A methodology for merging development assistance data from IATI and OECD-DAC CRS using SQLite

Contents

Introduction

Abstract

This is a methodology to build a database containing IATI and OECD-DAC CRS development assistance data using SQLite. The methodology further outlines how to utilise existing IATI and CRS policy markers combined with keyword searches in project descriptions to build sub-sets of projects targeting certain thematic areas. While the methodology is often described in the context of our Women's Economic Empowerment (WEE) project, the methodology can be replicated and adjusted to target different areas of related research.

Our project

Our project mapped national and international funding to support women's economic empowerment (WEE), women's financial inclusion (WFI), and women's empowerment collectives (WECs) in three focus countries (Bangladesh, Kenya and Nigeria). Our methodology for identifying WEE-relevant projects utilised the two largest development assistance data sources: The International Aid Transparency Initiative (IATI) and the Organisation for Economic Co-operation and Development Development Assistance Committee (OECD-DAC) Creditor Reporting System (CRS). We identified potentially relevant projects by filtering for funders' OECD-DAC gender markers scores and by applying key-word searches on project titles & descriptions. After this, our team manually reviewed the projects to verify their relevance for women's economic empowerment, women's financial inclusion and women's empowerment collectives.

The data

In this section we will outline the two data sources that we will be merging in this methodology. We will describe the various differences between the two sources along with any data considerations that need to be made when using them. These topics will also be explored in more detail throughout the methodology.

Data sources

International Aid Transparency Initiative (IATI)

The International Aid Transparency Initiative (IATI) is a global standard for the transparent publication of records of how aid money is spent. Currently IATI has around 1,200 publishers. The term 'aid' is used in IATI data to capture all humanitarian and development assistance and will allow a broader scope of search than the ODA definition adopted by the OECD. Unlike the CRS database, IATI data does not have a publishing delay, this allows a more current analysis on funding trends. The IATI standard is comprised of elements, each with a clear definition for publishers to follow.

However, it has not been verified by an official source because publication is voluntary. Some publishers choose to aggregate their publications at country and aid type levels as well as project level in some cases. In addition, funders sometimes aggregate all of their commitments/disbursements from a single month/quarter under one publication. For this reason, monthly analysis of trends can be inaccurate unless further care is taken to remove such cases. Conducting analysis at an annual level will also make it possible to merge IATI data together with other sources, such as CRS, which also do not provide timely disaggregation.

OECD-DAC CRS (CRS)

The Organisation for Economic Co-operation and Development – Development Assistance Committee (OECD-DAC) Credit Reporting System (CRS) is the standard for Official Development Assistance (ODA) reporting globally. The definition of ODA has been carefully developed over time and has its own strengths and limitations. An important limitation to consider is that not all countries and organisations report to the CRS, including prominent funders such as China, Brazil and India, and many private finance institutions. Funding flows data was reported to the CRS by 121 providers in 2020, including official providers, private philanthropic providers and multilateral organisations. CRS data publication is always delayed by 12/18 months to allow time for it to be verified by government sources so it has strong reliability. Whilst CRS data is increasingly disaggregated down to project level many publishers still only publish sector level aggregations. The CRS has also pioneered the use of various policy markers such as the gender policy marker allowing some insights into where funders are directing funding towards WEE/WFI/WEC and gender integration (GI).

Other data considerations

Timeliness

Due to the differences in the timeliness of data between the data sources which also intersects with the different aggregation levels, we have to apply the lowest common denominator which in this case is annual, so data will be analysed by annual spending trends. Furthermore, since the publication rates of funders in IATI varies, we have to be careful when analysing recent data, as taking recent funding amounts at face value would likely underestimate the total amount of funding once the funders who only publish quarterly/yearly have been included.

Frequency of use of variables

Another area which can complicate analysis while using merged IATI and CRS data is in the frequency of use of certain variables across the two data sources. For example, in 2020, only 40% of projects in the IATI standard included an OECD gender marker, compared to 72% in CRS. This creates difficulties when conducting analysis on these variables from data from both sources, as it can be difficult to ensure your analytical results are indeed due to the nature of the projects, rather than reporting differences between data sources. We will explore this topic further later in the methodology.

Overview of methodology for merging IATI and CRS data sources

In this section, we will outline the various steps involved in merging the two data sources along with brief descriptions of what these steps involve. Each of these steps will be expanded upon in the detailed methodology.

1. Extracting IATI/CRS data

The first step in this methodology will outline how to download the raw data for both the IATI and CRS databases. We will outline how to download both datasets separately and also outline any additional software that will be required during the data merging process. Data has to be downloaded at a country level in IATI so if your project looks at global funding then this stage might be more arduous than if you are analysing country trends. It should also be noted that a global database would be extremely large and you would likely need significant computing power to perform analysis on it, potentially even requiring external computing power such as cloud computing.

2. Importing missing data

The next step in the methodology will outline how any additional IATI data (such as Policy Markers and Project Descriptions) can be downloaded and added into the current database.

3. Combining data sources

Once the data has been imported and the additional IATI data has been added, the data needs to be combined so that we have one merged database containing data from both data sources.

4. Data splitting

Both the IATI and CRS databases include a wide range of development assistance data. It is therefore necessary for us to apply some filtering steps to our merged dataset so that we can analyse the appropriate activities that are relevant for our project. The variables that we will filter by will be: 'Finance Type', 'Transaction Type' & 'Flow Type'.

5. Duplicate removal

Some publishers report activities to both the IATI and CRS databases. We therefore need to identify where projects have been reported twice and remove one of them. Our approach for this was to identify publishers that reported to both IATI and CRS and remove all the publisher's activities from one of the data sources depending on which data source had the best data quality.

6. Double counts

Especially when using IATI data, we need to be aware of two organisations reporting the same funding flow to the same data source. This would result in the same project being accounted for twice in the data, therefore we have to take steps to identify such projects and remove them.

7. Data cleaning

Funders are often inconsistent in the way they publish their activities in each of the data sources, and sometimes even in the same source. This is most apparent in the 'Reporting Organisation' and 'Provider Organisation' columns where funders are often very inconsistent in their publishing. We will outline the process of removing these inconsistencies in detail for many different columns in the data.

Methodology

Overview

In this section we will lay out our detailed methodology for each of the steps outlined in the overview above. We will include any SQL syntax that we used where appropriate and link example files for some of the data cleaning steps.

Variable names

In this methodology, the displayed code will include generalised variable names such as *country_one_IATI_data*. These variable names will have to be updated at every stage of the methodology to match your projects variables, for example, *Pakistan_IATI_data*.

File names

The combined IATI and CRS database will be named as *combined* at every stage in this methodology, however, when implementing this code, you should keep a form of version control as you move through the steps. For example, *combined_v1* then *combined_v2* and so on... The code will need to be updated at each stage to reflect this.

Software

Microsoft Excel

The first piece of software that you will need for merging IATI and CRS data will be Microsoft Excel. A basic understanding of MS Excel and it's functions should be sufficient for the uses outlined in this methodology.

SQLite Browser

Most of the merging process will be implemented using a relational database and the programming language, SQLite. To run queries in SQLite, we must first download an SQLite browser. This can be downloaded here.

1. Extracting IATI/CRS Data

First, we will outline the steps for downloading the raw IATI and CRS data separately, including any data considerations that need to be considered.

Extracting IATI data

There are a few different ways to download IATI data, however the one that we have found to be easiest is using the IATI Country Development Finance Data Tool. Unfortunately, this does not allow you to download data for multiple countries at once, so if your project analyses global funding, then you will need to go through and download each country individually. This would be a very lengthy process when taking in to account the importing of additional data and other data merging steps so we would only recommend this for a project with a large amount of available resources. The IATI country data can be downloaded from the Country Finance Data tool website. This will allow you to download a Microsoft Excel file of the country data which can be saved as a CSV file to allow it to be later opened in an SQLite database.

Code for merging IATI country datasets

To make the process of adding extra variables into the data more streamlined, we will merge our IATI country datasets into one using the code below. This step can be ignored if your project only looks at a single country. This outlines an example of merging three countries IATI data, however, it can easily be adapted to the number of countries in your project.

```
1
2  CREATE TABLE combined_IATI AS
3      SELECT * FROM country_one_iati_data UNION ALL
4      SELECT * FROM country_two_iati_data UNION ALL
5      SELECT * FROM country_three_iati_data
6
```

Extracting CRS data

Unlike IATI data, CRS data can be downloaded as a bulk download for every country at once. However, they do disaggregate their data by year so you will need to separately download data for each if your project focuses on multi-year funding flows. To download CRS yearly data, you can use the OECD Bulk Download. This will allow you to download a zipped 'txt' file which can be opened in excel and converted to a CSV to allow it to be later opened in an SQLite database. Alternatively, a txt file can be opened directly into the SQLite browser, however, we found this to be more time consuming than using excel.

Code for merging CRS annual datasets

If your project will look at funding data from multiple years, then it is now necessary to merge each of our CRS annual datasets into one merged dataset. We will also use this as an opportunity to filter by the countries that we wish to keep in our data. The code below is an example of how you will merge the CRS datasets. This example takes CRS datasets from the years (2015, 2016, 2017, 2018 & 2019) and merges them into one and also filters by the three countries (country_one, country_two & country_three).

```
1
2  CREATE TABLE combined_CRS AS
3  SELECT * FROM combined_CRS_2015 UNION ALL
4  WHERE "Recipient Country" in (country_one, country_two, country_three)
5  SELECT * FROM combined_CRS_2016 UNION ALL
6  WHERE "Recipient Country" in (country_one, country_two, country_three)
7  SELECT * FROM combined_CRS_2017 UNION ALL
8  WHERE "Recipient Country" in (country_one, country_two, country_three)
9  SELECT * FROM combined_CRS_2018 UNION ALL
10 WHERE "Recipient Country" in (country_one, country_two, country_three)
```

```
SELECT * FROM combined_CRS_2019
WHERE "Recipient Country" in (country_one, country_two, country_three);


```

Column headings

Table 1 shows the full list of CRS and IATI variables that we will include in the final merged database along with brief descriptions of each. This is not the full list of CRS and IATI variables that exist, however, since the two data sources are being merged, we decided to only include variables that had a direct equivalent across both data sources and that were relevant to the scope of our project. The IATI 'gender' and 'description' columns marked in green are not included in the download described above and hence will have to be imported separately, the methodology of which will be outlined in the following section. Although it can be done at a later stage, we recommend deleting any extra columns that you will not use from your datasets now as it will make them cleaner and reduce their file sizes.

