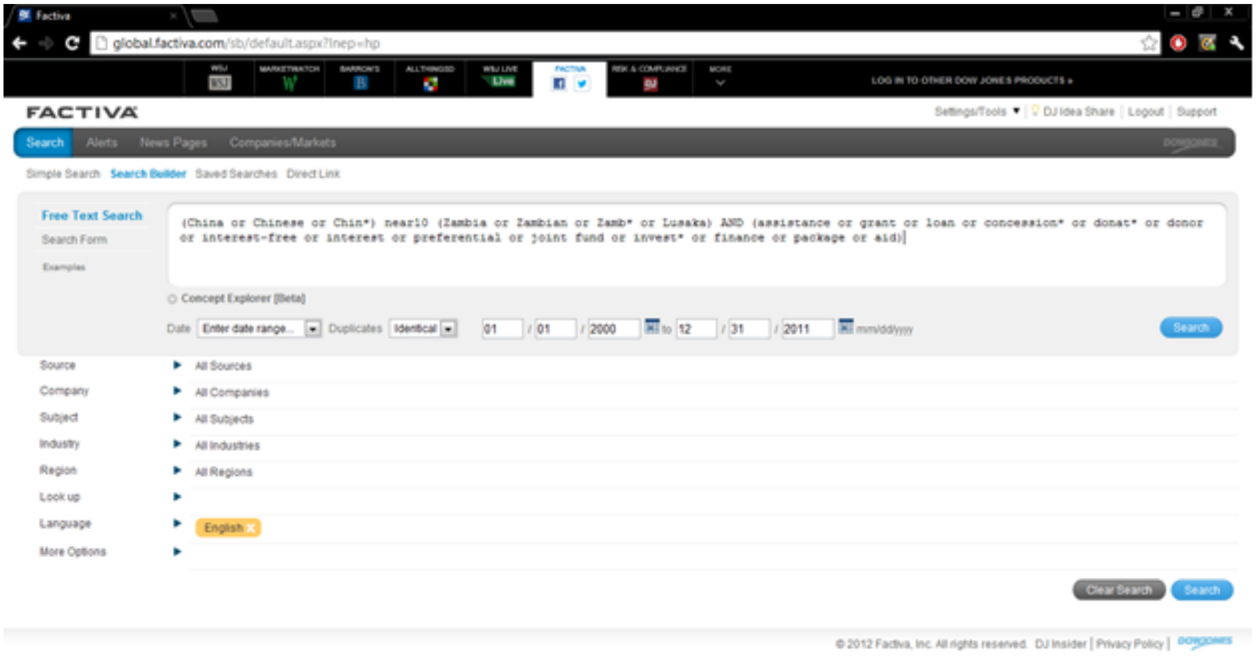
Each coder must systematically assess the intention of a given project based on specific criteria outlined below.

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

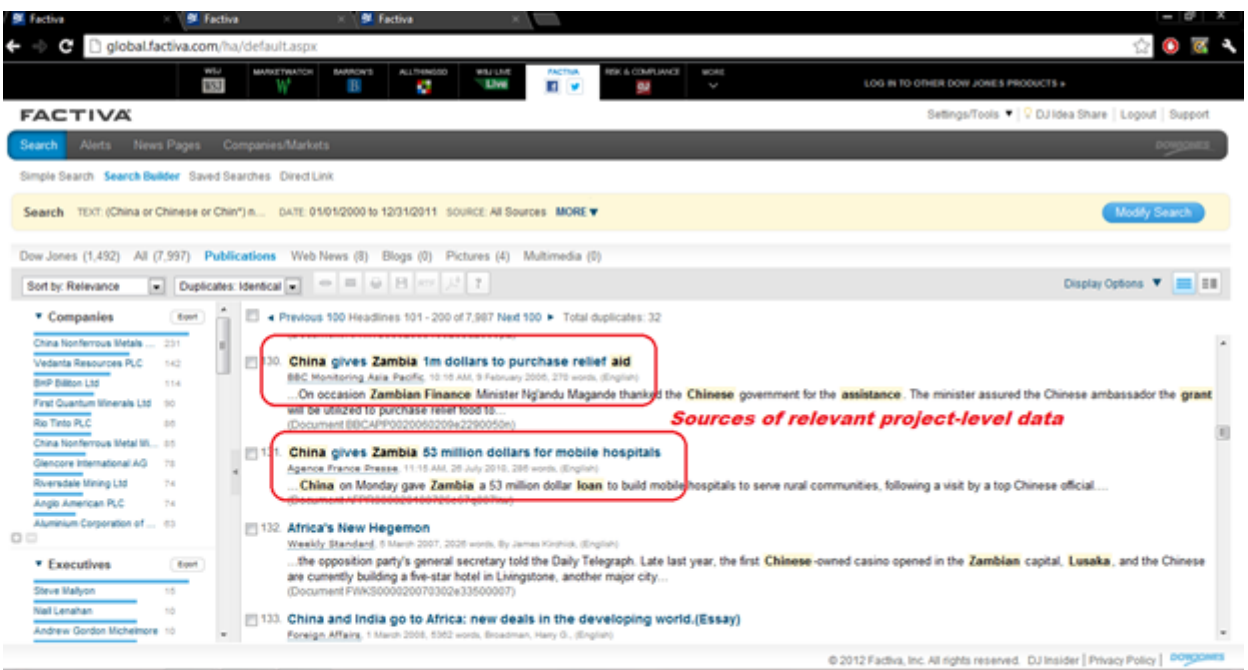
[23] "Is it ODA?" OECD Factsheet, November 2008. Accessed at: http://www.oecd.org/investment/aidstatistics/34086975.pdf

