AidData

AidData Data Management Plan
Version 1.0
October 28, 2015

Authors:
Scott Stewart (sstewart@aiddata.org)
Brooke Russell (brussell@aiddata.org)

Approved By:
Brad Parks, Co-Executive Director of AidData

10/28/2015

Table of Contents – Data Management Plan Documents

AidData Data Management Documents

Data Collection Documents
Donor Systems Data Collection Management Plan
Recipient Systems Data Collection Management Plan
Reporting Organization System Data Collection Management Plan
Pure Aggregates Data Collection Management Plan
Product Definitions
Portal Data Product
Core Research Release Product Definition
Geocoded Data Research Release Product Management Plan
IATI Product Definition

The AidData Data Management Documents provide detailed information on our data collection, standardization, and publishing procedures for all standard data products.

AidData Data Management Master Document

Authors:
Scott Stewart (sstewart@aiddata.org)
Brooke Russell (brussell@aiddata.org)

Document History and Version History

Version	Authors	Date Approved	Approved By	Description
1.0	Scott Stewart, Brooke Russell	10/28/2015	Brad Parks, Co-Executive Director	This is the initial public release of the document.

Purpose

The purpose of the **AidData Data Management Master Plan (DMP)** is to address the management of data produced by AidData, from the time of their data collection, to their entry into permanent archive. This document is the Master Document, which defines at a top level, the processes that AidData undertakes in the collection and production of data. The Master document has many support documents which go into greater detail about each of these processes. These documents and their relation to the Master Document are shown in Figure 1, Data Management Plan Structure.
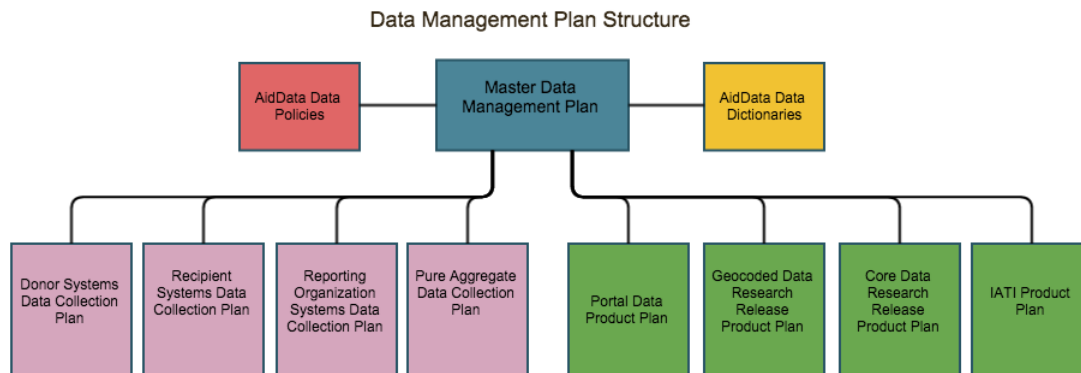
Figure 1: Data Management Plan Structure and related documents

Within the AidData organization, Data will typically flow from a **Data Collection Activity (DCA)**, potentially through a **Value Addition Activity (VA)**, and then to final **Product(s) and Archive.** During this data flow, data may be transformed at multiple steps, and not all intermediate data may be appropriate for public release or archive. Figure 2 represents the **Data Processing Flow** through these discrete lifecycle stages.

AidData Data Product Flow

Fig. 2, Data Processing Flow

1. Data Collection Activities (DCA)

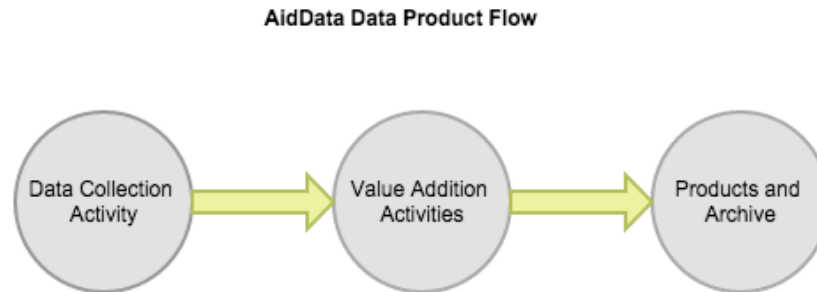A Data Collection Activity is any purposeful addition to AidData's core data holdings. It is the first stage of any Data Process Flow. Data Collection Activities can take place along two distinct product lines: **Project-Level Data** and **Pure Aggregate Data**. The sections and referenced documents below define each of these DCAs, with attendant Data Processing Flows. For definitions of common terms used in this document and AidData, please see the ***AidData Term Dictionary***.

1.1 Project-Level Data Collection Activities

Project-Level Data Collection Activities are any effort to gather specific project information. These DCA's are currently split into three (3) distinct subcategories: **Donor Systems Data Collection Activities**, **Recipient Systems Data Collection Activities** and **Reporting Organization Systems Data Collection Activities.** Common to all Project-Level Data Collection Activities are the baseline operations in Table 1. Additional operations, transforms, schedules and caveats for each individual Data Collection Activity are listed in the individual data collection plans for each subcategory.

