AidData: Tracking Development Finance

The College of William and Mary, Brigham Young University, and Development Gateway

Daniel Strandow, Michael Findley, Daniel Nielson, and Josh Powell

16 June 2010

ISBN: -------

The UCDP, AidData, and World Bank Institute codebook on geo-referencing World Bank Aid Projects

Version 1.1[1][2]

[2] This World Bank user's manual has been adapted from the UCDP/AidData geo-referencing aid codebook (2010)

1 Introduction

This codebook details how aid events that are available from the World Bank Operations Portal can be assigned latitude and longitude coordinates, i.e. be geo-referenced, under the UCDP/AidData coding rules. The rules are derived from the UCDP Geo-referenced Event Dataset (GED) Codebook version 1.0 (Sundberg et al., 2010) which covers the geo-referencing of violent events. The system has been adapted and complemented by additional rules to enable the coding of aid projects rather than battles. The UCDP GED is used as a starting point as it permits us to identify and record a hierarchy of locations differentiated by various precision scores.

Sources vary in the precision that locations are reported; sometimes the exact location is named and in other instances the general area is reported. Following UCDP, the system of geo-referencing used by UCDP/AidData can therefore cope with coordinates in four main levels, ranging from point locations, through two administrative divisions, to the country level. Eight precision categories are connected to the coordinates in order for researchers to select subsets of the dataset that contain different levels of precision. The main objective is to record all locations to which aid dollars are committed or distributed. Locations that benefit indirectly are not coded, unless the geographic locations of the indirect areas are significant enough to be clearly spelled out in project documents.

2 Motivation

The availability of aid locations at a sub-national level provides four main benefits to donor organizations: ease of management, dialogue with recipients, coordination with other donors, and increased transparency.

Ease of Management

Localized data can provide donors with a better understanding of the allocation of aid within the country, highlighting any potential financing gaps, displaying inequities of aid distribution, and ensuring that aid flows to those who need it most. Once mapped, city-level data presents management with a simple platform to assess its current performance and to plan its future projects.

Dialogue with Recipients

If made public, mapped sub-national data can provide recipients with a better understanding of how and where aid should be working for their benefit. Recipients will be better able to carry on a dialogue with donors to direct aid to areas which may have been otherwise neglected, and which need the aid most.

Coordination with Other Donors

Once multiple donors have made their mapped data available to the public, donor organizations will have a clearer picture of the aggregate distribution of aid within a country. Donors will be able to coordinate their efforts to specialize in certain regions, or to co-finance projects, eliminating potentially wasteful project duplication.

Increased Transparency

Mapped data will also hold both donors and recipient governments accountable to their intended beneficiaries. Researchers, auditors, and recipient citizens will be able to verify that projects are being implemented in their intended locations, reducing waste and fungibility and increasing aid effectiveness.

3 Locating aid events in time and space

The geo-referenced dataset that has so far resulted from this coding scheme is compatible with the original version of AidData as well as a variety of donor sources and GIS systems. This means that UCDP/AidData's unit of analysis is primarily funding commitments as opposed to distributed aid or calendar days (PLAID, 2010). Since data on the exact dates of funding commitments are sparse most geo-referenced locations can only be related to the year that a specific commitment was made.

In order to preserve AidData's unit of analysis, and the corresponding IDs, aid projects intended for several locations are not divided into several new events with multiple project IDs. Instead, when there is more than one location per project, we include an additional row of data with the same project information, but separate sets of coordinates for every additional location.

4 About the geo-referencing of locations

Each aid project may have location information on several levels. First, the project abstract often contains some location information, but may not reflect all of a project's intended targets.   After searching through the abstract for location information, the most recent updates of all project documentation are reviewed (this can include Project Appraisal Documents, Project Papers, etc.). In geo-referencing the AidData data set, titles and descriptions are also reviewed for relevant geographic information. Location information is frequently contained in more than one project document, necessitating a review of all possible sources of information. For this reason, unless a document definitively states that all locations are contained therein, the most recent of each document type is reviewed.

In the data set, the first column contains the number of locations that the project reaches (numbloc). The second and third columns in each set of coordinates contain the latitude (lat) and longitude (long) of the location. The fourth column in a set specifies the first-order administrative division (ADM1) and the fifth a second order division (ADM2) as a string variable. In addition, the official Geoname and Geoname ID (from geonames.org) are included in the sixth and seventh columns. If the information of an event only gives information on the administrative division, and not the exact location, then the centroid point of the administrative division is entered into the latitude and longitude columns. Lastly, in each set of coordinates, the precision of the coordinates is specified (Precision), which we discuss in greater depth below. For additional locations, the project information is replicated in the following row, then the next set of coordinates and Geonames is entered.

Figure 1 illustrates how the coding sheet looks for a project in which the recipient locations are clear and straightforward to geo-reference. The example references part of a World Bank transport improvement project with eleven locations, including Miritini and Kisumu, Kenya.

numbloc	Lat	Long	ADM1	ADM2	Geoname	GeoNameID	Precision
11	-4	39.56667	Coast	Mombasa	Miritini	177896	1
11	-0.1	34.75	Nyanza	Kisumu	Kisumu	191245	1

Figure 1 Example of clear locations.

In Figure 1, the coordinates of the first location are recorded in the first and second column. The fourth and fifth columns display the province and district in which the points are located. The sixth column is the Geoname of the point location, while the seventh column is the unique Geoname ID for the point. Finally, the precision code of 1 indicates that the points correspond to a specific place, in this case a town.

If there is no direct mention of any location in the sources, and the title and abstract do not indicate that aid is granted to the central government or national in nature, aid is assumed to go to the country in general or to the main administrative centre (most often the formal capital) depending upon the nature of the project. The country coordinates are coded with precision 7 which indicates that the location is unknown. This means that it is up to the researchers using the data to decide if unclear aid locations should, by default, be excluded, or be assumed to go to the entire country. See the advanced rules in section 6.

The coordinates are determined through Geonames, and the American National Geospatial Intelligence Service (NGA) is used as a secondary data source. Geonames and NGA both provide online services, namely www.geonames.org and the GEOnet Names Server (GNS), which contains names and coordinates of various administrative divisions, populated places, waterways, and objects. (http://geonames.nga.mil/ggmagaz/geonames4.asp) The latitude and longitude coordinates are recorded with a six decimal precision. The map projection used is the standard World Geodetic System 1984 (WGS 84) (Sundberg et al, 2010). Google Earth is also utilised in order to complement the primary and secondary sources. Note that some cases have no second order administrative divisions connected to the first order administrative divisions in GNS. Geonames contains this information more often, and second order divisions are included whenever they can be located, but for this version of the dataset the focus is on collecting information on the first order administrative divisions for all locations under the country level.