Appendix F: Screenshots to demonstrate search and data collection processes

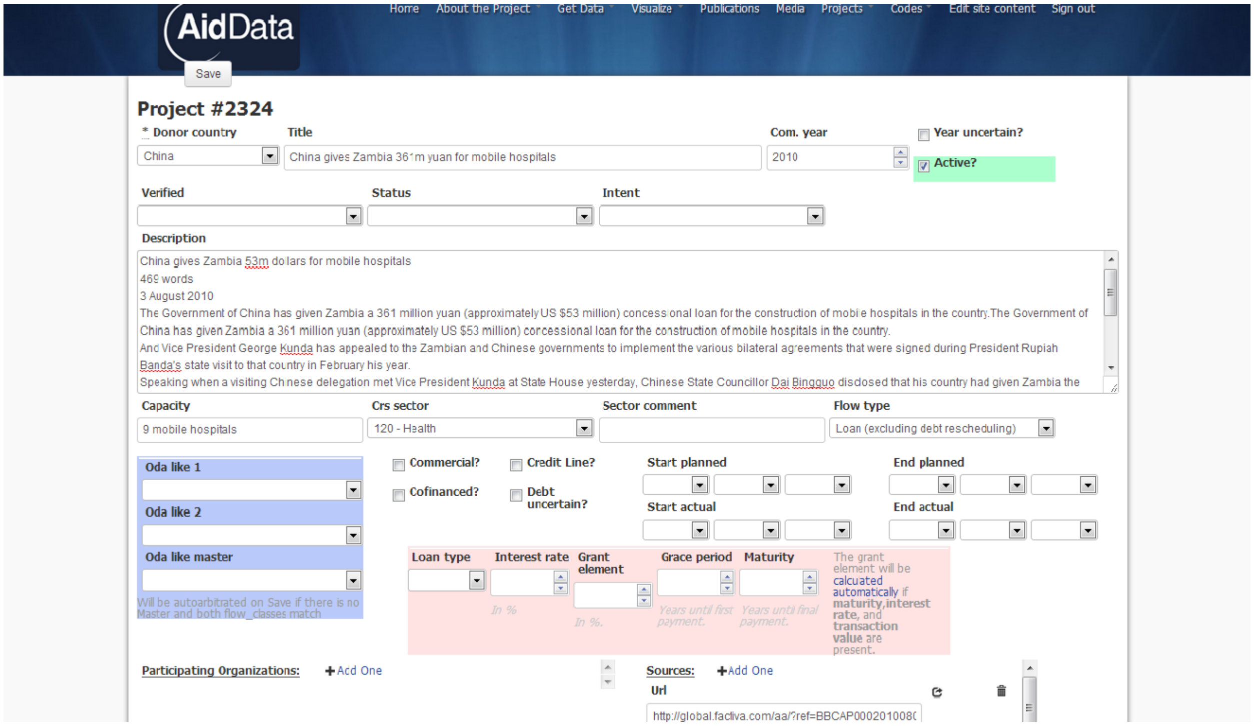
Screenshot #1: Stage One Factiva Search Terms



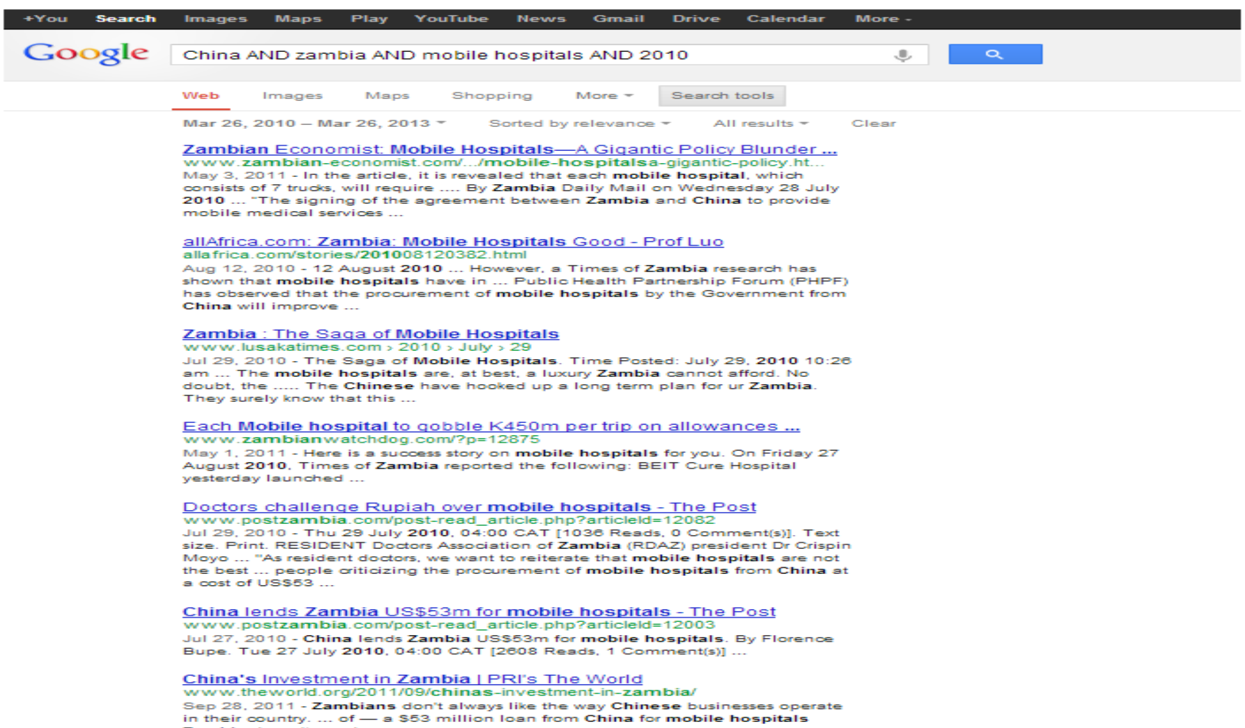
Screenshot #2: Stage One Factiva Search Yields