1.1.1 Donor System Data Collection Activities

Donor System Project-Level Data Collection Activities are any systematic effort to gather data directly from the Donor for inclusion into a data product. Processes to gather the data may include data reports or extracts provided directly to AidData from the Donor, access via API to a data store (such as the World Bank API), or web scraping of public donor web portals. The Donor Systems Data Collection Management Plan fully covers all aspects of data collection for each donor system in which AidData engages in a Data Collection Activity.

1.1.2 Recipient System Data Collection Activities

Recipient System Project-Level Data Collection Activities are any systematic effort to collect project-level data from an aid recipient for inclusion into a data product. Similar to Donor System Data Collection Activities, the processes to gather the data may include data reports or extracts provided directly to AidData from the aid recipient, access via API to a data store, or web scraping of recipient public web portals, such as public Aid Information Management

Systems (AIMs). The Recipient Systems Data Collection Management Plan fully covers all aspects of data collection for each recipient system that AidData engages in a Data Collection Activity for.

1.1.3 Reporting Organization System Data Collection Activities

Reporting Organization System Project-Level Data Collection Activities are any systematic effort to collect data from the from an organization that acts as an aggregator of project level aid, for inclusion into a data product. The most obvious example of this Data Collection Activity is the Organisation for Economic Cooperation and Development (OECD) Creditor Reporting System (CRS) data. The Reporting Organization System Data Collection Management Plan fully covers all aspects of data collection for each reporting organization system in which AidData engages in a Data Collection Activity.

1.2 Pure Aggregate Data Collection Activities

Pure Aggregate System Data Collection Activities are any systematic effort to collect data from the from data stores that have aid amounts that are stored/represented not at the project level, but as a single composite value (i.e. the total remittance flow amount to Honduras for the year 2008). Pure aggregates are not the same as derived aggregates, such as aggregating project level data into an aggregate amount (i.e. summing all the commitment amounts from the World Bank to Nepal for the year 2009). The Pure Aggregates Data Collection Management Plan fully covers all aspects of data collection for each Pure Aggregate source in which AidData engages in a Data Collection Activity.

2.0 Value Addition Activities

Value Addition Activities are any effort to add value to a particular data set. AidData currently supports four (4) Value Addition Activities: 1) Activity and Purpose Coding (the systematic addition of labels to project-level aid to categorize them for intended purpose); 2) Geocoding Activities (the systematic addition of location information to project-level aid record to assign geographic location); 3) Curation Activities (the linking, deduplication and presentation of project records gathered from different sources or data collection activities); and 4) Quality Assurance Activities (the transformation and verification of data according to standard AidData practice). A particular product may have one or more of these Value Addition Activities during its lifecycle.

2.1 Activity and Purpose Coding Activities

Activity and Purpose Coding Activities is the application of the AidData aid classification scheme to project-level aid. This is done at two levels: The Purpose Code Level and the Activity Code level. Each of these codes are assigned by a defined methodology for each Data Product, defined in the Data Product Management Plan for that particular product. This coding may be done by human coders, or my text classification algorithms. Each individual Data Product Management Plan will contain details about the specific methodology or algorithm used.

Regardless of the methodology, the codes assigned are standard, as defined in the AidData Activity Coding Codebook.

2.1.1 Purpose Coding Activities

The AidData Purpose Coding Activity is an extension of the widely used OECD CRS Purpose Codes. Whereas the OECD system seeks to capture a single overall purpose of any given aid project, AidData attempts to capture the overall purpose and each individual activity taking place as part of the development project. AidData's activity codes, like purpose codes, can be tracked as an extension of the broad, overarching sectors outlined by the OECD. Not all Data Products are Purpose Coded. Please see each individual Data Product Management Plan to verify if a Purpose Coding Activity is engaged in for a particular Data Product.

2.1.2 Activity Coding Activities

The AidData Activity Coding Activity scheme is similar to the Purpose Coding Activity, only at a more granular level. During the coding process, activity codes are attached to each project based upon the project information. These activity codes are a subset of their respective purpose codes, but the activity codes assigned to a project can come from various sectors. In general, every project record is assigned at least one activity code, but there is usually no limit on the number of activity codes that can be associated with a particular record.

2.2 Geocoding Activities

Geocoding Activities are the application of the AidData geocoding methodology to project-level aid. As with Purpose and Activity Coding, this value addition activity may be done by human coders, or by an algorithmic (machine learning) method. Each of these codes is assigned by a defined methodology for each Data Product, defined in the Data Product Management Plan for that particular product. This geocoding (sometimes cited as georeferencing) is usually done in accordance with the UCDP and AidData Codebook On Georeferencing Aid, unless otherwise stated in the Data Product Management Plan for the particular product.