5 About the precision categories

For the data to be useful for a wide range of applications it is crucial to make it possible to select sub-sets of the data based on varying criteria of precision. The first six categories detailed by the UCDP's Georeferencing Project Codebook (Sundberg and Lindgren, 2009, 13) are used here, with minor modifications. The seventh and eight precision categories are unique for the UCDP/AidData codebook. Precision categories 4.1, 4.2, 8.1, and 8.2 were developed to differentiate between projects that work at a local level and those that occur exclusively at a national or provincial scale.

1   The coordinates corresponds to an exact location, such as a populated place or a hill. The code is also used for locations that join a location which is a line (such as a road or railroad). Lines are not coded only the points that connect lines. All points that are mentioned in the source are coded.

2   The location is mentioned in the source as being "near", in the "area" of, or up to 25 km away from an exact location. The coordinates refer to that adjacent, exact, location.

3   The location is, or lies in, a second order administrative division (ADM2), such as a district, municipality or commune

4.1   The location is an entire first order administrative division (ADM1), such as a province, state or governorate.

4.2   The location is within a first order administrative division (ADM1), such as a province, state or governorate, but the location within the ADM1 is unknown.

5   The location can only be related to estimated coordinates, such as when a location lies between populated places; along rivers, roads and borders; more than 25 km away from a specific location; or when sources refer to parts of a country greater than ADM1 (e.g. "northern Uganda").

6   The location can only be related to an independent political entity, meaning the pair of coordinates that represent a country.

8.1   The location represents funding to a government agency.

8.2   The location represents funding to a government agency, but this is not the only location for the project.

6 Basic rules for geo-referencing

For a step-by-step explanation of the geo-coding process, see appendix A.

Geonames

Enter the location name into the Geonames search engine, selecting the desired country. If there is no hit, try an advanced search of feature classes or use the "fuzzy search" feature. If there are still no results, search again using NGA GNS. GNS often contains more alternate spellings than Geonames, and those alternate spellings can be re-entered into Geonames to obtain the Geoname and Geoname ID.

NGA GNS

Locations found through the GNS, Google Earth, or any other alternate source must be added into Geonames prior to geocoding (for instructions on updating Geonames, see Appendix E). Enter the location into the NGA GNS using the category "start with" and input search string "without diacritics". If there is no hit in the GNS, search again using the category "fuzzy search" (Sundberg and Lindgren, 2010). If there are still no results, search again using Google Earth or Google Maps. Sometimes Google is better in suggesting options for spelling names that are misspelled in the sources. Use the coordinates of the location with the modified spelling, if it appears to be the place intended in the source. Left click on the pair of coordinates suggested in the gazetteer in order to obtain the location in decimal form rather than degrees, minutes, and seconds. Only code a specific location once per row. For instance if there are funds going to farms somewhere in the location Bengo, as well as aid to hospitals somewhere in the same location, then Bengo is coded only once.

Area locations

Administrative divisions (ADM1 and ADM2) and countries are areas. The latitude and longitude representation of areas are estimated as the coordinates of the centroid point and the Geoname ID provides a boundary file for the administrative divisions relating to the point. The names of first order administrative divisions (ADM1) and second order administrative divisions (ADM2) are saved in the data as text/strings in the "ADM1" and "ADM2" columns. The precision code depends on the level of the area (3, 4, 6/7).

A particular problem when working with longer time series is that there are states that have revised the boundaries of their administrative divisions at some occasion during the years that are being

coded. The priority in such instances is to best approximate the area that is intended in the source. Thus, if a province is divided into several new provinces, each of the new provinces within the boundaries of the defunct province is coded. A more difficult case occurs when countries decrease the number of provinces. In this case, the current province which contains the territory of the defunct province is coded and a note is made of the defunct province as the intended recipient.

Point locations

Locations that are discrete points are for instance towns, cities, suburbs, hills, farms, and various installations. Occasionally the source will mention a location within a location, for instance a hospital in a capital. Since that does not happen often, and since the coordinates of the main location are the only ones available in most cases, record only the coordinates of the main location. Suburbs are borderline cases. Suburbs to cities should be considered to be locations in their own right and are coded if the coordinates are available (with precision 1). If the coordinates of a major suburb are not available use the coordinates of the main city and precision 2.

If a location cannot be pinpointed via the search functions in Geonames, the GNS, or Google Earth, for instance a dam, then use the coordinates of the closest populated place rather than for instance estimating a point in the lake through Google Earth. In that case use the appropriate precision code (2 or 5). However, if the dam itself can be visually located via Google Earth, then estimate the coordinates of the centre of that exact location and add a geographic reference into Geonames (see Appendix E).

When coding point and line locations, also record the names of the related administrative divisions, if possible. Sometimes those names are not available from the gazetteer. If that is the case, use alternate sources such as fallingrain.com to determine the relevant administrative division. First order administrative divisions must be coded when a coordinate represents a sub-provincial feature, while second order administrative divisions may be left vacant if a brief search does not return reliable results.

Line locations

If the location is a road, or a power line or similar connection between points, then code the point locations that are mentioned in the source as linking the road. Sometimes the name of the road (for instance Beira-Machipanda) is not the stretch of the road that receives the funding (for instance the

Inchope-Machipanda section). If the source specifies the stretch of road which receives the funding, then record all towns that are explicitly mentioned in the source and that connect the road as well as any provinces through which the road passes.

If connecting towns are not indicated, record the second-level administrative divisions (ADM2s) in which the road passes, if possible. If ADM2s cannot be determined, simply record the provinces through which the road passes. For instance, a road running from Nairobi, Kenya to Mombassa, Kenya would necessitate five codes: (1) Nairobi, with Nairobi Area listed as the ADM1 and precision code 1; (2) Mombassa, with Coast Province listed as ADM1 and precision code 1; (3) Nairobi Area Province with precision code 4; (4) Eastern Province (through which the road passes) with precision code 4; and (5) Coast Province with precision code 4. This is done to reflect the fact that funding is allocated across the entire road through each affected province, rather than simply to the point locations of the road ends.