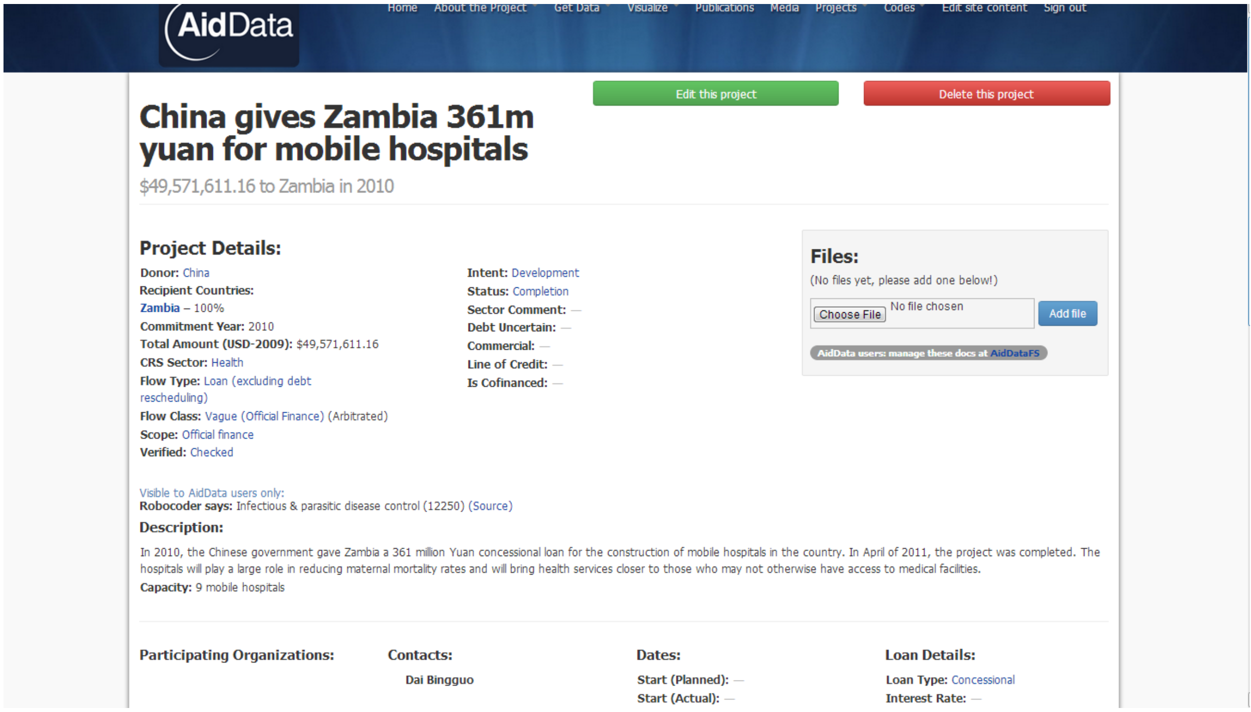
Screenshot #3: Create New Record to Input News Results from Factiva Stage One Searches



Screenshot #4: Google Search in Stage Two



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



Appendix G: Database Variable Definitions

Below is a glossary of the variables that AidData's TUFF 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

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 TUFF 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.

Published: The marker is only to be marked when a project has gone through all three stages of data collection and vetting process. Once a project is fully vetted, the program manager will check the marker and make the project published on the website.

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
- Grant
- 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
- Official Religious Aid 1
- Official Religious Aid 2

More details on these flow categorizations can be found in Appendix D. For details on Official Religious Aid, see Appendix O.

Note: While the above flow categories were designed to accurately represent the complexity of Chinese Official 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.

Umbrella: This field identifies projects where some agreement was signed between two countries but the funds were not allocated at the original agreement date, only to be allocated through smaller projects later on. Given the nature of TUFF data collection, it is likely that the smaller projects that relate to these umbrella projects were captured elsewhere in the dataset. To avoid double counting, we suggest users do not include rows that are marked as umbrella projects in financial analysis.

Recommended for Research: A binary marker that AidData includes to indicate whether the project is recommended for research and analysis. Only projects that meet the following criteria will be recommended: a) not an umbrella project; b) status is either "Pipeline: Commitment", "Implementation" or "Completion".

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."

Ground-truthed: A binary marker that indicates whether a project has been verified with in-country field research.

Organizations Variables

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: Cofinancing, 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.

Recipient Variables

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

Resources:

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: Government Sources (Donor/Recipient); Implementing/Intermediary Agency Source; Other Official Sources (non-Donor, non-Recipient) NGO/Civil Society/Advocacy; Academic Journal Article; Other Academic (Working Paper); Media Reports, including Wikileaks; Social Media, including unofficial blogs; Other.

Title: Name of the news report, journal article, or official reports.

Publisher: The ministry, news agency, NGO, or journal that has released the resource.

Publisher Location: The country of origin of the resource.

Author: When available, include the writer of the resource.

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_defl_2014: Deflated monetary equivalent of reported monetary amount in reported currency to 2014 U.S. Dollars.[24]

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: Represents the general nature of the loan. Categories covered include: interest-free; concessional; non-concessional; no information and some information.

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

Maturity: Reported duration of the project loan, in years

---

[24] To normalized financial commitments into 2009 dollars, we used GDP deflator data from the World Bank. Because the base year varies for some countries, we sometimes needed to adjust the base year to 2009. This change requires a simple algebraic operation. Find the value that when multiplied with the 2009 deflator will have a product of 100 (example 109.5217X=100). After finding X, multiply the other deflators by that value. If done correctly, the rate of change year to year will not change, but the base year will.

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).

Health of Record

Variables in this section provide information about the sources use to create project entries and missing fields and ambiguities in project records. See Appendix M for full scoring methodology.