IATI column name	CRS column name	Description
Reporting organisation	Donor name	Name of the organisation that reports the project to the database.
Provider organisation	Agency name	Name of the provider organisation/sub-agency that is implementing the project.
Title	Project title	Title of the Project
Reporting organisation type	-	
Aid Type	Aid_t	Aid type (e.g. budget scholarship, policy reform, core contributions etc.)
Finance Type	Finance_t	Finance Type (e.g. Loan, grant, equity, cash etc.)
Receiver organisation	ChannelReportedName	
Receiver organisation type	ChannelName	
Transaction Type	-	Transaction type (e.g. Commitment, Disbursement)
Recipient Country	Recipient Name	
Sector category (codes separated, 3-digit)	Sector codes (3-digit)	
Sector category (names separated, 3-digit)	Sector name (3-digit)	
Sector (code separated, 5-digit)	Purpose code (5-digit)	
Sector (name separated, 5-digit)	Purpose name (5-digit)	
Calendar Year	Year	
Description	LongDescription	
-	USD_Disbursement	
-	RegionName	
-	USD_Commitment	
IATI Identifier	CRSID	A unique ID for each project (Each project can have multiple activities)
Gender Marker	Gender	A marker for gender targeted projects
Value USD	-	A mixed value variable with disbursements and commitments.

Table 1: Column headers for IATI and CRS data

2. Importing missing data

After you have imported the core IATI and CRS data as outlined in the previous step, you can then download any additional columns that you would like to add to your data. This step is not necessary for CRS since the download described above includes all the variables that the OECD collects. In this section we will go through some of the IATI columns that can be added to the data along with the detailed methodology for joining them into the main IATI dataset.

Project descriptions

The project description column is an incredibly valuable variable to include in your dataset as it can give a detailed insight into the scope of each project. It can also be used to conduct keyword searches to identify projects relevant to your project's research. The quality of the descriptions provided often varies from organisation to organisation, however, as shown in Figure 1, the average length of description that organisations are publishing has increased over the last 5 years.

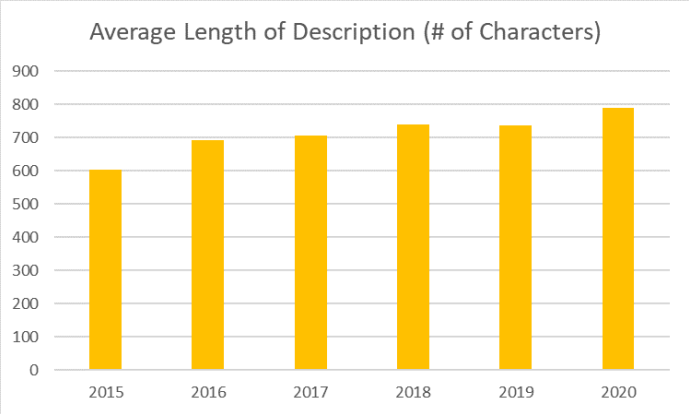


Figure 1: Average description length in IATI data by year

Adding project descriptions into the dataset

To import project descriptions, we will utilise IATI's query builder tool, which enables you to download activity level data for projects matching a chosen filter. Filtering by 'Recipient Country Code' will allow you to return all activities which list your target country as a recipient country. If you are analysing multiple countries, you will have to repeat this step for each. Once you have applied the filter, you will be able to download the data as a CSV file. When given the three options for the download, you should select the Activity download. To add the description column into the main dataset, you first need to create a new CSV file including only the *Description_Narrative* and *IATI_identifier* columns from the main CSV file. Figure 2 shows an example of how this CSV should be formatted. Then, repeat this step for the other countries so that you have a CSV file for each. The next step will be to merge the individual country CSV files into one. You will then be able to import this CSV file into the database browser and implement a join to merge them into the main dataset using the code shown below.

iati_identifier	description_narrative
41119-SN-O1-RT	UNFPA Senegal Improved programming for results activities,UNFPA Senegal Improved programming for results activities
41119-SN-S1-RT	UNFPA Senegal Policy and accountability activities,UNFPA Senegal Policy and accountability activities
41119-SN-S2-RT	UNFPA Senegal Quality of care and services activities,UNFPA Senegal Quality of care and services activities
41119-SN-S3-RT	UNFPA Senegal Gender and social norms activities,UNFPA Senegal Gender and social norms activities
41119-SN-S4-RT	UNFPA Senegal Population change and data activities,UNFPA Senegal Population change and data activities
41119-SN-S5-RT	UNFPA Senegal Humanitarian action activities,UNFPA Senegal Humanitarian action activities
41119-SN-S6-RT	UNFPA Senegal Adolescents and youth activities,UNFPA Senegal Adolescents and youth activities

Figure 2: Example CSV with description narrative column

Code for combining country description CSV files

This code will combine the individual country CSV files which include the description narrative into one merged table which we will be able to join back into the main IATI data.

```
1
2 CREATE TABLE combined_descriptions AS
3     SELECT * FROM country_one.descriptions UNION ALL
4     SELECT * FROM country_two.descriptions UNION ALL
5     SELECT * FROM country_three.descriptions
6
```

Code for joining descriptions into main IATI data

This code will join the descriptions from the merged description table (*combined_descriptions*) back into the main IATI data.

```
1
2 CREATE TABLE combined_IATI_V2 as
3     SELECT * from combined_IATI
4     LEFT JOIN combined_descriptions on combined_IATI.IATI_identifier =
  combined_descriptions.IATI_identifier
5
```

Policy markers

There is also a 'Policy Marker' variable available for publishers to use in IATI. The policy marker utilises the policy markers developed by the OECD. The policy markers each occupy a separate column in the CRS data, whereas in the IATI database they are represented in a single column with a corresponding codelist. This is shown in table 2.

Figure 3 shows the frequency of use of the gender marker in the IATI and CRS databases. There are a number of major donors that use the gender marker in their CRS publications but not in their IATI publications, for example United States Agency for International Development (USAID). This trend is consistent throughout the different policy markers and is something that should be considered when analysing IATI data compared with with CRS data.

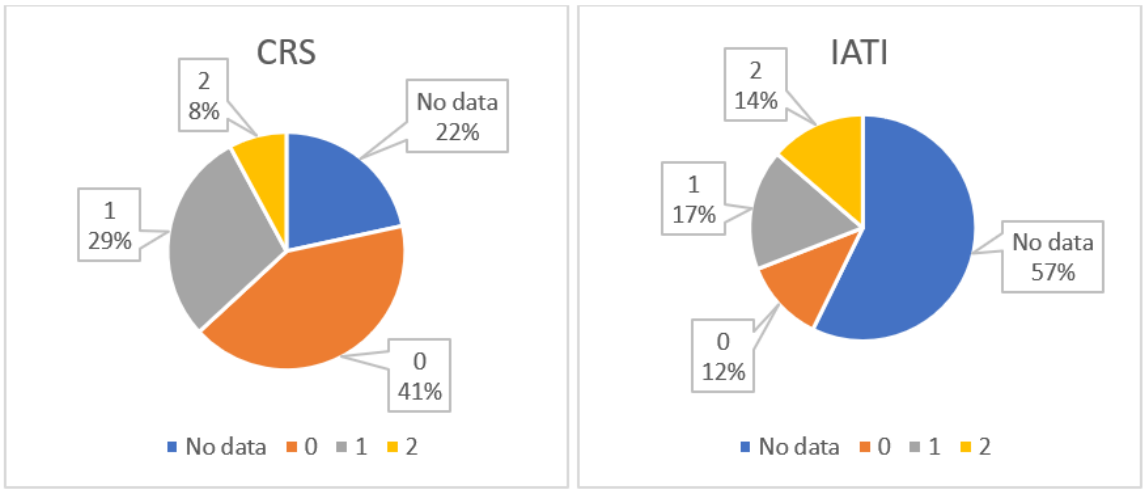


Figure 3: Use of the Gender marker across IATI and CRS

CRS Column Heading	IATI Policy Marker (Name)	IATI Policy Marker (Code)
Gender	Gender Equality	1
Environment	Aid to Environment	2
PDGG	Participatory Development/Good Governance	3
Trade	Trade Development	4
Biodiversity	Aid Targeting the Objectives of the Convention on Biological Diversity	5
ClimateMitigation	Aid Targeting the Objectives of the Framework Convention on Climate Change - Mitigation	6
ClimateAdaptation	Aid Targeting the Objectives of the Framework Convention on Climate Change - Adaptation	7
Desertification	Aid Targeting the Objectives of the Convention to Combat Desertification	8
RMNCH	Reproduction, Maternal, Newborn and Child Health (RMNCH)	9
DRR	Disaster Risk Reduction (DRR)	10
Disability	Disability	11
Nutrition	Nutrition	12

Table 2: Policy marker codelist

Policy significance

In addition to the 'Policy Marker' column, there is a 'Policy Significance' column which publishers use to state to what extent the project targets the policy area described in the 'Policy Marker' column. This codelist is the same as used in the CRS data and is displayed in table 3.