Islands, peninsulas and parks

When encountering islands, peninsulas or parks the proper precision code might be unclear. Do not code the island, peninsula or park separately if it is clearly a part of another location. For instance, Manhattan is a part of New York City and would not be counted as an Island or a suburb. For other situations use the available codes:

1= The island, peninsula or park is very small, or a particular point on an island is named and can be coded;

2= If a very small island, peninsula or park, or a point on a big area, is not specified more than near a point; If a island, peninsula or park falls entirely within an ADM2 it can be given this precision.

5=The island, peninsula or park is a bigger region that contains more than one ADM1; or if its location is unclear and further than 25 km from a named location; or if the coordinates can only be estimated between two named locations; If the island, peninsula or park falls within one ADM1, but more than one ADM2, the park should be given a precision 5 and the ADM1 should be filled in.

6=The island, peninsula or park is an independent political entity.

7 Advanced rules for geo-referencing

The advanced rules are designed to support the coder when sources are vague or unclear about which locations receive funding. The best option is always to try and find better sources with information on the location. When that is not possible, due to time or other constraints, advanced rules are used to code vague, unclear, and ambiguous locations. The advanced rules have been developed by weighting two criteria that are meant to balance each other:

- Be conservative in the coding and assign aid to larger or otherwise more significant locations rather than smaller and insignificant locations.
- Strive to locate coordinates that as much as possible reflect real locations (like towns) rather than artificial (like the centroids of administrative divisions).

Locations with ambiguous names

The sources often include the name of a location but are very sparse in specifying what type of location it is. In some cases that scarcity in specifying whether a location is an object, a town, or an administrative division, is combined with a wide selection of places that are named the same. In this case, perform a search with the GNS and select the location that is verified by GNS (i.e., has the same latitude and longitude). If an identical point cannot be located through GNS, use an additional source such as Google Earth or fallingrain.com.

When feasible, the best option in ambiguous cases is to consult area experts who know which place is most likely intended. In Mozambique there are, for instance, a number of places called Pemba. For someone who has lived in Mozambique the most likely location is the Pemba which is the seat of the administrative division Cabo Delgado. Until area experts can be consulted coders will however need to use the rules listed in the rest of this section.

One form of ambiguity is when there are several options in Geonames or the GNS, but neither is spelled the same as the location mentioned in the project's source documents. The source can for instance mention the location "Lang port" and the options in the gazetteer that is closest to that spelling may be "Lange" and "Langa". In this case, if Langa is closer to water than Lange, then go with Langa (since Lang has got a port according to the source). If it is difficult to determine whether Lang is just

misspelled in the source or if it is a place which is just not available in the gazetteer or Google Earth, the precision category will be a 6 (country level). The location is therefore treated as being no more exact than the country level but the potentially more exact coordinates are still saved for future reference.

When all options fail to provide the location in the source, there are other rules to follow, which are listed in Figure 2. These rules provide rules for arbitrating between two possibilities based on a prioritization of feature classes and are based on the two balancing criteria that are mentioned in the section above. When both potential points have identical feature classes, alternative sources should be used to corroborate one of the points. If other sources match both points or neither of the points, other factors should be considered such as the location of other points for the same project. For instance, there are three towns named Sabon Birnin in Nigeria, one each in the provinces of Kaduna, Sokoto, and Kebbi. If all of the other locations in the project documents are found in Kaduna province, the Sabon Birnin in Kaduna would be the most likely target, and should be selected. Once the location has been arbitrated satisfactorily, use a slightly more conservative precision code for the selected geographic feature (i.e. 2 for populated place and 5 for a district or province).

	Object	Populated place, unofficial name	Populated place, official (BGN) name	Seat of administrative division	Capital	ADM2	ADM1
Object	Use alternate sources and consider locations of other project components	Choose the populated place. Precision code = 1.	Choose the populated place. Precision code = 1.	Choose the seat of the administrative division. Precision code = 1.	Choose the capital. Precision code = 1.	Choose the centroid of ADM2 Precision code = 3.	Choose the centroid of ADM1. Precision code = 4.
Populated place, unofficial name	..	Use alternate sources and consider locations of other project components	Choose the populated place, official (BGN) name. Precision code = 1.	Choose the seat of the administrative division. Precision code = 1.	Choose the capital. Precision code = 1.	Choose the centroid of ADM2 Precision code = 3.	Choose the centroid of ADM1. Precision code = 4.
Populated place, official (BGN) name	..	..	Use alternate sources and consider locations of other project components	Choose the seat of the administrative division. Precision code = 8.	Choose the capital. Precision code = 8.	Choose the centroid of ADM2 Precision code = 3.	Choose the centroid of ADM1. Precision code = 4.
Seat of administra-tive division	..	..	..	Use alternate sources and consider locations of other project components	Choose the capital. Precision code = 8.	Choose the seat of the administrative division. Precision code = 8.	Choose the seat of the administrative division. Precision code = 8.
Capital	..	..	..	..	..	Choose the capital. Precision code = 8.	Choose the capital. Precision code = 8.
ADM2	..	..	..	..	..	The most likely is selected. Precision code = 4.	ADM1 is selected over ADM2. Precision code = 4.
ADM1	..	..	..	..	..	..	The most likely is selected. Precision code = 6.

Figure 2 Selecting a location when several alternatives have the same name

Vague area locations

In some instances the source will mention "most provinces" or similar vague locations. It is precise enough that the coder knows the general region that has aid committed to it, but unclear enough that the exact areas are hard to pinpoint – hence the term vague. When feasible, this project should be flagged by the coder (and temporarily left uncoded) for a supervisor to contact the implementing agency for clarification. If resources do not allow for specific inquiries when some vague geographic information is available, like "northern provinces", then code all provinces along the northern border (with precision 4). The principle is to be conservative, ensuring that the correct provinces receive a geographic reference.

When it is possible to locate a centroid point for a vague/large region through Geonames or the GNS, or by estimating it, precision code 5 is used. The reason is that a centroid of a large region is less precise then for instance two centroids of two provinces. An example is in Guinea where Haute Guinea refers to an "upper" region in Guinea. The GNS has a centroid point so the coordinates for that point is used along with precision code 5. When using precision code 5, the ADM1 field should be left blank and each province containing a portion of that geographic feature (such as a National Park or Ecological Reserve) should receive a separate code (similar to the rules for coding roads through multiple provinces). In many instances it is impossible to estimate which areas are intended (for instance "most provinces"). In such cases, when it is known that the locations are areas that appear to cover a big unspecified swath of land, code the country level coordinates once and use precision code 6.[3]

Cases with at least one clear location and one vague area location

If the source specifies one location and then notes that other locations will also receive aid, for instance "most provinces", then do not forget to code those vague cases. In that case there will thus be two sets of coordinates, one for the explicitly mentioned source, and one for the country that collects the unspecified locations.