Source Triangulation Score: Rates projects based on the types of sources (media, academic, donor/recipient government, other official, NGO).

Data Field Completeness Score: Rates projects based on the number of missing data fields critical to understanding the project. It also penalized projects with "Vague" makers like "Vague Official Finance" for flow class or "Vague TBD" for flow type.

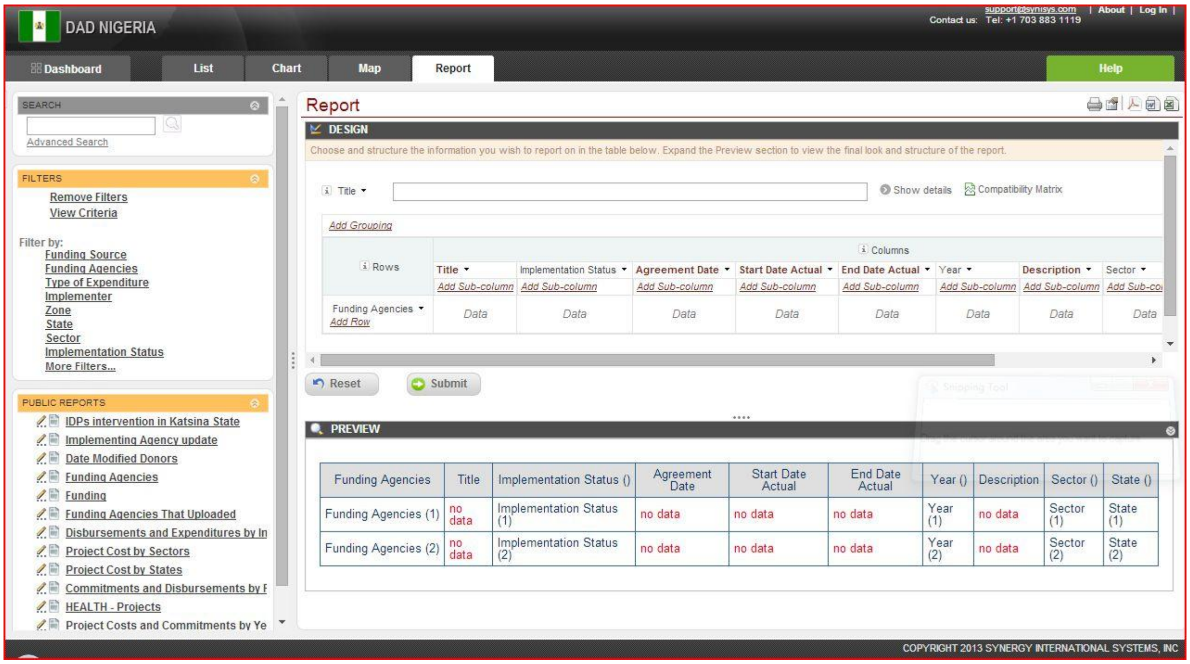
Appendix H: Mining Aid Information Management Systems

The diversity of recipient operated tracking systems for incoming aid means that RAs must have a range of strategies for extracting data on a donor of interest from these systems. This appendix offers several strategies for how to access donor-specific information on a typical aid information system.

Export all projects for as a spreadsheet

One way to extract project information for a donor of interest is to export all project records from an aid information management system. This option is attractive because it provides a stable list of projects for a recipient. Some AIMS will sometimes be removed from public view or have records removed without notification.

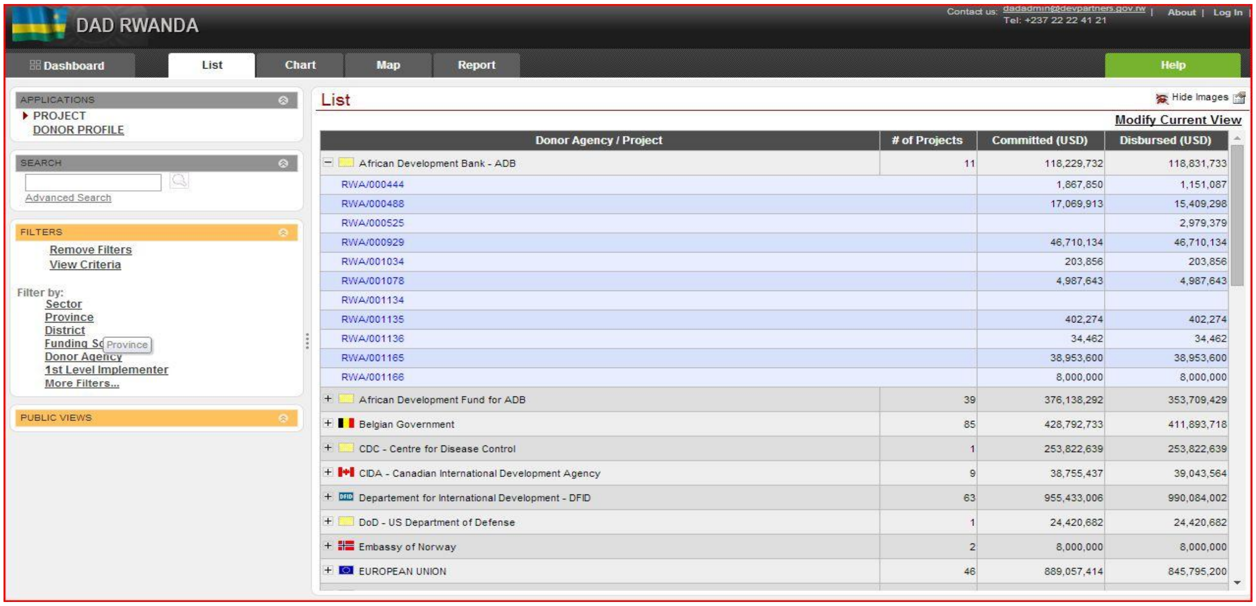
1. To export the all project records from an AIMS, go to the "Reports" page and select the project fields for export.
2. For the first column select "Funding Agencies." In the following columns, select all other variables provided (this will vary by system). These typically include: Title, Description, Commitment Amount, Disbursement Amount, Sector, Location, Status etc.
3. Once you have selected the variables of interest, export the finalized spreadsheet as a csv.
4. Next, the project manager or senior RA can assign the csv. to an RA to add new project records or amend existing entries.