Value	Description
2	principal objective
1	significant objective
0	not targeted
(blank)	not screened

Table 3: Policy significance values

There is also a unique policy significance marker 3 in IATI which represents 'principal objective AND in support of an action programme', however, this is for desertification-related aid only and as it had no direct comparison in CRS, so we chose not to include this in our data.

Adding policy markers into the dataset

To add a policy marker column into the main dataset, you should utilise the data download function of d-portal. The filtering in d-portal will allow you to filter by a given policy marker, policy significance pairing. In this example we will outline how to add the Aid to Environment policy marker, however, the process is exactly the same for each of the policy markers and indeed could be repeated multiple times if you wish to add multiple policy markers into your main data.

To create a list of projects that include the policy marker that you are interested in, first filter in d-portal by all the relevant recipient countries, then filter by *policy marker = Aid to Environment (not targeted) (0_2)* to return a list of projects that are marked as not targeting the environment. Then you are able to download a CSV including project data for these projects. This can then

be repeated for the other significance values: *Aid to Environment (significant objective) (1_2)* and *Aid to Environment (principal objective) (2_2)*. Once this is completed, you should have 3 CSV files which include project data for each of the policy marker/significance pairings. For each of the CSV files, you should then remove all of the columns except *IATI_identifier*, and add an additional column name *Aid_to_Environment* in which you should add the policy marker significance value. For our project, we used the numerical coding for this, i.e. 0,1,2 as this is the same format that is used in the CRS data, however, if you would rather use the written form i.e. not targeted, significant objective, principle objective you would just need to change the CRS coding to this too.

The next step is to append the 3 individual CSV files into one country table, then repeat for all the countries. We can then append the country tables into one combined table. We are then able to join this table back into the main data using the SQLite code below.

Code to create country CSV files with all three siginifcance value (0,1,2)

This code will take the three policy marker significance value CSV files and merge them into one country table which includes all the policy marker significance values. If your project looks at multiple countries, you will need to apply this code for each country, adjusting the code accordingly for each.

```
1
2  CREATE TABLE combined_policy_marker_country_one AS
3      SELECT * FROM country_one_policy_marker_1 UNION ALL
4      SELECT * FROM country_one_policy_marker_2 UNION ALL
5      SELECT * FROM country_one_policy_marker_3
6
```

Code to combine country policy marker tables

This code will take the individual country tables which include all the policy marker significance values and merge them into a combined table which includes all the countries. This example merges tables from 3 countries, however, it could easily be modified to include more/less countries.

```
1
2  CREATE TABLE combined_policy_marker AS
3      SELECT * FROM combined_policy_marker_country_one UNION ALL
4      SELECT * FROM combined_policy_marker_country_two UNION ALL
5      SELECT * FROM combined_policy_marker_country_three
6
```

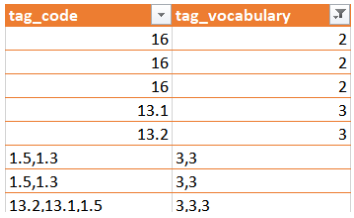
Code to join policy markers into main IATI data

This code will join the policy markers significance values from the table (*combined_policy_marker*) back into the main IATI data.

```
1
2  CREATE TABLE combined_IATI_v2 AS
3      SELECT combined_IATI
4      LEFT JOIN combined_policy_marker ON combined_IATI_v2.IATI_identifier =
   combined_policy_marker.IATI_identifier
5
```

SDG goals/targets

In addition to the policy markers, reporting organisations are also able to mark their activities in IATI against the 17 UN Sustainable Development Goals (SDGs) and within these, the 169 SDG targets. This is done through the *Tag_Code* and *Tag_Vocabulary* columns. The Tag Vocabulary column specifies which codelist the Tag Code has come from, 2 = SDG goals and 3 = SDG targets. The codelists can be found for UN Sustainable Development Goals (SDG) Goals & UN Sustainable Development Goals (SDG) Targets on the IATI website. The SDG targets are formatted such that they include the SDG goal too, e.g. SDG target 1.5 is a subset of SDG goal 1. An example of how this appears in the data is shown in figure 4. Publishers can also specify multiple SDG goals/targets for a given activity, like in the last three rows of figure 4.



tag_code	tag_vocabulary
16	2
16	2
16	2
13.1	3
13.2	3
1.5,1.3	3,3
1.5,1.3	3,3
13.2,13.1,1.5	3,3,3

Figure 4: Data import of SDGs

The equivalent column in the CRS data is *SDG-focus* which follows the same format as the tag_code column in IATI. However, consistent with what we observed with the policy markers, the frequency of use in the CRS data of the SDG marker is much higher; in Pakistan an SDG target was specified in 55% of CRS activities compared with only 7% in IATI.

Downloading SDG goals/targets

To download the SDG goals and targets we will use same data that we downloaded using the database query for the project descriptions. As with the project descriptions, it will be necessary to do this for each country separately. Next, it will be necessary to filter by tag_vocabulary = 2,3 or any combination thereof (e.g. 3,3,3 or 2,2 etc.). You should then create a new CSV with only the *tag_code* and *IATI_identifier* columns. Once this has been completed for each of the countries, the next step will be to append the individual country CSV files into one combined table. We will then be able to join the SDG goals back into the main IATI dataset.

Code to combined SDGs country CSV files into one merged table

This code will take the individual CSV files that contain the SDG *tag_code* values for each country and combined them into one table for every country.

```
1
2  CREATE TABLE combined_SDG AS
3      SELECT * FROM SDG_country_one UNION ALL
4      SELECT * FROM SDG_country_two UNION ALL
5      SELECT * FROM SDG_country_three
6
7
```

Code to join SDG values into main IATI data

This code will join the SDG *tag_code* values from the combined country table back into the main IATI data.

```
1
2  CREATE TABLE combined_IATI_v2 AS
3      SELECT combined_IATI
4      LEFT JOIN combined_SDG ON combined_IATI.IATI_identifier = combined_SDG.
   IATI_identifier
```

```
5
6
```

SDG disaggregation

As briefly touched upon earlier, reporting organisations are able to specify multiple SDG goals/targets for a given activity. These are represented in the data as a comma separated value in a single cell. This unfortunately makes filtering by the SDG goal/target more difficult because two activities which both target the same SDG can have different values (e.g. 1.5, 1.3 and 1.5, 6.2). In addition to this, it is not posible to publish both SDG goals and SDG targets for a single activity. This means that in any instance that an activity has an SDG target, it would not be included in any filtering for its SDG goal. For example, an activity with *tag_code* = 1.5 (i.e. SDG target 1.5) would not be included in any filtering for *tag_code* = 1 (i.e. SDG goal = 1) despite SDG target 1.5 being a subset of SDG goal 1. There are a number of possible ways that the SDG goals/targets could be separated into separate columns, for example, using comma separation, however, in our project we did not explore this avenue as we were not interested in filtering by SDGs.

3. Combining data sources

Once we have imported all the additional datapoints that we require, we must combine our CRS and IATI datasets so that we have a final appended dataset.

Appending IATI and CRS data

When merging our CRS and IATI data, it is necessary that we chose a column heading name that will be used for each new merged column. For example, the *Reporting Organisation* column in IATI and *Donor Name* column in IATI will be merged and named *Reporting Organisation*. The choices that we made are represented by the bold column headings in table 1.

Sector and purpose name columns

In order to add new *Purpose Name* and *Sector Name* columns which will include the text description for the *Purpose Code* and *Sector Code* columns, we must first create tables which include the numerical value and the text description for each sector/purpose code. An example for the sector name table is shown in table 4. These tables can be created using the OECD list of sector codes and purpose codes which can be found here. In the code listed below the tables are referenced as *IATI_sector* for the sector code table, and *IATI_sector_category* for the purpose code table.

Code	Name
111	Education, Level Unspecified
112	Basic Education
113	Secondary Education
114	Post-Secondary Education
121	Health, General
122	Basic Health
123	Non-communicable diseases (NCDs)
130	Population Policies/Programmes & Reproductive Health
140	Water Supply & Sanitation