Unclear locations

A location is unclear if it is not certain that it is an area and if the exact location cannot be determined. Even if there is no direct mention of any location in the source material, a project is still geo-referenced. Users of the dataset can therefore decide whether or not, or how, to include unclear locations in

---

[3] If the source mentions locations that are unspecified points then see the rules for "cases with no clear location".

analyses. An example of a project with an unclear location is a World Bank local government support project in Tanzania which requires that local governments apply and be approved for support. In order to properly code this project, the coder should temporarily apply a national code with precision 7, then email the project manager and request that the Task Team Leader (TTL) be contacted for clarification.

numbloc	Lat	Long	ADM1	ADM2	Geoname	GeoNameID	Precision
1	-6	35			United Republic of Tanzania	149590	7

Figure 3 Example of unclear location.

Precision code 7 indicates that locations in unspecified areas somewhere in the entire country could be recipients.

On the other hand, if the source makes it clear that a central government institution receives the funding then the capital is coded, albeit with precision 8.1, or, if only a component of the project is to be distributed to an institution of the central government and the project has additional locations, code the country with a precision 8.2.[4] See figure 4.

numbloc	Lat	Long	ADM1	ADM2	Geoname	GeoNameID	Precision
1	-4.2669	15.28327	Brazzaville		Brazzaville	2260494	8.1

Figure 4 Example of when a central government institution receives funding

Finally, if the source makes it clear that funding will take place throughout the country, the country coordinates are used with precision 6.

numbloc	Lat	Long	ADM1	ADM2	Geoname	GeoNameID	Precision
1	8	-5			Republic of Côte d'Ivoire	2287781	6

Figure 5 Example of when a project is national in scope

---

[4] One could assume that precision code 7 should be used rather than 8 in this case. It is however more convenient for users to easily locate the most unclear cases by just selecting all observations where column "Precision" is 7. Furthermore, assuming that the capital is the recipient if a government institution receives the funding, is a much clearer decision than assigning the national level only, when rural locations are recipients.

Cases with at least one clear location and one unclear location

If the coordinates of one location of a project (a row in the dataset) are known and if there is one additional location in that event which is unclear (like "some towns") then, unlike when areas are merely vague, that unclear place is just not coded. No assumed country level or capital coordinates or anything similar are added in the way that is done when the entire event is unclear.[5]

However if it is entirely clear that one part of the funding goes to unnamed populated places *throughout* the country, then the country level is also coded (with precision 6). In those cases it must however be clear that the locations cover most of the country to warrant inclusion alongside a clear location with precision code 6.

A summary of coding rules for vague or unclear locations

Figure 6 lists which coordinates and precision categories to choose in three different situations.

Precision code	Capital is assumed to be location if:	Country is assumed to be location if:
8.1		The project title or description makes it clear that the central government ministries or financial institutions are the recipients.
8.2		Central government ministries or financial institutions are the recipients of one component of the project, but the project also contains other locations
7		The project title or description fields suggest that aid goes to a rural area (and most likely not to the capital).
		If the capital has no access to waterways and if the project title or description indicates that the aid goes to harbour activities.

---

[5] The clear location is considered to provide enough location information. If a set of unclear coordinates for the country and the capital would be added every time there is a completely unclear location mentioned parallel with clear locations – such a rule would complicate the coding and the use of the data well beyond its benefits.

Note that that "central government ministries or financial institutions" only include those that are fully controlled by the government. Government "programs" or similar cooperative arrangements or organisations, which are composed of a wide variety of NGO's, central or local government agencies, and companies are not assumed to be located in the capital (with precision 8). Such recipients are considered unclear and coded as such (country coordinate with precision code 7).

References

PLAID 1.9.1 Codebook and User's Guide. February 2010 Development Release of the

PLAID Database

Sundberg, Ralph, Mathilda Lindgren and Ausra Padskocimaite 2010. UCDP Geo-referenced Event Dataset (GED) Codebook. Version 1.0, 6 April.

National Geospatial Intelligence Agency (Updated 2010) GEOnet Names Server (GNS) http://geonames.nga.mil/ggmagaz/geonames4.asp

GeoNames. www.geonames.org

Google Inc. (2009) Google Earth (Version 5.0) http://earth.google.com/index.html

Google Inc. Google Maps http://maps.google.com/

Strandow, Daniel and Josh Powell, Jeff Tanner, Michael Findley 2010, The Geography of Foreign aid and violent armed conflict. Manuscript

Appendix A: The Geo-Coding Process

Step 1: Open Assignments Sheet in Google Documents

The coding assignment sheet allows management to assign World Bank projects to coders and allows coders to submit their work remotely.

Step 2: Open "View Project" Link

Link is drawn directly from the World Bank's operations portal and connects to the main project page, allowing access to project abstract and documents.

Step 3: Select Desired Project Document

In this case, the coder is accessing the Integrated Safeguards Data Sheet (ISDS). When coding a project, the ISDS typically provides a quick, but incomplete overview of project locations. Often, the ISDS is restricted to mentioning provinces, but neglects district, city, or other sub-provincial location information.

Step 4: Open PDF File of Project Document

The PDF version of each project document contains searchable text, making it easier to find location information.