Pulling information directly from AIMS

Because some aid information systems will not allow public users to export project-level data. In these instances, researches can still access project-level information on the AIMS to create records.

1. To access project-level information on an AIMS, go to the "list" page and find projects for the donor of interest.
2. From the "List" page, it is possible to access individual project records. These records contain all the information that would be included in an export of the database.
3. Transfer information from the project page into a new database entry on china.aiddata.org.


DAD RWANDA
Contact us: dadadmin@devpartners.gov.rw | About | Log In
Tel: +237 22 22 41 21
Dashboard
List
Chart
Map
Report
Help
APPLICATIONS
PROJECT
DONOR PROFILE
SEARCH
Advanced Search
FILTERS
Remove Filters
View Criteria
Filter by:
Sector
Province
District
Funding So
Donor Agency
1st Level Implementer
More Filters...
PUBLIC VIEWS
List
Hide Images
Modify Current View
Donor Agency / Project
# of Projects
Committed (USD)
Disbursed (USD)
African Development Bank - ADB 11 118,229,732 118,831,733
RWA/000444 1,867,850 1,151,087
RWA/000488 17,069,913 15,409,298
RWA/000525 2,979,379
RWA/000929 46,710,134 46,710,134
RWA/001034 203,856 203,856
RWA/001078 4,987,643 4,987,643
RWA/001134
RWA/001135 402,274 402,274
RWA/001136 34,462 34,462
RWA/001165 38,953,600 38,953,600
RWA/001166 8,000,000 8,000,000
African Development Fund for ADB 39 376,138,292 353,709,429
Belgian Government 85 428,792,733 411,893,718
CDC - Centre for Disease Control 1 253,822,639 253,822,639
CIDA - Canadian International Development Agency 9 38,755,437 39,043,564
Departement for International Development - DFID 63 955,433,006 990,084,002
DoD - US Department of Defense 1 24,420,682 24,420,682
Embassy of Norway 2 8,000,000 8,000,000
EUROPEAN UNION 46 889,057,414 845,795,200

Appendix I: Google Scholar Searches

Objective: Refine project entries with information from high-end publications, reports, and peer-reviewed journal articles. Increase diversity of resources supporting a project entry.
Process:

1. **Type search term into Google Scholar search ba**r
    a. Master Search

[donor country] OR [donor demonym] OR [donor countr*] AND [recipient country] OR [recipient demonym] OR [recipient countr*] OR [capital of country] 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
Example 1: China OR Chinese OR Chin* AND 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

Example 2: Qatar OR Qatari or Qat* AND Yemen OR Yemeni or Yeme* OR Sana'a 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

2. **Browse titles for relevant articles**
    b. Click on any articles relating to "international development" or "development assistance"
    c. Check if resource record exists - type [article URL] into china.aidata.org search bar in "Resources" page
    d. *If* article contains relevant project-level information:
        i. Create resource record for resource
        ii. Use "save" feature below article title to save to "My library" in Google Scholar
3. **Extract relevant information from articles**
    e. If project record exists:
        i. Match resource with project record
        ii. Revise project descriptions accordingly
        iii. Add additional information not found in previous resources
    f. If project record does not exist:
        i. Create complete record with relevant information
        ii. Cite resource
4. **Engage in relevant post-scraping work**
    g. Scrape "works cited" or "bibliography" page for *each* relevant article – repeat steps 2-4
    h. Use "related article" feature for *each* relevant article – repeat steps 2-4
    i. Use "cited-by" feature for *each* relevant article – repeat steps 2-4
    j. *Optional*: Use "Web of Science" feature for each relevant article

Advantages:
1. Provides a systematic process to scrape through high-end publications
2. Creates a repository for high-end publications through "My Library" feature
3. Generates indices of citations to track relevant academic literature through "Cited By" feature

Additional Features: Google Scholar Alerts
To further streamline the collection of high-end publications, we employ alerts for *general searches* in Google Scholar and for the *citations* index of each article. Doing so allows us to track the newest publications and further develop our repository of high-end publications.

For each search entry into Google Scholar:

b. Check "list alerts" to ensure that an alert does not already exist for the search entry
c. Create an entry – type general search into alert query