Table 4: Sector code & Sector name pairings

Code for merging IATI and CRS datasets

This code will merge the IATI and CRS datasets that we have created in the previous section.

```
1
2
3  CREATE TABLE combined  AS SELECT
4       'IATI' AS [Data Source],
5
6       "Reporting Organisation" AS "Reporting Organisation",
7       "Provider Organisation" AS "Provider Organisation",
8       "Title" AS "Title",
9       "Reporting Organisation Type" AS "Reporting Organisation Type",
10      "Aid Type" AS "Aid Type",
11      "Finance Type" AS "Finance Type",
12      "Receiver Organisation" AS "Receiver Organisation",
13      "Receiver Organisation Type" AS "Receiver Organisation Type",
14      "Transaction Type" AS "Transaction Type",
15      trim(substr("Recipient Country or Region", instr("Recipient Country or Region
    " || '-', '-') + 2)) AS "Recipient Country",
16      "sector category" AS "Sector code (3-digit)",
```

```
    COALESCE(iati_sector_category.name, "sector_category_name") AS "Sector name (3-digit)",
    "Sector" AS "Purpose code (5-digit)",
    COALESCE(iati_sector.name, "sector_code_name") AS "Purpose name (5-digit)",
    "Calendar Year" AS "Year",
    "Description" AS "Description",
    NULL AS "USD_Disbursement",
    NULL AS "RegionName",
    NULL AS "USD_Commitment",
    "iati_identifier" AS "Unique ID",
    "Policy Marker" AS "Policy Marker",
    "Value (USD)" AS "Value (USD)",
    "flow type" AS "Flow Type",
    NULL as "Currency",
    "provider organisation type" AS "Provider Organisation Type",
    "multi country" AS "Multi Country",
    "humanitarian" AS "Humanitarian",
    "Calendar Quarter" AS "Calendar Quarter",
    "Calendar Year and Quarter" AS "Calendar Year and Quarter",
    "URL" AS "URL",
    "Value (EUR)" AS "Value (EUR)",
    "Value (Local currrency)" AS "Value (Local currrency)",
FROM combined_IATI
LEFT JOIN iati_sector ON iati_sector.code = sector_code
LEFT JOIN iati_sector_category ON iati_sector_category.code = sector_category
UNION ALL
SELECT
    'CRS' AS `Data Source`,

    "DonorName" AS "Reporting Organisation",
    "AgencyName" AS "Provider Organisation",
    "ProjectTitle" AS "Title",
    NULL AS "Reporting Organisation Type",
    "Aid_t" AS "Aid Type",
    "Finance_t" AS "Finance Type",
    "ChannelReportedName" AS "Receiver Organisation",
    "ChannelName" AS "Receiver Organisation Type",
    '' AS "Transaction Type",
    "RecipientName" AS "Recipient Country",
    "SectorCode" AS "Sector code (3-digit)",
    COALESCE(iati_sector_category.name, "SectorName") AS "Sector name (3-digit)",
    "PurposeCode" AS "Purpose code (5-digit)",
    COALESCE(iati_sector.name, "PurposeName") AS "Purpose name (5-digit)",
    "Year" AS "Year",
    ShortDescription || CASE WHEN substr(ShortDescription, -1, 1) = '.' THEN '. ' ELSE ' ' END || LongDescription AS Description,
    NULL AS "Target Groups",
    "USD_Commitment" * 1000000 AS "USD_Disbursement",
    "RegionName" AS "RegionName",
    "USD_Commitment" * 1000000 AS "USD_Commitment",
    "CrsID" AS "Unique ID",
    "Policy Marker" AS "Policy Marker",
    NULL AS "Value (USD)",
    "FlowCode" || ' - ' || "FlowName"        AS "Flow Type",
    NULL as "Currency",
    NULL AS "Provider Organisation Type",
    NULL AS "Multi Country",
    NULL AS "Humanitarian",
    NULL AS "Calendar Quarter",
    NULL AS "Calendar Year and Quarter",
    NULL AS "URL",
    NULL AS "Activity Start Date",
    NULL AS "Value (EUR)",
    NULL AS "Value (Local currrency)"
```

```
79 FROM combined_CRS
80 LEFT JOIN iati_sector ON iati_sector.code = "PurposeCode"
81 LEFT JOIN iati_sector_category ON iati_sector_category.code = "SectorCode"
82
83
84
85
```

To add any additional columns from your IATI and CRS data into your merged data you will need to add in a new line in the format

```
1
2 (Column_heading_in_IATI) AS (New_column_heading) -- Add into IATI section
3 (Column_heading_in_CRS) AS (New_column_heading) -- Add into CRS section
4
5
```

4. Data splitting

The combined IATI and CRS database is an excellent resource to analyse total spending flows within each country, however, in order to aid the data analysis, some data will need to be split. The most important distinction that needs to be made when analysing development assistance data is that of grants and loans. Given the very different functions of grants and loans in aid financing, it is very unwise to collate the two when analysing funding trends. For our project, we also chose to filter by Flow Type and Transaction Type.

Finance type

For our project we were only interested in making the distinction between Grants and Loans, however, the finance type column allows publishers to specify more granular information on the type of financing within the definition of 'grants and loans'. These 'finance types' encompass a variety of different funding flows including concessional loans, standard grants, equity, guarantees and standard loans (see table 5 for the variations of codes used). To allow us to split up Grants and Loans, we must first group each of the finance type values into Grants or Loans. To do this we should create a table of distinct finance types and add the additional columns 'Finance Type Filter Grants Y/N' and 'Finance Type Filter Loans Y/N'. For each of the distinct finance types we then need to decide whether it falls under Grants or Loans populate the other columns accordingly. An example of what this table will look like is shown in Figure 5.

Finance Type	Finance Type Filter Grant Y/N	Finance Type Filter Loan Y/N
110	Y	N
110 - Standard grant	Y	N
1100	N	Y
1100 - Guarantees/insurance	N	Y
111 - Subsidies to national private ...	N	N

Figure 5: Finance type filtering

A significant number of activities did not include a 'finance type'. This appears in the data as 'no data'. Rather than remove all of these projects from our analysis, we decided to mark these in our filtering with a *'B'*. We would then only include these projects if the other filters (Flow type & Transaction type) returned a *'Y'*. We will go into this filtering in more detail later in this section

Code for joining *Finance Type* filters into main data

The code below will join in the *Finance Type Filter Grant Y/N* and *Finance Type Filter Loan Y/N* columns into the main data.

```
1
2  UPDATE TABLE "combined" SET "finance type" = trim("finance type");
3  UPDATE TABLE "Finance_type_filter" SET "finance type" = trim("finance type");
4  CREATE TABLE "combined_v2" AS
5  SELECT * FROM combined
6  LEFT JOIN "Finance_type_filter" ON combined."finance type" = "finance_type_filter
   "."finance type";
7
8
```

Commitment versus disbursements (Transaction Type)

Funding organisations publish both commitments data of committed spend and disbursements of actual spend in the IATI standard and the CRS database. These two funding flows require separate analysis, for this reason they should be split in the data sets and their funding should be analysed separately. Donors specify whether a transaction is a disbursement or commitment using the 'transaction type' column, therefore, we need to filter by 'transaction type' to select which we would like to include in our analysis. For our project, we chose to analyse commitments data for Non-grants and disbursements for Grants data as this maximised the amount of funding we would capture. However, it would be easy to modify the code to include different funding flows. Because we chose to analyse different transaction types for Grants and Non-grants, we needed to apply different filtering for each, much like we did for the 'finance type' filtering. Figure 6 shows an example table for the 'transaction type' filtering. Table 6 includes the full list of transaction types and our filtering for each.

Transaction Type	Transaction Type Filter Grant Y/N	Transaction Type Filter Loan Y/N
	Y	Y
1 - Incoming Funds	N	N
2 - Outgoing Commitment	N	Y
3 - Disbursement	Y	N
4 - Expenditure	Y	N
A	N	Y
P	Y	N
budget - Budget	N	N

Figure 6: Transaction type filtering

Code for joining *Transaction Type* filters into main data

The code below will join in the *Transaction Type Filter Grant Y/N* and *Transaction Type Filter Loan Y/N* columns into the main data.

```
1
2  UPDATE TABLE "combined" SET "transaction type" = trim("transaction type");
3  UPDATE TABLE "Transaction_type_filter" SET "transaction type" = trim("transaction
     type");
4  CREATE TABLE "combined_v2" AS
5  SELECT * FROM combined
6  LEFT JOIN "Transaction_type_filter" ON combined."transaction type" = "
   transaction_type_filter"."transaction type";
7
8
```

Flow Type

The 'flow type' column is used by funders to describe the type of resource flow using the OECD-DAC distinction between ODA (official development assistance) and other types of resource flow. Similar to the Finance/Transaction type filtering, we also need to filter by 'flow type' separately for grants and non-grants to keep the relevant resource flows for each. Figure 7 shows an example table of this filtering.
Similar to our finance type filtering, there are many flow types that we only chose to keep for grants or non-grants if they also had the relevant finance and flow types. These are also marked with a *'B'*.

Flow Type	Flow Type Filter Grant Y/N	Flow Type Filter Loan Y/N
	B	B
10 - ODA	B	B
11 - ODA Grants	Y	N
13 - ODA Loans	N	Y
14 - Other Official Flows (non Export ...	B	B
19 - Equity Investment	N	Y
20 - OOF	B	B
21 - Non-export credit OOF	B	Y
30 - Private Development Finance	B	Y
35 - Private Market	B	B
36 - Private Foreign Direct Investment	B	B
40 - Non flow	B	B
50 - Other flows	B	B
No data	B	B

Figure 7: Flow type filtering

Code for joining *Flow Type* filters into main data

The code below will join in the *Flow Type Filter Grant Y/N* and *Flow Type Filter Loan Y/N* columns into the main data.

```
1
2  UPDATE TABLE "combined" SET "flow type" = trim("flow type");
3  UPDATE TABLE "Flow_type_filter" SET "flow type" = trim("flow type");
4  CREATE TABLE "combined_v2" AS
5  SELECT * FROM combined
6  LEFT JOIN "Flow_type_filter" ON combined."flow type" = "flow_type_filter"."flow
   type";
7
8
```

Grants/Loans to Keep

Now that we have our three filters (*Finance type, Transaction type & Flow type*) that we wish to apply, it is necessary for us to create two columns that will consolidate these filters into one for grants and one for non-grants. We will label these two columns *Grants to Keep* and *Loans to Keep* and they will hold a binary value with 1 signifying a activity to keep and 0 signifying a activity for removal.

As briefly outlined before, we wish to keep activities that are marked 'Y' for all three of the filters. The exception to this will be when a activity is marked 'Y' for two and marked 'B' for the other. In this case we will also keep the activity.