2.3 Curation Activities

Curation Activities are any systematic linking of multiple versions of project-level aid records from multiple sources. In this manner, AidData hopes to create a 'best of breed' view of a particular Data Product. This activity is most often engaged in for the Portal Data product, where multiple sources (for example, OECD CRS records from the World Bank and web scraped or API records from the World Bank) are substituted on a donor-year level to create a more enhanced data product. At the donor-year level, this means we show projects for a donor from one source for specific years and another source for other years. At its most detailed level, Best of Breed would allow us match data from multiple sources (even in the same year) and choose the best source for each field individually. *Note: Currently, AidData only publishes one source per donor-year. The full version of the 'best of breed' model has not yet been implemented in the published Portal data. Before doing so, a more accurate way of matching rows among different sources is needed.*

2.4 Quality Assurance Activities

Quality Assurance Activities are any processes undertaken to standardize a Data Product, or to attempt to improve the accuracy and compatibility, or usability of a Data Product. Quality Assurance Activities fall into six (6) sub-Activities: Currency Conversion and Deflation; 2) Field Standardization and Crosswalks; 3) Deduplication; 4) Field Consistency; 5) Spatial Consistency; 6) Record Verification and Triangulation; and 7) Health of Record. Each of these activities is detailed below. Each Data Product may make use of any or none of these Activity types during data production. The application and particular methodologies of these activities are detailed in the Data Product Management Plan for each product.

2.4.1 Currency Conversion and Deflation Activities

Currency Conversion and Deflation Activities are the standardization of financial amounts to common, comparable currency. It is AidData standard practice to convert all currency to USD and deflate that currency to a common year, so that year-over-year financial amounts may be compared. For example, if a project had a commitment amount of 230,00 Euros (EUR) made in 2005, AidData will first convert that to US Dollars (USD) and then deflate that to a common year USD (for example, USD 2011). Please see the ***AidData Exchange Rates and Deflation Methodology*** for a detailed version of the methods used by AidData. Each Data Product Management Plan will delve into greater details about the application of a specific deflation method for the Product.

2.4.2 Field Standardization and Crosswalk Activities

Field Standardization and Crosswalk Activities are quality assurance processes that standardize fields from collected or derived datasets into AidData standard field names and content, per the AidData Data Field Dictionary and the AidData Naming Convention. For example, a Donor Systems data set might report a disbursement date in the format 'YY-MM-DD', whereas the AidData standard date format is ISO 8601 format ('yyyy-mm-dd'). A Field Standardization Activity will convert that date to the specified format in the final Data Product. Similarly, a Crosswalk Activity transforms a field. However, instead of changing the format of the field, the Crosswalk Activity translates the contents of a field (or fields) into a standard AidData field (as defined in the Data Field Dictionary). This can be thought of as moving a field from one vocabulary (the incoming vocabulary) to the AidData vocabulary. An example of a Crosswalk Activity would take an Aid Information Management System (AIMS) Purpose categorization phrase 'Health Infrastructure' and match that to that AidData Sector Code of '12230 (Basic Health Infrastructure)'.

2.4.3 Deduplication Activities

Deduplication Activities are the application of heuristics to detect duplicate project level records (or project locations for geocoded aid Data Products). While not a common problem for most data sources, these types are important when joining two sets of data, which might have some temporal or functional overlap. The application and particular methodologies of these activities are detailed in the Data Product Management Plan for each product.

2.4.4 Field Consistency Activities

Field Consistency Activities are the checking of fields for internal consistency (against a published schema--or intra-field consistency) and/or against other fields (inter-field consistency). Both inter- and intra-field consistency checks are largely derived from the AidData Data Field Dictionary. Each Data Product Management Plan will detail any and all of these checks performed in the generation of the Product.

2.4.5 Spatial Consistency Activities

Spatial Consistency Activities are the checking of spatial fields for consistency and are generally a subclass of an inter-field consistency checks described above. A typical spatial consistency check would be checking a geocoded project location against a geographic boundary to verify that it appears within the stated boundary (for example, if in dealing with aid to Uganda, and it is noticed that a project location appears outside the national boundaries for Uganda, this would be considered a failed Spatial Consistency check). The application and particular methodologies of these Activities are detailed in the Data Product Management Plan for each product.

2.4.6 Record Verification and Triangulation Activities

Record Verification and Triangulation Activities are Activities which attempt to verify the veracity of a project record or field by checking a secondary source for the same information. The most prototypical AidData Data Product which uses this methodology is the application of the AidData TUFF Methodology for tracking under-reported development finance flows using media and other third-party sources. Each Data Product Management Plan will detail any and all of these checks performed in the generation of any given Product.

2.4.7 Health of Record Activities

Health of Record Activities are Activities that classify and record changes any meaningful information about the collection, quality assurance or production of a Data Product record. In broad terms, the goal is to give a user a reasonable idea about how the data was collected, and what - if any - AidData Activities might have changed the content of the Product record. Each Data Product Management Plan will detail any and all of these checks performed in the generation of any given Product.

3.0 Products and Archive

An AidData Data Product is any data holding that has been released to the public for use. AidData's primary product lines fall into two (2) categories: 1) **Portal Data**, or data which appears on the AidData website(s), or 2) **Static/Snapshot Releases,** or standalone data products for download. Each individual Data Product is accompanied by an attendant Data Product Management Plan. All data products must follow the AidData Naming Convention**.**

3.1 Portal Data Products

Portal Data Products are products displayed and accessible through an AidData web portal of some form. The two Products in this category are the AidData Portal Product and the TUFF Portal Data Product. Please see their respective Data Product Management Plans for details about the each product.