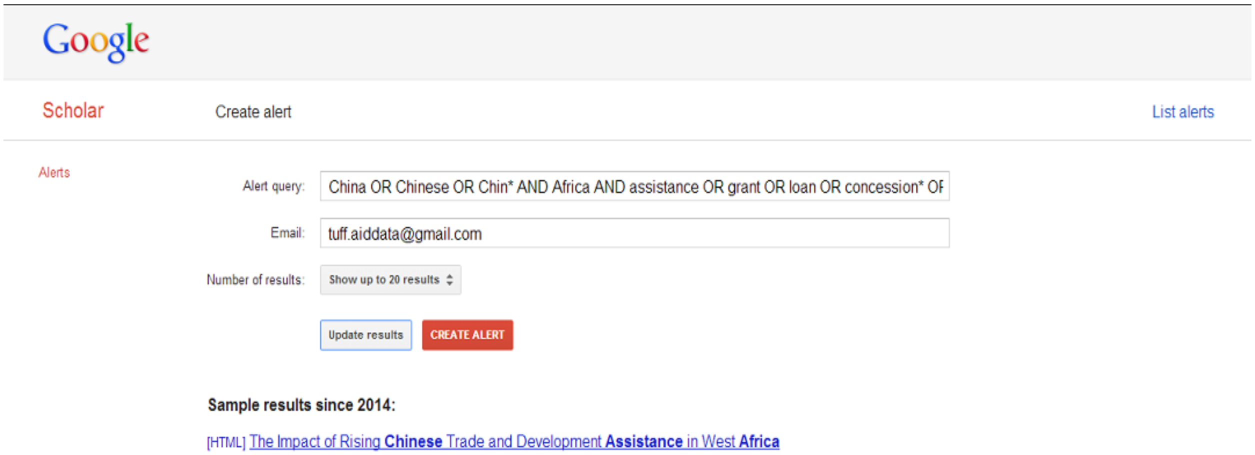


Figure 2: Example of Alert for Search Query

For each article's "cited by" page:

a. Check "list alerts" to ensure that an alert does not already exist for the article's "cited by" page
b. Click "create alert" at the bottom of an article's "cited by" page

Appendix J: Automated Decision Rules, Logical Inconsistency Checks and Project Templates

These coding rules are designed to be a "cheat sheet" for coders when tracking Chinese Official Finance. They should also be used by a project manager before the publication of a dataset to check for human coding errors.
To check if these any of these decision-rules have been violated, following the following steps.

- Use the "*Search and Filter*" page to find rule violations. For example, to check for violations of Export-Import Bank rules, you would select "Export-Import Bank of China" as a funding agency and check all "flow types" except "Loan" "Export Credit" or "Strategic/Supplier Credit."
- Export these projects and assigned an RA to correct the errors.
- Repeat this process for each decision-rule, illogical combination and project template until database fields are standardized.

Supplier: Certain government agencies are responsible for providing certain modalities of financing or targeting their financing in certain sectors. These decision rules draw heavily from the *2014 White Paper on Chinese Foreign Aid* and other high level academic research on China's Official Finance portfolio.

Export Import Bank of China

- If "funding agency" = Export-Import Bank, then "flow" = Loan OR Export Credit OR Strategic/Supplier Credit OR Debt Rescheduling[25] (United States Senate 2011)
- If "loan type" = concessional, then "funding agency" = Export-Import Bank (Corkin 2011; Davies et al. 2008; "China's Foreign Aid" 2014)
- If"funding agency" = Export-Import Bank, then "flow" ≠ any other flow type
- If "funding agency" = Export-Import Bank and sources indicate the loan is concessional, then "flow class" = "ODA-Like"

China Development Bank

- If "funding agency" = China Development Bank, then "flow" = Loan OR Export Credits[26]
- If "funding agency" = China Development Bank, then "flow class" = OOF-like (Downs 2011)
- If "funding agency" = China Development Bank, then "loan type" ≠ Concessional (Davies et al. 2008; cdb.com 2015)

China Africa Development Fund

- If "funding agency" = China Africa Development Fund, then "flow" = Official Investment [27] [28]

People's Bank of China

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

Other Logically Inconsistent Variable Combinations:

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

---

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

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

[27] *Brautigam 2009*, pg. 205.

[28] *Brautigam 2011*, pg. 4.

Project Templates:

Project Type	Sector	Flow Type	Flow Class	Intent	Notes
Medical Teams	Health	Free-standing technical assistance	ODA-like	Development	
Agricultural Technology Demonstration Centers	Agriculture, Forestry, Fisheries	Grant or Loan or Export Credit	Depends on flow type	Development	If loan terms uncertain, Vague OF
Special Economic Zones	Trade and Tourism	Joint Venture	OOF-like	Commercial	
Confucius Institutes/Classrooms	Education	Grant	OOF-like	Representational	
Military Aid	Government and Civil Society	Grant, Loan, Scholarship/training, Free-standing Technical Assistance or Export Credit	Military	Representational	
Training/Technical Assistance	Varies	Free-standing technical assistance	ODA-like	Development	
Debt Cancellation	Government and Civil Society	Debt Forgiveness or rescheduling	ODA-like	Development	If amount uncertain check "debt uncertain"
Malaria Treatment Centers/Anti-Malarial Drugs	Health	Grant or Loan or Export Credit	Depends on flow type	Development	If loan terms uncertain, Vague OF
De-mining	Government and Civil Society	Free-standing technical assistance	ODA-like	Development	
School Construction	Education	Grant or Loan or Export Credit	Depends on flow type	Development	If loan terms uncertain, Vague OF
Hospital Construction	Health	Grant or Loan or Export Credit	Depends on flow type	Development	If loan terms uncertain, Vague OF
Government Buildings	Government and Civil Society	Grant or Loan or Export Credit	Depends on flow type	Development	If loan terms uncertain, Vague OF
Road/port/railway construction	Transport and Storage	Grant or Loan or Export Credit	Depends on flow type	Development	If loan terms uncertain, Vague OF
Dam/Hydropower projects	Energy Generation and Supply	Grant or Loan or Export Credit	Depends on flow type	Development	If loan terms uncertain, Vague OF
Grants for "Economic and Technical Cooperation"	Unallocated/unspecified	Grant	ODA-like	Development	