Step 5: Search "locat" within Project Document
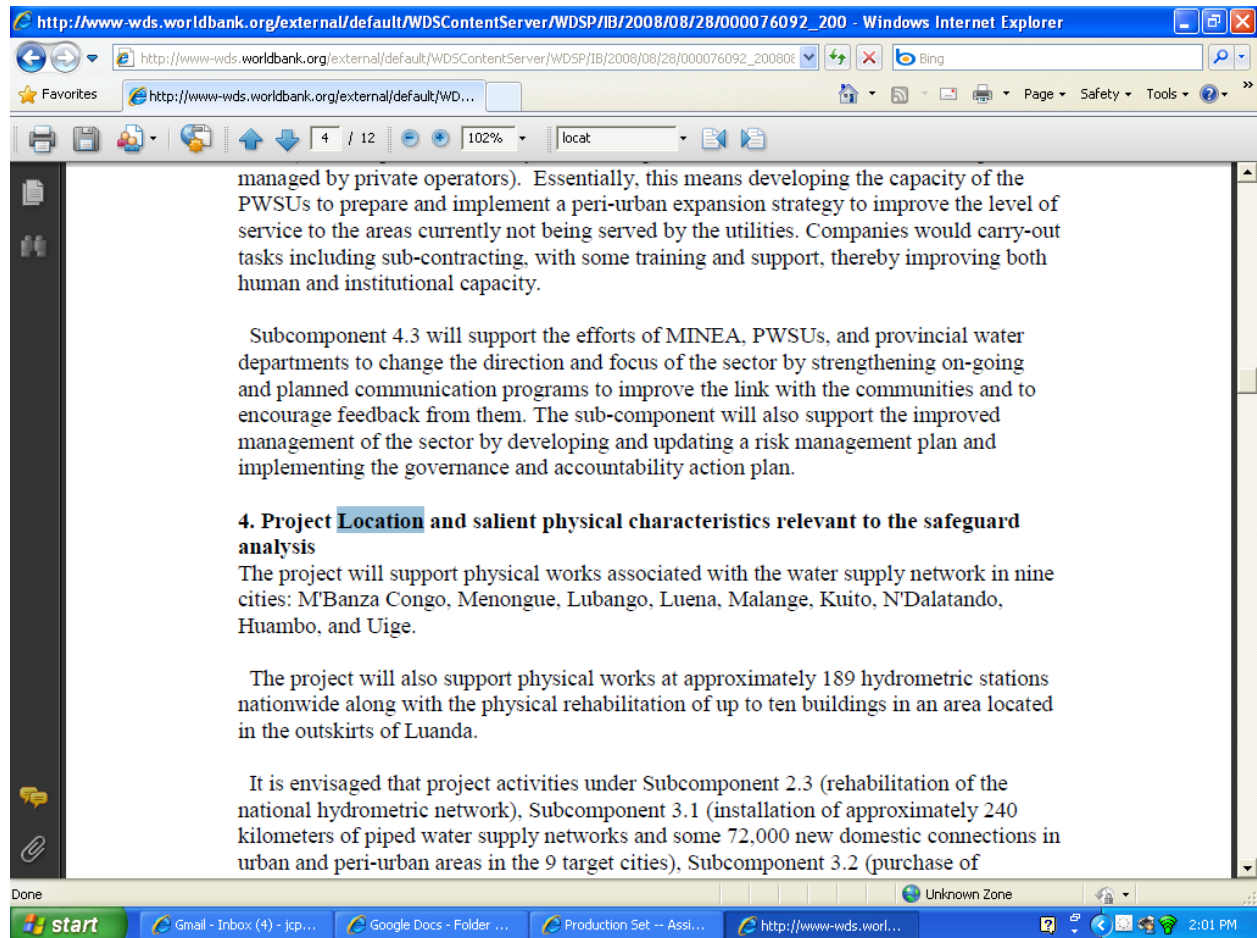
Searching "locat" rather than "location" allows the search to find "location" "located" "locate" etc. The search function, however, should not be the only method used to review project documents. Documents should also be searched for tables, maps, or any other sources of geographic information.