Code for creating *Grants to Keep* and *Loans to Keep* columns

The code listed below will consolidate the three columns *Finance Type Filter Grant Y/N*, *Transaction Type Filter Grant Y/N* & *Flow Type Filter Grant Y/N* into a single column that will signify which projects should be kept (*Marked 1*) and considered as grants. The code should be applied to create a column *Loans to Keep* by changing the variables in the code.

```
1
2  ALTER TABLE combined ADD COLUMN "Grants to Keep";
```

```
3  UPDATE combined SET "Grants to Keep" = 0;
4  UPDATE combined SET "Grants to Keep" = 1
5  WHERE ("finance type filter grant Y/N" = "Y" and "flow type filter grant Y/N" = "
   Y" and "transaction type filter grant Y/N" = "Y") or ("finance type filter grant
   Y/N" = "B" and "flow type filter grant Y/N" = "Y" and "transaction type filter
   grant Y/N" = "Y") or ("finance type filter grant Y/N" = "Y" and "flow type filter
    grant Y/N" = "B" and "transaction type filter grant Y/N" = "Y");
6
7
```

5. Duplicate removal

Many funders publish their activities to both the IATI and CRS databases. To ensure that this doesn't lead to activities being included in the final data twice, we must make the decision to keep only one of the data sources for each funder. First, however, we must identify which funders are publishing to both databases. Large organisations often use the "Reporting Organisation" column to describe the organisation and the "Provider Organisation" column to describe the Agency/Ministry. For this reason, we decided to look for "Reporting Organisation" and "Provider Organisation" pairings that published across both data sources.

Identifying funders that publish to both IATI and CRS

In our project, we used a number of techniques for identifying which funders were publishing to both IATI and CRS. Below we have laid out these techniques and have also included any supporting documents that we created during the process.

Direct comparison of funders across data sources

The first method we used for identifying duplicate pairings involved creating a list of distinct "Reporting Organisation" and "Provider Organisation" pairings for both IATI and CRS, like shown in figure 8.

Reporting Organisation	Provider Organisation	Data Source
50 Eight Limited [GB-COH-07291220]	50 Eight Limited [GB-COH-07291220]	IATI
50 Eight Limited [GB-COH-07291220]	World Vision UK [GB-CHC-285908]	IATI
ACTION AGAINST HUNGER SPAIN [ES-CIF-G81164105]	ACTION AGAINST HUNGER SPAIN [ES-CIF-G81164105]	IATI
ADD International (Action on Disability and Development) [GB-COH-294860]	ADD International (Action on Disability and Development) [GB-COH-294860]	IATI
ADD International (Action on Disability and Development) [GB-COH-294860]	Sightsavers [GB-CHC-207544]	IATI
ADRA Denmark [DK-CVR-20074035]	ADRA Denmark [DK-CVR-20074035]	IATI
AECOM Limited [GB-COH-GB-COH-01846493]	AECOM Limited [GB-COH-GB-COH-01846493]	IATI
AECOM Limited [GB-COH-GB-COH-01846493]	DFID	IATI

Figure 8: Distinct pairs of Reporting/Provider Organisations in IATI

A comparison between the two lists can help identify any funders that publish to both data sources. This can initially be done automatically by running searches to find Reporting/Provider Organisation pairings that occur in both lists. However, since many funders use different names in each data source, this will only help identify a small proportion of the duplicates. Our project team manually compared the two lists to identify as many duplicates as possible. It should be noted that one Reporting/Provider organisation pairing in IATI does not always match with only one Reporting/Provider organisation paring in CRS, and vice-versa. This will become clearer when observing the 'Duplicate Quality Decisions' table. This manual process is unfortunately a lengthy one and does require some prior knowledge of funding organisations in order to accurately identify all the duplicate pairings. We have attached to this report our 'Duplicate Quality Decisions' table which will hopefully shorten the process for you, however, since new funders are being added to IATI constantly, this list will become incomplete very quickly.

Searching for titles that appear in both data sources

Another technique that we used to identify duplicates was to search for titles that appeared across both data sources. This technique can be useful to identify duplicate pairings that have not been picked up during the previous stage, however, care needs to be taken when analysing the projects to decide whether it is a duplicate or not. This is due to the fact that projects can have multiple

funders who publish their own contributions to the project separately from each other and therefore their activities are in fact representing different funding to the same project. To ensure that we were not marking false duplicates, we looked through the d-portal entries for the project titles existing across both data sources and deduced whether it was in fact the same funding being described twice or separate funding altogether.

Code for finding project titles that appear in both data sources

The code below returns a list of project titles that appear across both data sources, along with their reporting and provider organisations. This list can be used to identify duplicate reporting/organisation pairings, however, the vast majority of the project titles returned by this code will not be due to duplicates. Firstly, the same project title could feasibly be used for unrelated activities. For example, a title like "Project Support" is relatively common and would not indicate a duplicate across sources. Secondly, as mentioned before, projects can have multiple funders who publish their own contributions to the project separately from each other. Care therefore needs to be taken when using this method to identify duplicates and further analysis of a project would need to be done to ensure that a project was indeed a duplicate.

```
1
2  CREATE VIEW "Duplicate_titles" AS
3  SELECT "Title" from combined
4  GROUP BY "title"
5  HAVING count(DISTINCT("data source")) = 2;
6  SELECT "title", "reporting organisation", "provider organisation", "data source",
     "unique id"  FROM combined
7  WHERE "title" IN (SELECT * FROM titles)
8  GROUP BY "title", "reporting organisation", "provider organisation", "data source
   "
9
10
```

Creating a 'Duplicate Quality Decisions Table'

When identifying our duplicate reporting/provider organisation pairings, we will want to add them into a table so that we can easily make our quality decisions later. It is important when doing this step that you copy the exact reporting and provider names including any symbols or spacing so that when we go to remove projects later on, we are able to easily implement joins in SQL using this table. We found that creating this table in excel was the easiest. We simply created a table with the reporting and provider organisations and used a new column *Match_Number* to specify which pairings had duplicates between them. Figure 9 shows an example of how this should be formatted.

Quality decision

Once we had identified duplicate Reporting/Provider Organisation pairings across IATI and CRS, it is necessary to remove one of the sources from the final data to ensure no duplicate projects remain in our final dataset. For our project, we made separate decisions for grants and non-grants data for each of our countries. This was due to the fact that organisations often have different departments responsible for the publication of aid data for each country and also often separated by Grants and Non-Grants too, resulting in varying quality of data for each of these categories. Our methodology for deciding which source to keep involve a number of factors, some of which were specific to the overall requirements for our specific project. Factors that we analysed included, but were not limited to:

- Number of Projects

Match number	Reporting org.	Provider org.	Data sources
1	**Asian Development Bank**	Asian Development Bank	CRS
1	**Asian Development Bank**	Technical Assistance Special Fund (TASF)	CRS
1	**Asian Development Bank [46004]**	Asian Development Bank	IATI
1	**Asian Development Bank [46004]**	Asian Development Bank [46004]	IATI
1	**Asian Development Bank [XM-DAC-46004]**	Asian Development Bank	IATI
1	**Asian Development Bank [XM-DAC-46004]**	Asian Development Bank [XM-DAC-46004]	IATI
2	**Australia**	Australian Government	CRS
2	**Australia**	Miscellaneous	CRS
2	**Australia - Department of Foreign Affairs and Trade [AU-5]**	Australia - Department of Foreign Affairs and Trade [AU-5]	IATI
3	**Bill & Melinda Gates Foundation**	Bill & Melinda Gates Foundation	CRS
3	**Bill & Melinda Gates Foundation [DAC-1601]**	Bill & Melinda Gates Foundation [DAC-1601]	IATI

Figure 9: Example duplicate table

- Total funding amount
- Average Length of Description
- Frequency of use of a Policy Marker

Figure 10 shows how we would begin to inform a quality decision using the example of Bill and Melinda Gates Foundation. In this example our decision was to keep the IATI data, since the total disbursements was significantly higher, however due to the requirements of our project, there were many instances where we sacrificed higher total disbursements when the average length of the project descriptions was much higher. This was due to the fact that our project involved conducting key word searches on project descriptions, however, the quality decision process can be adjusted to prioritise the variables which are of the most importance of your projects research.

	DataSource	ReportingOrganisation	ProviderOrganisation	Total Disbursments	Distinct Titles	Description Length
1	CRS	bill & melinda gates foundation	bill & melinda gates foundation	1528.99192606	816	223.742371185593
2	IATI	bill & melinda gates foundation [dac-1601]	bill & melinda gates foundation [dac-1601]	4882.25412019638	673	191.432221986267

Figure 10: Example duplicate decision

Code for creating Match number summary tables

This code will create a summary table for both *Data Sources* for each *match_number*. The code will return each of variables listed below for both data sources in order to aid your decision on which data source to keep.

- Total number of projects
- Average description length
- Percentage of projects with description length over 100 characters (100 can be changed to any number as required)
- Percentage of projects that use the Gender Marker (Can be changed to any Policy marker easily)
- Total Disbursements/Commitments (USD)

This example is for grants as we decided to split our decisions by grants and non-grants, however the code should be modified to create the summary tables for non-grants projects.

```
1
2  ALTER TABLE combined ADD COLUMN "description indicator";
3  UPDATE combined SET "description indicator" = 0;
4  UPDATE combined SET "description indicator" = 1
5  WHERE length(description) > 100;
6  ALTER TABLE combined ADD COLUMN "Gender marker indicator";
7  UPDATE combined SET "Gender marker indicator" = 0;
8  UPDATE combined SET "Gender marker indicator" = 1
9  WHERE "gender marker" = 0 OR "gender marker" = 1 or "gender marker" = 2;
10 CREATE VIEW "Quality Decision Table" AS
11 SELECT "match_number", "data source" , avg(length(description)) as "Average
   Description Length", avg("Description indicator") as "Description Proportion",
   avg("Gender marker indicator") as "Gender Marker Proportion", count(title), sum(
   "Value (USD)")+sum("USD_Disbursement") as "Total Value (USD)" from combined
12 WHERE "match_number" > 0
13 GROUP BY "data source", "match_number";
14 ALTER TABLE combined DROP COLUMN "description indicator";
15 ALTER TABLE combined DROP COLUMN "Gender marker indicator";
16
17
```

Year check swaps

The publication to each data source by funder also varies by year. For example, in Bangladesh, IFAD published to CRS between 2015-2019. But in IATI data they only publish for the years 2015-16 and 2018-19. Therefore, there is only an overlap of data for four years. IFAD's data quality was richer in IATI but for the year 2017 IFAD only reported to CRS. To make sure that we don't miss out on IFAD's data for 2017, it is necessary for us to make a swap on our quality decision for 2017. Another possible approach to these quality decisions would be to make separate decisions for each year, however, we decided against this method as it was not feasible for the size of our project.