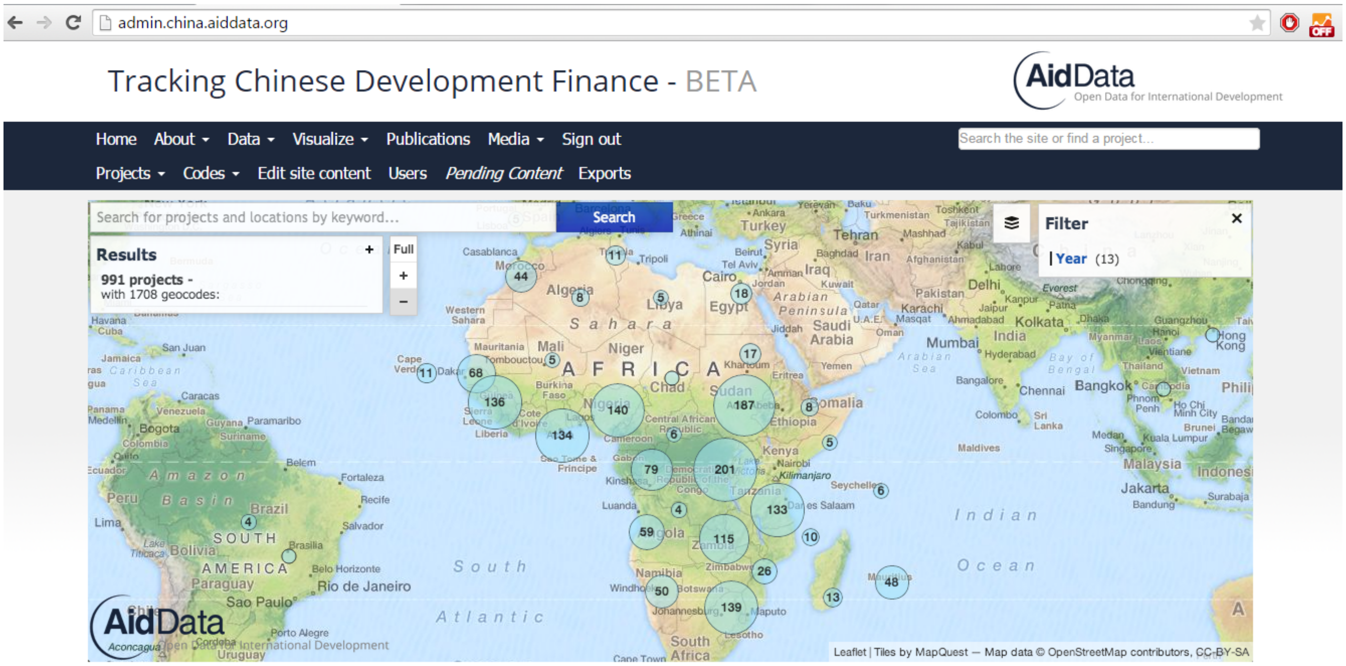
Interest-free loans for "Economic and Technical Cooperation"	Unallocated/unspecified	Loan	ODA-like	Development
Export Credits	Varies	Export Credit	OOF-like	Commercial

Appendix K: Georeferencing TUFF-Based Data

This document will provide instructions for geocoding any data generated using the TUFF methodology. Coders will use ccaps.aiddata.org/toolkit to geocode, following the same double-blind methodology with arbitration.

The Online Dataset

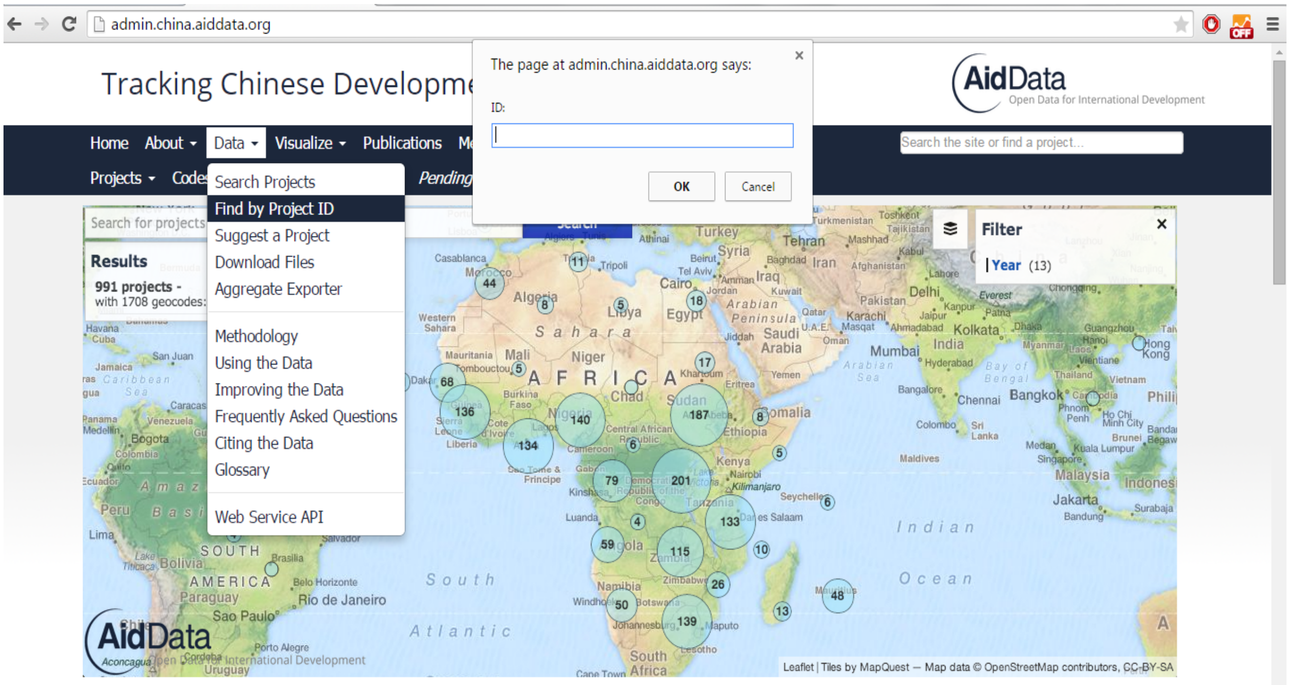
Coders will use the searchable dataset on admin.china.aiddata.org as the geocoding resource. **Project Descriptions** and



Sources will appear on every project page on the website—these specific fields will contain the location information from which you will code. Project descriptions are short summaries of news reports—usually a paragraph long—written by RAs on the TUFF team. The Sources are links to actual news reports from which the project descriptions are paraphrased. Project descriptions generally contain all relevant location information, but it's always a good idea to glance at the sources to verify that all location information has been identified.