3.2 Static/Snapshot Releases

Static/Snapshot Releases are data products that are available as stand-alone data products via download through the AidData portal. There are three (4) types of these products: 1) **AidData Core Research Releases**; 2) **Non-DAC Research Releases**; 3) **Geocoded Data Research Releases**; and 4) **TUFF Data Research Releases.** Please see the Data Product Management Plan for each of these releases for details about each product and product lifecycle.

3.3 AidData Product Levels

AidData Data products are built using a hierarchical data processing levels, numbering 1-4 (with sub-processing Levels like Level 1a). Each processing Level is built upon the previous processing Level (meaning the Level 1 product is derived from the Level 0 product and so on). Table 1 below lists the processing levels used by AidData in Data Production, with their notional meanings and descriptions. The processing level will be reflected in the name of the data product, as described in the AidData Naming Convention**.**

AidData Data Level	Description (AD)
Level 0	Minimally processed raw data to Level 0 field names and table structure per the product line, converted to UTF-8. This product is for internal use only. **This data will never be part of a public release.**
Level 1	Data post processed to Level 1 field names; Data decomposed to proper table structure by the product definition; financials deflated to base years where possible. Sectors crosswalked to AidData sectors. Geocoded data is spatially scrubbed (verified to be in the correct boundary). Fields checked for intra and inter-field consistency. Level 1 stats generated. QA flags added. This Level may include ancillary data included in the release, but is not quality assured by AidData. **This is the first data product which is available as a public release.**
Level 1A	A single table, denormalized version of the Level 1 product, with provided assumptions. This is a 'joined' product of Level 1 constituent tables or data.
Level 2 (Planned Product)	Level 1 + any quality assured ancillary data (e.g. evaluation data); Aggregates and rollups of data by a data dimension (e.g. by Sector or Donor)
Level 2A (Planned Product)	Rasterizations - Continuous surface representations of our Aid Information. For geocoded data, a geospatial representation of (a) the total dollars of aid we estimate are at a given location, and (b) the number of projects. This product additionally provides at least one surface of the uncertainty in our continuous estimates.
Level 2B (Planned Product)	Simulated Products - These products are further refined versions of the Level 2 product line. For example, a product might produce a continuous surface estimate of aid projects that is weighted to account for slope, population, or road networks. These products will also provide a surface of uncertainty.

Table 1, AidData Data Processing Levels.

3.4 AidData Versioning Schema

AidData supports a strict versioning schema for all MetaData and Static/Snapshot Data Products released (beginning in 3Q 2014). Versions are determined by the following Semantic Versioning flow (Fig. 3). All data produced by AidData goes through this flow. Important actors in this flow are the **AidData Data Team** and the **Data Recommendation Workgroup (DRWG)**. The Data Team is responsible for functional and technical management of the data (production, version control, archive and dissemination). The DRWG is a multidisciplinary internal working group which is charged with vetting the data, and has final approval over the public release of any data product.

So, a typical AidData Product will combine both a processing level and a version according to the full naming convention (e.g. 'UgandaAMP_GeocodedResearchRelease_Level1_v1.2.zip').
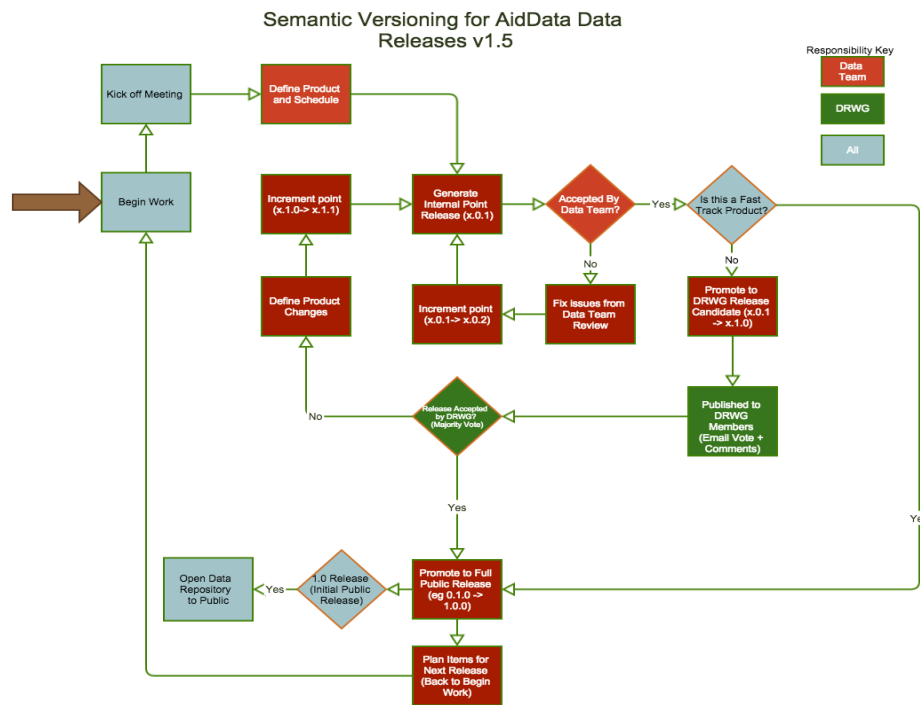
Fig 3., AidData Semantic Versioning Flow

3.5 AidData Archive Activity

Beginning in 3Q 2015, AidData archives all source and final data products into a data archive. Both source documents and final data products have stable uniform resource locators (URLs). In the case of final products, the URL of the product is found in the Data Management plan for the product. In the case of source documents, these URLs are most often located in the data products themselves as a "source" field (where applicable and available).