Code for conducting year check swaps

The code included below will make these year checks and change the duplicate decisions in the main data accordingly. This code will have to be applied separately for each of the countries in the data and will also have to be applied separately for grants and non-grants. After each iteration the new *combined* data will be a new version with where the suffix (*e.g. combined_v2*) should be adjusted adjusted accordingly. The code will therefore need to be adjusted between iterations to reflect this, along with adjustments for the country name and grants/non-grants.

```
1      -- First trim and lowercase the tables to aid joining
2
3  UPDATE Combined SET "Reporting Organisation"= TRIM("Reporting Organisation"), "
   Provider Organisation" = TRIM("Provider Organisation");
4
5  UPDATE Combined SET "Reporting Organisation"= LOWER("Reporting Organisation"), "
   Provider Organisation" = LOWER("Provider Organisation");
6
7
8  -- Then create a view of the Ethiopia subset of combined
9
10 CREATE VIEW "combined_country_one" AS
11
12 SELECT *
13
14 FROM combined
15
```

```
16 WHERE "recipient country" = "Country One" AND "grants_to_keep" = 1;
17
18
19 -- Create a list of match number/ year combinations that are marked for removal
20
21 CREATE VIEW "Duplicate_delete_country_one" AS
22
23 SELECT "year","match_number","Data source", "reporting organisation", "Provider
   organisation"
24
25 FROM "combined_country_one"
26
27 WHERE "duplicate grants keep" = 0
28
29 GROUP BY "match_number", "data source","year";
30
31 -- Create a list of match number/ year combinations that are marked for keeping
32
33 CREATE VIEW "Duplicate_keep_country_one" AS
34
35 SELECT "year","matchnumber","Data source", "reporting organisation", "Provider
   organisation"
36
37 FROM "combined_country_one"
38
39 WHERE "duplicate grants keep" = 1
40
41 GROUP BY "match_number", "data source","year";
42
43 -- Find match number/year combinations that are marked for removal that do not
   exist in       duplicate_delete_match
44
45 CREATE VIEW "year_swap" as
46
47 SELECT Duplicate_delete_country_one.Match_number AS Matchnumber_delete,
   Duplicate_delete_country_one.Year AS "Year_delete", Duplicate_keep_country_one.
   year AS "year_keep",Duplicate_delete_country_one."Data Source",
   Duplicate_delete_country_one."reporting organisation",
   Duplicate_delete_country_one."provider organisation"
48
49 FROM Duplicate_delete_country_one
50
51 LEFT JOIN Duplicate_keep_country_one ON Duplicate_delete_country_one.Match_number
    = Duplicate_keep_country_one.Match_number and Duplicate_delete_country_one.year
   = Duplicate_keep_country_one.year
52
53 WHERE "year_keep" ISNULL;
54
55
56 -- Create new table with the match numbers joined back in when we want to change
   the duplicate keep to 1.
57
58 CREATE VIEW "combined_v2" AS
59
60 SELECT * FROM Combined
61
62 LEFT JOIN year_swap ON combined."year" = year_swap."year_delete" AND combined."
   reporting organisation" = year_swap."reporting organisation" AND combined."
   provider organisation" = year_swap."provider organisation" AND combined."data
   source" = year_swap."data source";
63
64 -- Update the table accordingly (Change years as necessary)
65
```

```
66 UPDATE combined_v2 SET "duplicate grants keep" = 1 WHERE "Matchnumber_delete" IS
   NOT NULL AND "year" IN (2015, 2016, 2017, 2018, 2019, 2020);
67
68 -- Delete unnecessary columns
69
70 ALTER TABLE combined_v2
71
72 DROP COLUMN "year_delete";
73
74
75 -- Repeat for Loans/Other countries
76
77
```

6. Double counts

We defined double counting as when two separate organisations publish the same activities resulting in the same funding being represented in the data. This usually occurs when both the funding organisation and another implementing partner of a project both decide to publish data to IATI for that project. To identify instances when double counting occurred, we first identified any organisations which have been listed, in the data, as both a 'reporting organisation' and a 'provider organisation'. We would then, remove the projects for which the organisation was listed as the 'provider organisation'. So, for example, all activities in IATI where 'Netherlands MFA (NL, MFA)' was the 'reporting organisation' would be left in that data but, any activities which list another organisation as the 'reporting organisation' (that is not the NL, MFA) and which lists the NL, MFA as the 'provider organisation' will be removed. This is under the assumption that if the NL, MFA is listed as a provider by other publishers that: 1. The NL, MFA is already reporting the activity and 2. the publisher has published only that activity in its entirety. The OECD-DAC remove double counts in the CRS data so there the steps laid out in this sections should only be applied to IATI data.

Identifying the organisations with the highest double counts

A search for the number of times each 'reporting organisation' is listed as a 'provider organisation' should be repeated for each country in order to identify the publishers with the highest instances of double counting. In our project, we only considered reporting organisations that were listed as a provider organisation by at least 5 other reporting organisations, as this represents the overwhelming majority of doubly counted projects. Figure 11 shows an example of the start of a list produced by the code below.

	Provider Organisation	count(distinct title)
1	United States [US-USAGOV]	4287
2	GlobalGiving.org [US-EIN-300108263]	1557
3	European Commission - International...	1410
4	UK - Foreign, Commonwealth and ...	1007
5	United States Agency for Internation...	957

Figure 11: Example list of reporting organisations that have been listed as a provider organisation at least 5 times.

Code for identifying double counts

This code will produce a list of reporting organisations that have been listed as a provider organisation at least 5 times.

```
1
2  CREATE VIEW "reporting" AS
3  SELECT DISTINCT reporting organisation FROM combined;
4
5  SELECT "provider organisation", COUNT(DISTINCT title) FROM combined
6  WHERE "provider organisation" IN (SELECT * FROM "reporting") AND "data source" =
   "IATI"
7  GROUP BY "provider organisation"
8  HAVING COUNT(DISTINCT title) > 5
9  ORDER BY COUNT(DISTINCT title) DESC
10
11
```

Identifying the variations of the reporting organisation name

Once the list of reporting organisations with the highest number of double counts has been identified, a more thorough search of the double count activities will be made. The search will be made to account for differences in spelling and language, for example, 'Netherlands MFA' is sometimes

published as ‘Nederlands Ministerie van Buitenlandse Zaken’. In fact, for our set of countries we found that ‘Netherlands MFA’ was associated with 10 written variations. To find these variations, we found it useful to consult the list of distinct reporting organisations that we created earlier and search for relevant words that might return related organisations. For example, we might search for ’Netherlands’ or ’Nederlands’ and look through the results to see if any of them were indeed equivalent. It can also be helpful to look at the website of the given agency as they might list any translations of their name. Once we have compiled the full list of written variations we can create a table for reference showing the number of associated reporting organisations for each. Figure 12 is an example of such a table for ’Netherlands MFA’.

	ProviderOrganisation	Count of distinct reporting organisations
	Filter	Filter
1	[min. of foreign affairs]	1
2	ministerie van buitenlandse zaken	1
3	ministry of foreign affairs of the ...	1
4	netherlands - ministry for foreign trad...	1
5	netherlands - ministry of foreign affai...	64
6	netherlands - ministry of foreign affai...	1
7	netherlands ministry of foreign affairs	2
8	netherlands ministry of foreign affairs...	1
9	netherlands ministry of foreign affairs...	3
10	netherlands. ministerie van ...	1

Figure 12: Count of reporting organisations that list each of the written variations of ’Netherlands MFA’ as a provider organisation.

Marking double count projects for removal

Once we have identified all the written variations of each relevant provider organisation, we can start recording them in a table so that we can mark their activities for removal from the data. We can simply create a list in excel of the provider organisations that we identify. This can then be exported as a CSV file ("Double counts") so that we can join this back into the main data.

Code for joining double count information back into main data

Once we have imported the CSV into the db browser, the code below will add a column into the main data *"Double Counts Keep"* which will be populated with a 0 if the activity is a double count and a 1 otherwise.

```
1
2  ALTER TABLE combined ADD COLUMN "Double Counts Keep";
3  UPDATE TABLE combined SET "Double Counts Keep" = 1;
4  UPDATE TABLE combined SET "Double Counts Keep" = 0
5  WHERE "provider organisation" IN (SELECT * FROM "double counts")
6
```

7. Data cleaning

Funders are often inconsistent in the way that they publish their data, especially to IATI. It was therefore necessary to apply some data cleaning to ensure consistency when we conducted our analysis. In this section, we will outline the steps that we took to clean each of the variables in our data. The process that we used for each variable followed the same basic structure, and hence should be very easily modified to apply to any additional variables that you choose to use in your project.

Reporting Organisation & Provider Organisation

The Reporting/Provider Organisation columns was where we saw the most inconsistency in funders publishing style. Figure 13 shows an example of these inconsistencies in the 'provider organisation' column and how we cleaned them up in our final data. To create the lists of cleaned up reporting/provider organisation names, our approach was simple. We first created a list of distinct reporting/provider names and then went through and made our choice for the cleaned-up name. We will include all of the tables containing these choices in this report which should speed up the process greatly, however, as new organisations are being added to IATI every day, a thorough check of all reporting/provider organisation names will still have to be made.