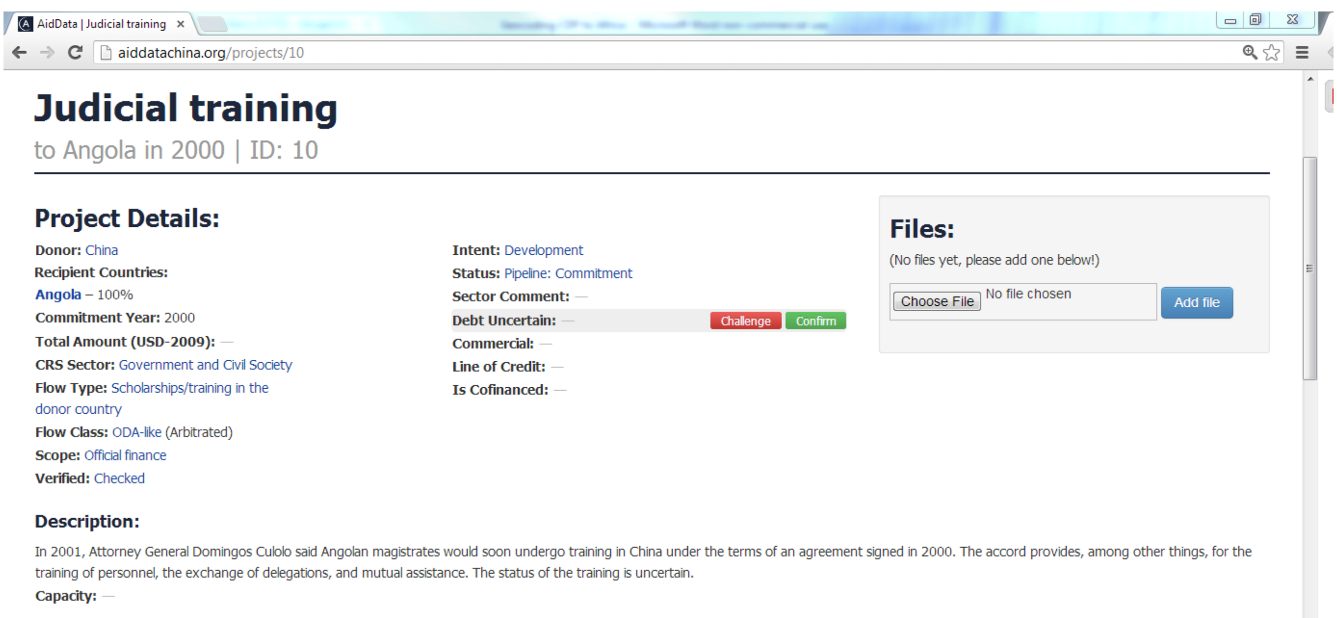
Using "admin.china.aiddata.org"

Once you navigate to the admin.china.aiddata.org, click on the **Data** tab on the top of the screen. A dropdown menu will appear. Click **Find by Project ID**.



A search field will appear in the middle of the screen. Fill in the project ID (found in your toolkit task list) and press enter to navigate to the unique project page.

Project pages will look something like this:



The title of the project will appear in bold at the top. **Make sure that the title, country, and project ID on the header match the title, country, and project ID of the project you're coding in the toolkit.** Once you've matched the project page to your toolkit project, follow these general steps:

4. Read the **Description** field carefully for location information. You may write this description in the "Activity Description" field in Geo Coding form.
5. Next to **Recipients** field, in parentheses, is a field displaying any sub-national data for a project. Record all this information.

6. Read all **Sources** to confirm location information. Remember, the project description is a guide. Always default to sources, as they are the media reports that ultimately validate each project. If you have access to Factiva sources, please use this Factiva source to geocode. However if you do not have access to Factiva and it is the main source for the project you are geocoding, please flag the project. Please provide the source name and the URL in the "Source" and "Source URL" fields. If a project has multiple sources, enter main source in these fields and enter all other sources in the "Notes" field.
7. Geocode each location in the toolkit. If the description and sources lack sub-national location information, use a national level precision code ("6" v "8" depending on end user).

Infrastructure Projects

When projects mention infrastructure that has already been completed, code them as precisely as you can. Coding infrastructure projects at the city level is adequate, but if it is possible to add a more specific "street-level" latitude and longitude, please do so. This will be easier some types infrastructure (stadiums, hospitals) and more difficult for others (aquaculture research centers). Follow the same procedures you would normally follow to add new location to the toolkit (using google earth and the geonames database). The general goal is to take steps towards making the data "crowdsource-friendly".

Appendix L: AidData's Deflation Methodology

1 Currency Conversion and Deflation Purpose

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

2 Exchange Rates

2.1 Exchange Rate Methodology

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

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

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

3 Deflators

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

*Inflation * Change in Buying Power = Deflator*

3.1 Inflation

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

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

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

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

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

$P_{year}\ I_{year}$

[29] Available at http://www.oecd.org/dac/aidstatistics/informationnoteonthedacdeflators.htm.

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

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

P_{year} I_{year}

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

The following sources are used to compile the inflation values.

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

3.2 Change in Buying Power

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

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

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

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

$$1273.9 / 804.4 = 1.58$$

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

3.3 Finalized Deflators

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

Inflation * Change in Buying Power = Deflator

4 Examples on Using GDP Deflators

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

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

Next, divide it by the AidData deflator:

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

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