Step 6: Open Coding Helper and Search for Desired Location
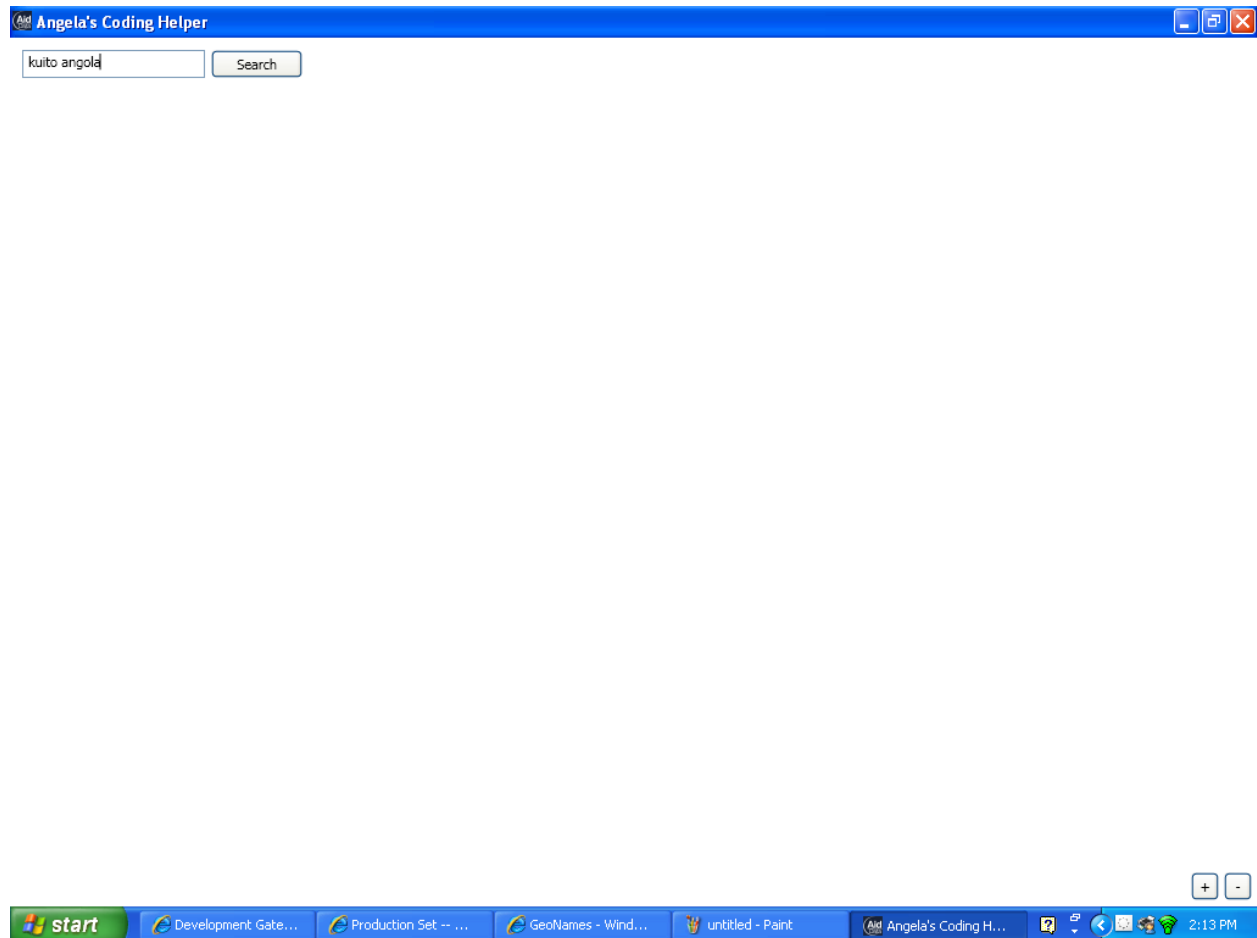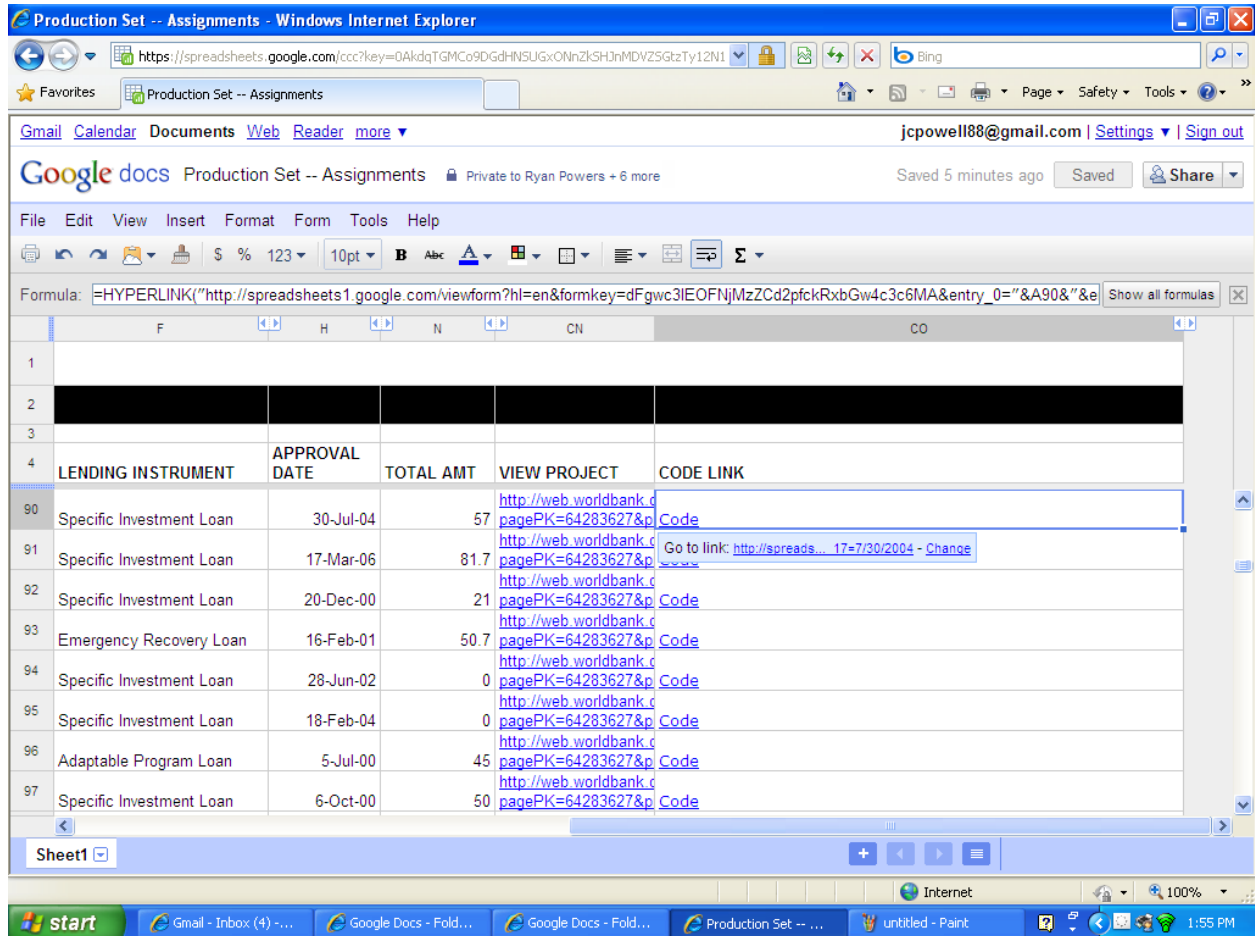
In this case, the location is Kuito, Angola.