Donor Systems Data Collection Management Plan

Authors:
Scott Stewart (sstewart@aiddata.org)
Brooke Russell (brussell@aiddata.org)

Document History and Version History

Version	Authors	Date Approved	Approved By	Description
1.0	Scott Stewart, Brooke Russell	10/28/2015	Brad Parks, Co-Executive Director	This is the initial public release of the document.

Purpose

The purpose of the **Donor Systems Data Collection Management Plan** is to address the management of the collection process for Donor Data. This document is related to the Data Management Plan suite of documents as defined in **Figure 1: Data Management Plan Structure** and related documents.

This document is concerned with the **Data Collection Activities (DCA)** pertaining to **Donor Systems—**specifically donors outside reporting organization systems such as the **OECD-CRS.** These donors either do not report to the CRS, or have data collected by AidData directly from the donor. As shorthand, these donors are often referred to as non-DAC donors because most of them are not members of the OECD's Development Assistance Committee. We seek to disambiguate these distinctions by using 3 categories for Donor Systems: 1) **Bilateral Non-DAC Donors;** 2) **Multilateral Non-DAC Donors;** and 3) **Multilateral Improved Donors**

1. Bilateral Non-DAC Donors

AidData systematically engages in data collection for Bilateral Non-DAC Donors. We seek to continually increase the number of donors we have data from as well as the total coverage within each donor—including adding additional years of coverage, additional financing agency coverage, or field coverage.

1.1 Bilateral Non-DAC Donor Selection Criteria

For our core data, AidData focuses on collecting data on any official development finance flows under the following criteria:
1. Bilateral flows from official agencies within the donor government to developing countries or other countries.
2. The flow must have some development-related purpose.
3. Multilateral flows are only included from donor data in cases where the government earmarked flows to multilateral agencies. Core contributions to

multilateral agencies are not currently collected for inclusion in the core data (please see the ***AidData Term Dictionary*** for more details on exclusions).

1.2 Bilateral Non-DAC Donor Prioritization Criteria

In order to ensure the highest impact in its Non-DAC Bilateral Donor Data Collection, AidData prioritizes its DCAs based upon the following considerations and dimensions of the data:

1. **Contractual or Grant Obligations,** meaning whether AidData under an obligation to collect and disseminate the data as part of a contract or grant.
2. **Portfolio/Research Impact,** or how unique, usable, or cornerstone the data is to our users and researcher base.
3. **Ease of Integration**, meaning how difficult it will be to gather, normalize and standardize this data to AidData product types.
4. **Data Quality**, or how assured we can be that the data truly is representative and curated. This includes an assessment of the total number of projects, the number of fields available, and the quality of the descriptive information in the data.
5. F**inancial Information,** meaning that data has financial information such as commitment and disbursement amounts. Data without financial information can still be collected, but data with financial information will generally be prioritized first.
6. **Temporal Coverage**, or the length of time covered by the data source. More recent data is prioritized higher, but increasing AidData's historical coverage is also a priority.
7. **Portfolio Coverage**, or the proportion of the donor's total development assistance portfolio the source provides. AidData is dedicated to increasing the proportion of each donor's development assistance portfolio through collecting data from all relevant financing agencies and mechanisms.
8. **Lack of CRS Coverage**, meaning the donor does not report to the CRS (currently or in the past). Donors with CRS data should likely be considered for data coverage through the Reporting Organizations Systems.

In general these considerations have been presented in order of importance for selecting and prioritizing data collection activities for all Donor Systems, though specific dimensions may take on greater importance with some donors.

1.3 Bilateral Non-DAC Donor Holdings and Collection Methodology

The following table lists all the current and planned holdings for Donor Systems Data Collection. This table is updated in the second quarter (2Q) of each year.

Bilateral Non-DAC Donors are collected using 2 methods:

1) Published Donor Information. Usually this information is available on donor websites or through publications such as Annual Reports, Donor Datasets, or other Donor

Documents. The quality of these sources varies by donor. Donors reporting methods through these types of publications also vary by donor.

2) Direct Contact with Donor. This includes donors where we have established relationships and they have agreed to send us their data directly. Again these sources vary in quality depending on the donor. Table 2 lists the Bilateral Non-DAC Donor systems in which AidData is currently engaged in Data Collection Activities.