You should create a CSV file with every distinct reporting organisation and its cleaned name, as shown in Figure 13. You will then be able to import it into the db browser and then implement the code below to join the cleaned names into the main data. This process will have to be repeated for the provider organisation column.

Original	Cleaned
department for international development [gb-1]	United Kingdom - Foreign, Commonwealth and Development Office
department for international development [gb-gov-1-300788]	United Kingdom - Foreign, Commonwealth and Development Office
department of international development (dfid) [gb-1]	United Kingdom - Foreign, Commonwealth and Development Office
dfid	United Kingdom - Foreign, Commonwealth and Development Office
dfid (the department for international development) [gb-1]	United Kingdom - Foreign, Commonwealth and Development Office
dfid [1 gb-1 10]	United Kingdom - Foreign, Commonwealth and Development Office
dfid [7374]	United Kingdom - Foreign, Commonwealth and Development Office
dfid [dfid gb-1]	United Kingdom - Foreign, Commonwealth and Development Office
dfid [dfid-gb-1]	United Kingdom - Foreign, Commonwealth and Development Office
dfid [dfid]	United Kingdom - Foreign, Commonwealth and Development Office
dfid [gb-1-203559]	United Kingdom - Foreign, Commonwealth and Development Office
dfid [gb-1]	United Kingdom - Foreign, Commonwealth and Development Office
dfid [gb-gov-1-300083]	United Kingdom - Foreign, Commonwealth and Development Office
dfid [gb1]	United Kingdom - Foreign, Commonwealth and Development Office
dfid [gpaf-imp-057]	United Kingdom - Foreign, Commonwealth and Development Office

Figure 13: Provider Name cleaning

Code for joining cleaned reporting organisation names back into the main data

The code below will join the cleaned reporting organisation names from the CSV file 'reporting_cleaned) back into the main data.

```
1
```

```
2  UPDATE TABLE "combined" SET "reporting organisation" = trim(lower("reporting
   organisation"));
3  UPDATE TABLE "reporting_cleaned" SET "reporting organisation" = trim(lower("
   reporting organisation"));
4  CREATE TABLE "combined_v2" AS
5  SELECT * FROM "combined"
6  LEFT JOIN "reporting_cleaned" ON "combined"."reporting organisation" = "
   reporting_cleaned"."reporting_organisation";
7
```

Aid Type & Finance Type

The Aid Type and Finance Type cleaning should be much less time consuming as there are typically far fewer different unique values that they can take. Figure 14 shows an example of how the aid type would be cleaned up. The process for joining the cleaned Aid/Finance Types into the main data will be exactly the same as described above for the reporting organisation, however, the code will have to be updated to reflect the different variables.

Original	Cleaned
A01	A01 - General budget support
A01 - General budget support	A01 - General budget support
A02	A02 - Sector budget support
A02 - Sector budget support	A02 - Sector budget support

Figure 14: Aid Type cleaning

Conclusion

For more information on the Women's Economic Empowerment project and to read the full reports, please visit our WEE page here.

Finance Type	Description	Non-grants Y/N	Grants Y/N
110 - Standard grant	Grants are transfers in cash or in kind for which no legal debt is incurred by the recipient.	N	Y
422 - Reimbursable grant	A contribution provided to a recipient institution for investment purposes, with the expectation of long-term reflows at conditions specified in the financing agreement. The provider assumes the risk of total or partial failure of the investment; it can also decide if and when to reclaim its investment.	N	Y
410 - Aid loan excluding debt reorganisation	*Withdrawn in 2022 – description no longer available*	Y	N
411 - Investment-related loan to developing countries	*Withdrawn in 2022 – description no longer available*	Y	N
413 - Loan to national private investor	*Withdrawn in 2022 – description no longer available*	Y	N
421 - Standard loan	Transfers in cash or in kind for which the recipient incurs legal debt (and the resulting claim is not intended to be traded). Since payment obligations on standard loan are senior obligations, i.e. creditors are entitled to receive payments against their claims before anyone else, they are also referred to as senior loans.	Y	N
431 - Sub-ordinated loan	A loan that, in the event of default, will only be repaid after all senior obligations have been satisfied. In compensation for the increased risk, mezzanine debt holders require a higher return for their investment than secured or more senior lenders.	Y	N
510 - Common equity	A share in the ownership of a corporation that gives the owner claims on the residual value of the corporation after creditors' claims have been met.	Y	N
511 - Acquisition of equity not part of joint venture in developing countries	*Withdrawn in 2022 – description no longer available*	Y	N
520 -Shares in collective investment vehicles	Collective undertakings through which investors pool funds for investment in financial or nonfinancial assets or both. These vehicles issue shares (if a corporate structure is used) or units (if a trust structure is used).	Y	N
432 - Preferred equity	Equity that, in the event of default, will be repaid after all senior obligations and subordinated loans have been satisfied; and will be paid before common equity holders. It is a more expensive source of finance than senior debt, a less expensive source than equity.	Y	N
433 - Other hybrid instruments	Including convertible debt or equity.	Y	N
1100 - Guarantees/insurance	*No description available*	Y	N

Finance Type	Description	Non-grants Y/N	Grants Y/N
111 - Subsidies to national private investors	*Withdrawn in 2022 – description no longer available*	N	N
210 - Interest subsidy	A payment to soften the terms of private export credits, or loans or credits by the banking sector.	N	N
310 - Capital subscription on deposit basis	Payments to multilateral agencies in the form of notes and similar instruments, unconditionally encashable at sight by the recipient institutions.	N	N
311 - Capital subscription on encashment basis	Payments to multilateral agencies in the form of notes and similar instruments, unconditionally encashable at sight by the recipient institutions.	N	N
610 - Debt forgiveness: ODA claims (P)	*No description available*	N	N
611 -Debt forgiveness: ODA claims (I)	*No description available*	N	N
618 - Debt forgiveness: Other	*No description available*	N	N
621 -Debt rescheduling: ODA claims (I)	*No description available*	N	N
912 - Purchase of securities from issuing agencies	*Withdrawn in 2022 – description no longer available*	N	N
Blank		B	B

Table 5: Finance Type Filtering

Transaction Type	Non-grants Y/N	Grants Y/N
1 - Incoming Funds	N	N
2 - Outgoing Commitment	Y	N
3 - Disbursement	N	Y
4 - Expenditure	N	Y
A - Authorised	Y	N
P - Paid	N	Y

Table 6: Transaction Type Filtering

Flow Type	Non-grants Y/N	Grants Y/N
19 - Equity Investment	Y	N
11 - ODA Grants	N	Y
13 - ODA Loans	Y	N
14 - Other Official Flows (non-export credit)	B	B
30 - Private Development Finance	Y	B
21 - Non-export credit OOF	Y	B
50 - Other flows	B	B
20 - OOF	B	B
35 - Private Market	B	B
Blanks	B	B

Table 7: Flow Type Filtering

Match number	Reporting org.	Provider org.	Data sources
1	Asian Development Bank	Asian Development Bank (only)	CRS
1	Asian Development Bank [XM-DAC-46004]	Asian Development Bank	IATI
2	Australia	Australian Government + Misc.	CRS
2	Australia - Department of Foreign Affairs and Trade [AU-5]	Australia - Department of Foreign Affairs and Trade [AU-5]	IATI
3	Bill & Melinda Gates Foundation	Bill & Melinda Gates Foundation	CRS
3	Bill & Melinda Gates Foundation [DAC-1601]	Bill & Melinda Gates Foundation [DAC-1601]	IATI
4	Canada	Global Affairs Canada (only)	CRS
4	Canada - Global Affairs Canada — Affaires mondiales Canada [CA-3]	Canada - Global Affairs Canada — Affaires mondiales Canada [CA-3]	IATI
5	Canada	International Development Research Centre	CRS
5	Canada - International Development Research Centre/-Centre de recherches pour le développement international [XM-DAC-301-2]	Multiple agencies	IATI
6	CDC Group plc [GB-COH-03877777]	CDC Group plc [GB-COH-03877777]	IATI
6	United Kingdom	CDC Capital Partners PLC	CRS
7	UK - Department for Business, Energy and Industrial Strategy (BEIS) [GB-GOV-13]	UK - Department for Business, Energy and Industrial Strategy (BEIS) [GB-GOV-13]	IATI
7	UK - Department for Environment, Food and Rural Affairs [GB-GOV-7]	UK - Department for Environment, Food and Rural Affairs [GB-GOV-7]	IATI
7	UK - Foreign, Commonwealth and Development Office [GB-GOV-1]	UK - Foreign, Commonwealth and Development Office [GB-GOV-1]	IATI
7	United Kingdom	ALL (excluding Scottish/Welsh; CSSF & PF)	CRS
8	Central Emergency Response Fund	Rapid Response	CRS
8	United Nations Central Emergency Response Fund (CERF) [XM-OCHA-CERF]	United Nations Central Emergency Response Fund (CERF) [XM-OCHA-CERF]	IATI
9	Denmark	All	CRS
9	Denmark - Ministry of Foreign Affairs, Danida [XM-DAC-3-1]	Denmark - Ministry of Foreign Affairs, Danida [XM-DAC-3-1]	IATI
10	EU Institutions	European Commission (ONLY)	CRS
10	European Commission - Humanitarian Aid & Civil Protection [XI-IATI-EC_ECHO]	European Commission - Humanitarian Aid & Civil Protection [XI-IATI-EC_ECHO]	IATI
10	European Commission - International Partnerships	European Commission - International Partnerships	IATI