Step 7: Select Desired Location from Search List
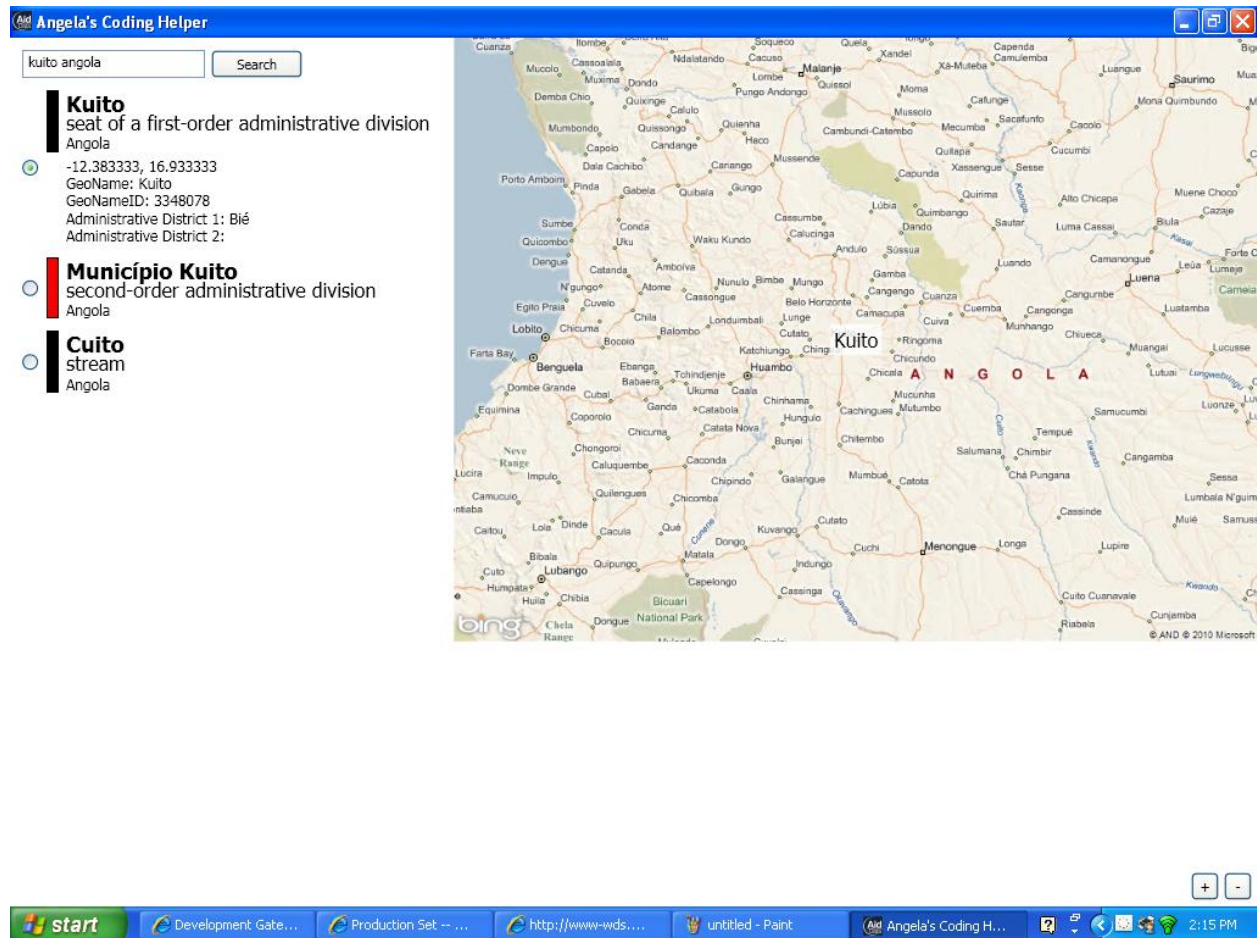
Having reviewed the documentation, the "seat of a first-order administrative division" (provincial capital) is the most likely location. Note: If a location found in project documentation is not found in geonames. This location will need to be identified using other sources and added to geonames. (see Instructions for Updating Geonames)

Step 8: Open "Code Link" in Assignments Sheet

The "code link" will access the coding interface, which is used to enter the geographic information for the project. Multiple codes can be entered for the same project by selecting the "code link" again after submitting the first location.

Step 9: Place Cursor in "Latitude" Field and Press CTRL+ALT+V

Pressing Ctrl+Alt+V when proper location is selected in coding helper populates all location fields automatically eliminating the need to copy and paste each geographic field individually.

Step 10: Insert Source Information and Source Link

In this case, the source is the Project Appraisal Document (PAD). The link to the project document is copied and pasted from the internet browser. The notes section is used to clarify coding decisions (such as why a city was selected for coding rather than a province of the same name).

Step 11: Open Project Results Sheet and Review Location Information to Ensure Accuracy

Inexperienced coders will not be able to edit incorrect information and must email their supervisor with inaccuracies, while experienced coders will be granted editing capabilities within the results spreadsheet.

Appendix B: The Arbitration Process

Step 1: Open Results Sheet for Coder 1

This coding results sheet contains all of the geo-codes entered by the specified coder.

Step 2: Download Results Spreadsheet in Excel Format

Step 3: Open Excel Sheet of Coder's Results and Filter to Desired Country

The filter function can be found by selecting the "data" tab in excel, selecting the arrow in the "Country" column, then selecting the desired country.

Step 4: Copy Country Rows and Paste into New Excel Spreadsheet

After creating new file, save as form Coder 1_Angola.

Step 5: Upload Country File to Arbitration System
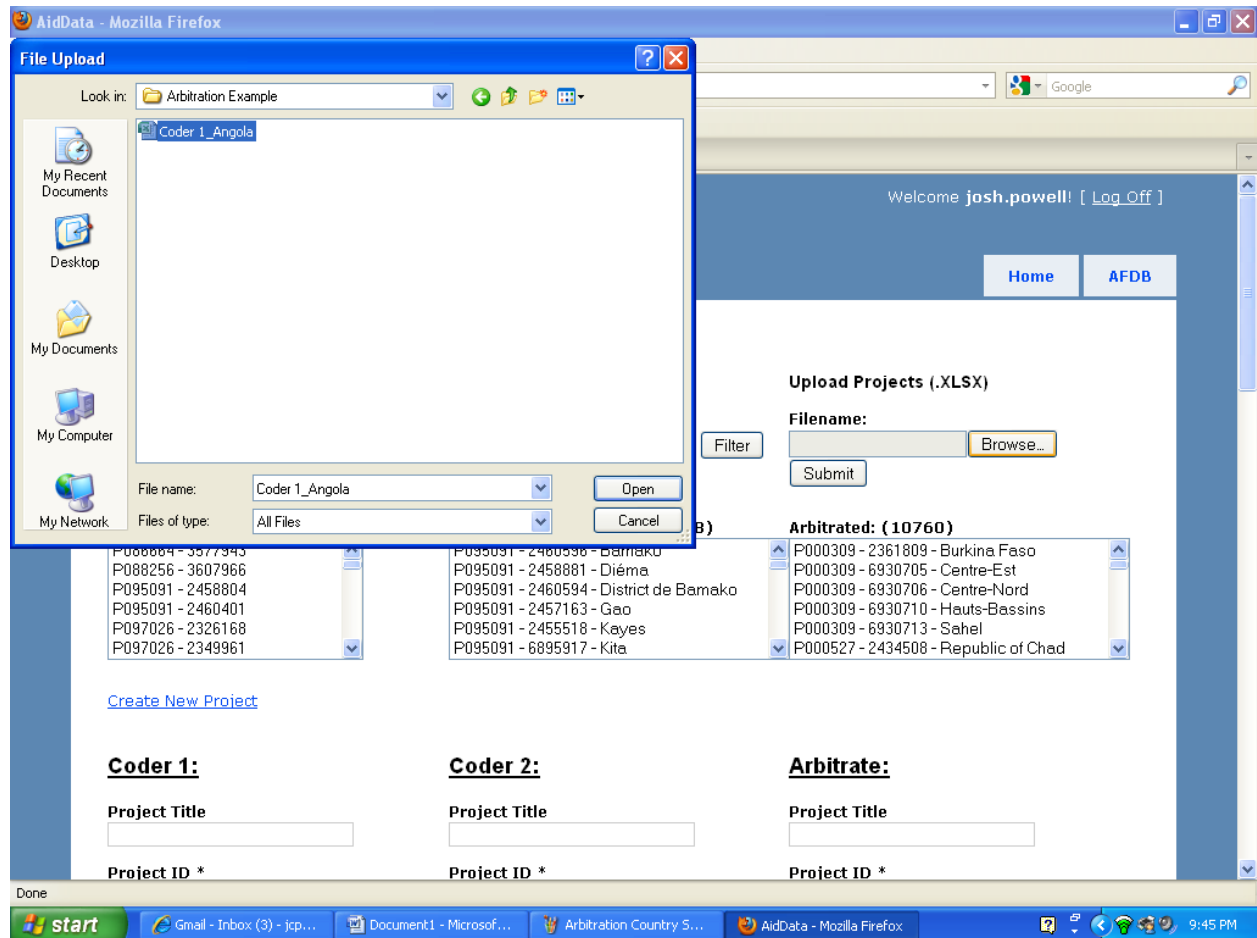
Step 6: Repeat Steps 1-5 for Coder 2 on Same Country