Non-DAC Bilateral	Method	Last Update	Planned Update	Notes
Brazil	Direct Contact	2011	--	
Chile	Direct Contact	2010	Q1 2016	
Colombia	Direct Contact and Published Information	2010	Q4 2015	
Cyprus	Published Information	2010	Q1 2016	
Czech Republic	Published Information	2011	Q1 2016	Now reports to the CRS (2013 on)
Estonia	Published Information	2011	Q4 2015	
Hungary	Published Information	2009	Q1 2016	
Iceland	Published Information	2011	Q1 2016	
India	Published Information	2011	Q4 2015	
Kuwait	Published Information	2011	--	KFAED now reports to the CRS
Latvia	Published Information	2011	Q4 2015	
Liechtenstein	Published Information	2011	Q4 2015	
Lithuania	Published Information	2011	Q4 2015	
Monaco	Published Information	2009	Q1 2016	
Poland	Published Information	--	Q4 2015	
Qatar	Published Information	2009	--	No updates are currently available (see TUFF dataset at aiddata.org/donor-datasets)
Romania	Published Information	2011	Q4 2015	
Saudi Arabia	Published Information	2011	Q4 2015	
Slovak Republic	Published Information	2009	Q1 2016	

Non-DAC Bilateral	Method	Last Update	Planned Update	Notes
Slovenia	Published Information	2011	Q1 2016	
South Africa	Published Information	2011	Q1 2016	
Taiwan	Published Information	2011	Q1 2016	
Thailand	Published Information	2011	Q1 2016	
Turkey	Published Information	--	Q4 2015	
United Arab Emirates	Published Information	2011	--	UAE now reports to the CRS

Table 2, Bilateral Non-DAC Donor Collection Schedule

1.4 Bilateral Non-DAC Donor Products and Archive

Table 3 lists the **Data Collection and Source Methodology (DCSM)** documents used for each Non-DAC Bilateral Donor as well as all the data products that include that donor information. The DCSM for each donor is an internal record of all data collection activities for that donor. The DCSMs follow the general AidData *Source Document and Archive Policy*. In relation to data products, Bilateral Non-DAC Donor data has generally been used for the production of Level 1 datasets (Research Releases and Portal Data). Quality Assurance and other Value Addition Activities are detailed in the related Data Product Plan. The key for the Data Products abbreviations are as follows:

- DD0 = Donor Data Level 0 (not published)
- P = Portal Data
- RR1 = Research Release Level 1

The **Data Collection and Source Methodology** documents describe the sources used for each donor as well as the steps taken to align the data to the standardized data structure of the **DD0** level. The data is then prepared for publication through feeding the data through the **Porta**l Data Product Definition procedure. After the data is published to the Portal, the data can be exported and fed into the **RR1** Product Definition procedure.

Non-DAC Bilateral	Data Collection and Source Methodology (DCSM)	Data Products
Brazil	Brazil DCSM	DD0, P, RR1
Chile	Chile DCSM	DD0, P, RR1
Colombia	Colombia DCSM	DD0, P, RR1
Cyprus	Cyprus DCSM	DD0, P, RR1
Czech Republic	Czech DCSM	DD0, P, RR1

Non-DAC Bilateral	Data Collection and Source Methodology (DCSM)	Data Products
Estonia	Estonia DCSM	DD0, P, RR1
Hungary	Hungary DCSM	DD0, P, RR1
Iceland	Iceland DCSM	DD0, P, RR1
India	India DCSM	DD0, P, RR1
Kuwait	Kuwait DCSM	DD0, P, RR1
Latvia	Latvia DCSM	DD0, P, RR1
Liechtenstein	Liechtenstein DCSM	DD0, P, RR1
Lithuania	Lithuania DCSM	DD0, P, RR1
Monaco	Monaco DCSM	DD0, P, RR1
Qatar	Qatar DCSM	DD0, P, RR1
Poland	Poland DCSM	DD0, P, RR1
Romania	Romania DCSM	DD0, P, RR1
Saudi Arabia	Saudi Arabia DCSM	DD0, P, RR1
Slovak Republic	Slovak Republic DCSM	DD0, P, RR1
Slovenia	Slovenia DCSM	DD0, P, RR1
South Africa	South Africa DCSM	DD0, P, RR1
Taiwan	Taiwan DCSM	DD0, P, RR1
Thailand	Thailand DCSM	DD0, P, RR1
Turkey	Turkey DCSM	DD0, P, RR1
United Arab Emirates	UAE DCSM	DD0, P, RR1

Table 3, Non-DAC Bilateral Data Collection and Source Methodology and Products

2. Multilateral Non-DAC Donor Systems

AidData collects information on multilateral flows to developing and other countries from *core* resources. Multilateral agencies are defined as: 1) all agencies on the OECD's Official List of International Organizations; or 2) Organizations that engage in development finance and the organization's board members are representatives of donor governments. Multilateral Non-DAC Donors include those multilateral agencies that do not currently (or in the past) report to the OECD-CRS.

2.1 Multilateral Non-DAC Donor Selection Criteria

For our core data, AidData focuses on collecting data on any official development finance flows from multilateral agencies under the following criteria:

1. Flows from official multilateral agencies using core, unearmarked resources (resources that have been pooled from donor contributions).
2. The flow must have some development-related purpose. We collect ODA flows, OOF Flows, Export Credits and Lines of Credit.