10	European Commission - Service for Foreign Policy Instruments [XI-IATI-EC_FPI]	European Commission - Service for Foreign Policy Instruments [XI-IATI-EC_FPI]	IATI
11	UNFPA	UNFPA	CRS
11	United Nations Population Fund [41119]	United Nations Population Fund [41119]	IATI
12	Finland	ALL (exc. Finnfund)	CRS
12	Finland - Ministry for Foreign Affairs [FI-3]	Finland - Ministry for Foreign Affairs [FI-3]	IATI
13	Food and Agriculture Organisation	FAO	CRS
13	Food and Agriculture Organization of the United Nations (FAO) [XM-DAC-41301]	ALL	IATI
14	France	Ministry of Foreign Affairs (ONLY)	CRS
14	France - Ministry for Europe and Foreign Affairs [FR-6]	France - Ministry for Europe and Foreign Affairs [FR-6]	IATI
15	Agence Française de Développement [FR-3]	Agence Française de Développement [FR-3]	IATI
15	France	French Development Agency	CRS
16	Gavi, the vaccine alliance [47122]	ALL	IATI
16	Global Alliance for Vaccines and Immunization	GAVI	CRS
17	Germany	Foreign Office (ONLY)	CRS
17	Germany - Federal Foreign Office [XM-DAC-5-7]	Germany - Federal Foreign Office [XM-DAC-5-7]	IATI
18	Germany	Kreditanstalt für Wiederaufbau; Bundesministerium f³r Wirtschaftliche Zusammenarbeit und Entwicklung (ONLY)	CRS
18	Germany - Ministry for Economic Cooperation and Development [DE-1]	ALL	IATI
19	Global Fund	Global Fund	CRS
19	The Global Fund to Fight AIDS, Tuberculosis and Malaria [47045]	The Global Fund to Fight AIDS, Tuberculosis and Malaria [47045]	IATI
20	IFAD	IFAD	CRS
20	International Fund for Agricultural Development (IFAD) [XM-DAC-41108]	International Fund for Agricultural Development (IFAD) [XM-DAC-41108]	IATI
21	International Development Association	IDA	CRS
21	The World Bank [44000]	International Development Association [44002]	IATI
22	International Finance Corporation	International Finance Corporation	CRS
22	International Finance Corporation [XM-DAC-903]	International Finance Corporation [XM-DAC-903]	IATI
23	International Labour Organisation	ALL	CRS

23	International Labour Organization (ILO) [XM-DAC-41302]	ALL	IATI
24	Ireland	Department of Foreign Affairs (ONLY)	CRS
24	Ireland - Department of Foreign Affairs and Trade [XM-DAC-21-1]	Ireland - Department of Foreign Affairs and Trade [XM-DAC-21-1]	IATI
25	AICS - Agonize Italiana per la Cooperazione allo Sviluppo / Italian Agency for Cooperation and Development [XM-DAC-6-4]	AICS - Agenzia Italiana per la Cooperazione allo Sviluppo / Italian Agency for Cooperation and Development [XM-DAC-6-4]	IATI
25	Italy	Direzione Generale per la Cooperazione allo Sviluppo (ONLY)	CRS
26	Japan	ALL (exc. Agriculture; prefectures)	CRS
26	Ministry of Foreign Affairs of Japan [XM-DAC-701-2]	Ministry of Foreign Affairs [JP-2] (ONLY)	IATI
27	Japan	Japan International Cooperation Agency (JICA) [XM-DAC-701-8]	CRS
27	Ministry of Foreign Affairs of Japan [XM-DAC-701-2]	Japanese International Co-operation Agency	IATI
28	Korea	Ministry of Foreign Affairs and Trade & Misc.	CRS
28	Republic of Korea [KR-GOV-010]	All (Exc. KOIC & Export-Import bank)	IATI
29	Korea	Korea International Cooperation Agency	CRS
29	Republic of Korea [KR-GOV-010] - KOICA	Korea International Cooperation Agency [KR-GOV-051]	IATI
29	Korea - Export-import bank	Export-Import Bank of Korea	CRS
29	Republic of Korea [KR-GOV-010] -Export-import bank	The Export-Import Bank of Korea(Economic Development Cooperation Fund) [KR-GOV-021]	IATI
30	Slovak Aid [XM-DAC-69-2]	Slovak Aid [XM-DAC-69-2]	IATI
30	Slovak Republic	Slovak Agency for International Deve (SAMRS)	CRS
31	Ministry of education, science, research and sport of the Slovak republic [XM-DAC-69-4]	Ministry of education, science, research and sport of the Slovak republic [XM-DAC-69-4]	IATI
31	Slovak Republic	Ministry of Education, Science, Rese (MSVVS)	CRS
32	Ministry of Foreign and European Affairs of the Slovak Republic [XM-DAC-69-1]	Ministry of interior of the Slovak Republic [XM-DAC-69-5]	IATI
32	Slovak Republic	Ministry of Interior	CRS
33	Netherlands	Ministry of Foreign Affairs of the Netherlands (ONLY)	CRS
33	Netherlands - Ministry of Foreign Affairs [XM-DAC-7]	Netherlands - Ministry of Foreign Affairs [XM-DAC-7]	IATI
34	New Zealand	Ministry of Foreign Affairs and Trade	CRS

34	New Zealand - Ministry of Foreign Affairs and Trade - New Zealand Aid Programme [NZ-1]	New Zealand - Ministry of Foreign Affairs and Trade - New Zealand Aid Programme [NZ-1]	IATI
35	Norad - Norwegian Agency for Development Cooperation [NO-BRC-971277882]	Norwegian Ministry of Foreign Affairs - Embassies	IATI
35	Norway	Ministry of Foreign Affairs	CRS
36	Norad - Norwegian Agency for Development Cooperation [NO-BRC-971277882]	Norad - Norwegian Agency for Development Cooperation [NO-BRC-971277882]	IATI
36	Norway	Norwegian Agency for Development Co-operation	CRS
37	Spain - Ministry of Foreign Affairs and Cooperation [ES-DIR3-E04585801]	Spain - Ministry of Foreign Affairs and Cooperation [ES-DIR3-E04585801]	IATI
37	Spain	Spanish central ministires	CRS
38	Spain	Spanish Agency for International Development Co-operation	CRS
38	Spanish Agency for International Development Cooperation (AECID) [ES-DIR3-EA0035768]	Spanish Agency for International Development Cooperation (AECID) [ES-DIR3-EA0035768]	IATI
39	Sweden	ALL (Exc. Swedfund)	CRS
39	Sweden, through Swedish International Development Co-operation Agency (Sida) [SE-0]	Sweden, through Swedish International Development Co-operation Agency (Sida) [SE-0]	IATI
40	Switzerland	Swiss Agency for Development and Co-operation (ONLY)	CRS
40	Switzerland - Swiss Agency for Development and Cooperation (SDC) [CH-4]	Switzerland - Swiss Agency for Development and Cooperation (SDC) [CH-4]	IATI
41	UNAIDS	UNAIDS	CRS
41	United Nations Joint Programme on HIV and AIDS Secretariat (UNAIDS) [XM-DAC-41110]	United Nations Joint Programme on HIV and AIDS Secretariat (UNAIDS) [XM-DAC-41110]	IATI
42	UNDP	UNDP	CRS
42	United Nations Development Programme (UNDP) [XM-DAC-41114]	ALL	IATI
43	UNHCR	UNHCR	CRS
43	United Nations High Commissioner for Refugees (UNHCR) [XM-DAC-41121]	United Nations High Commissioner for Refugees (UNHCR) [XM-DAC-41121]	IATI
44	UNICEF	UNICEF	CRS
44	United Nations Children's Fund (UNICEF) [XM-DAC-41122]	ALL	IATI

45	MasterCard Foundation	MasterCard Foundation	CRS
45	Mastercard Foundation [CA-CRA_ACR-817387277]	Mastercard Foundation [CA-CRA_ACR-817387277]	IATI
46	United Nations World Food Programme (WFP) [XM-DAC-41140]	ALL	IATI
46	WFP	WFP	CRS
47	United States- other depts.	ALL (Excl. Agriculture & Federal Trade Commission)	CRS
47	United States [US-USAGOV]	United States [US-USAGOV]	IATI
48	Millennium Challenge Corporation [US-18]	Millennium Challenge Corporation [US-18]	IATI
48	United States- MCC	Millennium Challenge Corporation	CRS
49	United States	Agency for International Development	CRS
49	United States Agency for International Development (USAID) [US-GOV-1]	United States Agency for International Development (USAID) [US-GOV-1]	IATI
50	World Health Organisation	ALL	CRS
50	World Health Organization [XM-DAC-928]	ALL	IATI
51	Charity Projects Ltd (Comic Relief)	Charity Projects Ltd (Comic Relief)	CRS
51	Charity Projects Ltd (Comic Relief) [GB-CHC-326568]	Charity Projects Ltd (Comic Relief) [GB-CHC-326568]	IATI
52	African Development Bank	ALL	CRS
52	African Development Bank [46002]	ALL	IATI
52	African Development Fund	ALL	CRS
53	International Bank for Reconstruction and Development	International Bank for Reconstruction and Development	IATI
53	The World Bank	International Bank for Reconstruction and Development	CRS
54	EU Institutions	European Investment Bank	CRS
54	European Investment Bank [XM-DAC-918-3]	European Investment Bank [XM-DAC-918-3]	IATI
55	Belgian Development Cooperation [XM-DAC-2-10]	Belgian Development Cooperation [XM-DAC-2-10]	IATI
55	Belgium	Directorate General for Co-operation and Development (ONLY)	CRS

Table 8: Duplicate donors table