Step 7: Filter to Desired Country for Arbitration

Filtering by country will restrict viewed projects to desired country.

Step 8: Filter to Desired Project ID

Arbitration should be performed one project at a time for efficiency and accuracy.

Step 9: Address Matches with Errors

"Matches with errors" are entries where both coders selected the same geoname and geoname ID, but other differences exist. Differences between coding entries are highlighted in red and must be resolved before the entry can be submitted to the final database. The notes section from both coders should provide sufficient information to determine the correct substance. However, should the notes field be blank or insufficient, arbiters can determine correct entries by opening the "Source URL" link and reviewing project documents for clarification.

Step 10: Resolve Differences and Submit Entry

Submitting the resolved information will move the geographic entry into the "arbitrated" column.

**Step 11: Address Projects without Matches and Verify Locations**

Projects without matches occur when one coder finds a location that the other coder does not. The coded location can be verified by opening the "Source URL" link and searching for the location name.

Step 12: After Verifying Location, Submit Project without Match

Submitting the verified location will move entry into the "arbitrated" column.

Step 13: After Arbitrating Country, Select "Export Arbitrated Records"
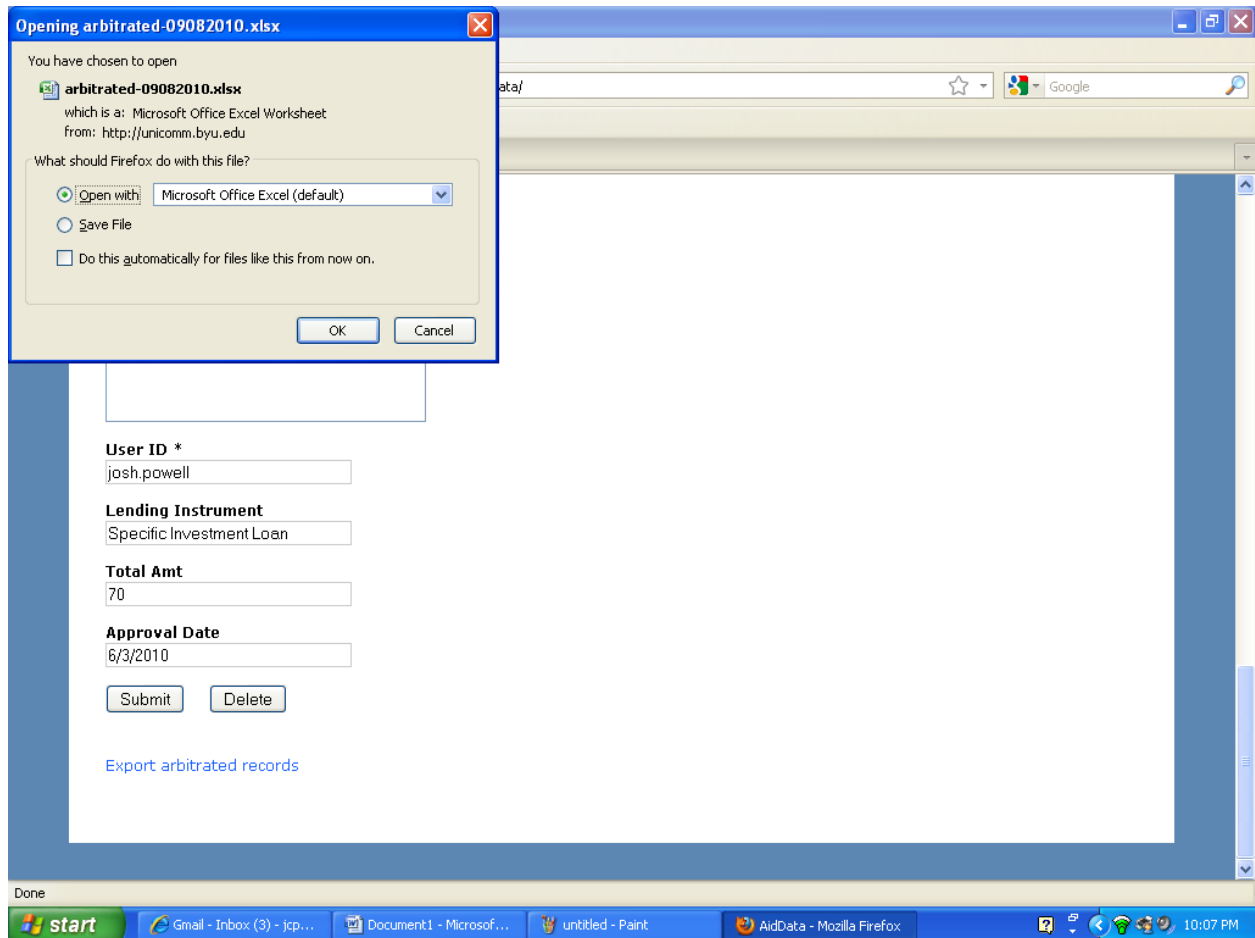
This will export all arbitrated records, separated into individual sheets by country.