2.2 Multilateral Non-DAC Donor Prioritization Criteria

In order to ensure the highest impact in its Non-DAC Multilateral Donor Data Collection, AidData prioritizes these DCAs based upon the following considerations and dimensions of the data:

1. **Contractual or Grant Obligations,** meaning whether AidData under an obligation to collect and disseminate the data as part of a contract or grant.
2. **Portfolio/Research Impact,** or how unique, usable, or cornerstone the data is to our users and researcher base.
3. **Ease of Integration**, meaning how difficult it will be to gather, normalize and standardize this data to AidData product types.
4. **Data Quality**, or how assured we can be that the data truly is representative and curated. This includes an assessment of the total number of projects, the number of fields available, and the quality of the descriptive information in the data.
5. **Financial Information,** meaning that data has financial information such as commitment and disbursement amounts. Data without financial information can still be collected, but data with financial information will generally be prioritized first.
6. **Temporal Coverage**, or the length of time covered by the data source. More recent data is prioritized higher, but increasing AidData's historical coverage is also a priority.
7. **Portfolio Coverage**, or the proportion of the donor's total development assistance portfolio the source provides. AidData is dedicated to increasing the proportion of each donor's development assistance portfolio through collecting data from all relevant financing agencies and mechanisms.
8. **Lack of CRS Coverage**, meaning the donor does not report to the CRS (currently or in the past). Donors with CRS data should likely be considered for data coverage through the Reporting Organizations Systems or the Multilateral Improved Donor category.

In general these considerations have been presented in order of importance for selecting and prioritizing data collection activities for all Donor System, though specific dimensions may take on greater importance with some donors.

2.3 Multilateral Non-DAC Donor Holdings and Collection Methodology

The following table lists all the current and planned holdings for Multilateral Non-DAC Donor Data Collection. This table is updated in the first quarter (1Q) of each calendar year.

Multilateral Non-DAC Donors are collected using information publicly published by the donor. Usually this information comes in the form of Annual Reports or an online database. The quality of these sources varies by donor. Donors reporting methods through these types of publications also vary by donor. Table 4 lists the Multilateral Non-DAC Donor systems in which AidData is currently engaged in Data Collection Activities.

Multilateral Non-DAC Donor	Method	Last Update	Planned Update
African Capacity Building Foundation (ACBF)	Donor Documents	2009	Q2 2016
Andean Development Corporation (CAF)	Annual Report, Donor Documents	2010	Q2 2016
Arab Bank for Economic Development in Africa (BADEA)	Annual Report	2011	Q2 2016
Caribbean Development Bank (CDB)	Annual Report	2011	Q2 2016
Congo Basin Forest Fund (CBFF)	Donor Documents	2011	Q2 2016
Islamic Development Bank (ISDB)	Annual Report	2011	Q2 2016
Multilateral Fund for the Implementation of the Montreal Protocol	Donor Website (World Bank)	2010	Q2 2016
Nigerian Trust Fund (NTF)	Annual Report	2011	Q2 2016
North American Development Bank (NADB)	Annual Report	2011	Q2 2016
United Nations Democracy Fund (UNDEF)	Annual Report	2011	Q2 2016

Table 4, Multilateral Non-DAC Donor Collection Schedule

1.4 Multilateral Non-DAC Donor Products and Archive

Table 5 lists the **Data Collection and Source Methodology (DCSM)** documents used for each Non-DAC Multilateral Donor as well as all the data products that include that donor information. The DCSM for each donor is an internal record of all data collection activities for that donor. The DCSMs follow the general AidData *Source Document and Archive Policy*. In relation to data products, Multilateral Non-DAC Donor data has generally been used for the production of Level 1 datasets (Research Releases and Portal Data). Quality Assurance and other Value Addition Activities are detailed in the related Data Product Plan. The key for the Data Products abbreviations are as follows:

- DD0 = Donor Data Level 0 (not published)
- RR1 = Research Release Level 1
- P = Portal Data

Multilateral Non-DAC Donor	Data Collection and Source Methodology	Data Products
Caribbean Development Bank (CDB)	CDB DCSM	DD0, RR1, P
Islamic Development Bank (ISDB)	ISDB DCSM	DD0, RR1, P
Nigerian Trust Fund (NTF)	NTF DCSM	DD0, RR1, P
North American Development Bank (NADB)	NADB DCSM	DD0, RR1, P
United Nations Democracy Fund (UNDEF)	UNDEF DCSM	DD0, RR1, P
Andean Development Corporation (CAF)	CAF DCSM	DD0, RR1, P
Arab Bank for Economic Development in Africa (BADEA)	BADEA DCSM	DD0, RR1, P
Multilateral Fund for the Implementation of the Montreal Protocol	Montreal Protocol DCSM	DD0, RR1, P
African Capacity Building Foundation (ACBF)	ACBF DCSM	DD0, RR1, P
Congo Basin Forest Fund (CBFF)	CBFF DCSM	DD0, RR1, P

Table 5, Non-DAC Multilateral Source Collection and Data Archive and Products

3. Multilateral Improved Donor Systems

AidData sometimes collects multilateral donor data directly from donor sources even if the donor reports to the OECD-DAC and CRS. We have done this in cases where there is a clear benefit to expending the resources to collect the data outside the CRS. Many of our Non-DAC Multilateral donors now report to the CRS, in which case we must continually evaluate the usability of sub-planting the OECD data with the data collected direct from donor sources going forward.

3.1 Multilateral Improved Donor Selection Criteria

For our core data, AidData considers collecting improved data for multilateral donors that report to the CRS under the following criteria:

1. The multilateral donor reports to the CRS,
2. There is a clear deficit in the CRS data (either in structure, coverage, or content), and
3. There is official data available from the donor that addresses this deficit

3.2 Multilateral Improved Donor Prioritization Criteria

In order to ensure the highest impact in its Multilateral Improved Donor Data Collection, AidData prioritizes these DCAs based upon the following considerations and dimensions of the data:

1. **Contractual or Grant Obligations,** meaning is AidData under an obligation to collect and disseminate the data as part of a contract or grant
2. **Portfolio/Research Impact,** or how unique, usable, or cornerstone the data is to our users and researcher base.
3. **Ease of Integration**, meaning how difficult it will be to gather, normalize and standardize this data to AidData product types.
4. **Data Quality**, or how assured we can be that the data truly is representative and curated. This includes an assessment of the total number of projects, the number of fields available, and the quality of the descriptive information in the data.
5. **Financial Information,** meaning that data has financial information such as commitment and disbursement amounts. Data without financial information can still be collected, but data with financial information will generally be prioritized first.
6. **Temporal Coverage**, or the length of time covered by the data source. More recent data is prioritized higher, but increasing AidData's historical coverage is also a priority.
7. **Portfolio Coverage**, or the proportion of the donor's total development assistance portfolio the source provides. AidData is dedicated to increasing the proportion of each donor's development assistance portfolio through collecting data from all relevant financing agencies and mechanisms.

In general these considerations have been presented in order of importance for selecting and prioritizing data collection activities for all Donor System, though specific dimensions may take on greater importance with some donors.

3.3 Multilateral Improved Donor Holdings and Collection Methodology

The following table lists all the current and planned holdings for Multilateral Improved Donor Data Collection. This table is updated in the third quarter (3Q) of each year.

Multilateral Improved Donors are collected using information publicly published by the donor. Usually this information comes in the form of Annual Reports or an online database. The quality of these sources varies by donor. Donors reporting methods through these types of publications also vary by donor. Table 6 lists the Multilateral Improved Donor systems for which AidData is currently engaged in Data Collection Activities. AidData will continually evaluate the usefulness of using donor-provided data over data available through the CRS.

Multilateral Improved Donors	Method	Last Update	Planned Update
African Development Bank (AFDB)	Annual Report	2011	Q2 2016 (subject to evaluation)
African Development Fund (AFDF)	Annual Report	2011	Q2 2016 (subject to evaluation)

Multilateral Improved Donors	Method	Last Update	Planned Update
Arab Fund for Economic & Social Development (AFESD)	Annual Report	2011	Q2 2016 (subject to evaluation)
Asian Development Bank (ASDB)	Annual Report	2011	Q2 2016 (subject to evaluation)
Asian Development Fund (ASDF)	Annual Report	2011	Q2 2016 (subject to evaluation)
European Bank for Reconstruction & Development (EBRD)	Annual Report	2010	Q2 2016 (subject to evaluation)
Inter-American Development Bank (IADB)	Annual Report	2011	Q2 2016 (subject to evaluation)
International Fund for Agricultural Development (IFAD)	Annual Report	2010	Q2 2016 (subject to evaluation)
Nordic Development Fund (NDF)	Annual Report	2009	Q2 2016 (subject to evaluation)
OPEC Fund for International Development (OFID)	Annual Report	2011	Q2 2016 (subject to evaluation)
Global Environment Facility (GEF)	Donor Documents	2011	Q2 2016 (subject to evaluation)
Global Fund to Fight Aids, Tuberculosis and Malaria (GFATM)	Donor Website	2011	Q2 2016 (subject to evaluation)
World Bank (All Associated Funds)	Donor Website	2011	Q2 2016 (subject to evaluation)
International Monetary Fund (IMF)	Donor Website	2011	Q2 2016 (subject to evaluation)

Table 6, Multilateral Improved Donor Collection Schedule

3.4 Multilateral Improved Donor Products and Archive

Table 7 lists the **Data Collection and Source Methodology (DCSM)** for each Multilateral Improved Donor as well as all the data products that include that donor information. The DCSM for each donor is an internal record of all data collection activities for that donor. The DCSMs follow the general AidData *Source Document and Archive Policy*. In relation to data products, Multilateral Non-DAC Donor data has generally been used for the production of Level 1 datasets (Research Releases and Portal Data). Quality Assurance and other Value Addition Activities are detailed in the related Data Product Plan. The key for the Data Products abbreviations are as follows:

- DD0 = Donor Data Level 0 (not published)
- RR1 = Research Release Level 1
- P = Portal Data

Multilateral Improved Donors	Data Collection and Source Methodology	Data Products
African Development Bank (AFDB)	AFDB DCSM	DD0, RR1, P
African Development Fund (AFDF)	AFDF DCSM	DD0, RR1, P
Arab Fund for Economic & Social Development (AFESD)	AFESD DCSM	DD0, RR1, P
Asian Development Bank (ASDB)	ASDB DCSM	DD0, RR1, P
Asian Development Fund (ASDF)	ASDF DCSM	DD0, RR1, P
European Bank for Reconstruction & Development (EBRD)	EBRD DCSM	DD0, RR1, P
Inter-American Development Bank (IADB)	IADB DCSM	DD0, RR1, P
International Fund for Agricultural Development (IFAD)	IFAD DCSM	DD0, RR1, P
Nordic Development Fund (NDF)	NDF DCSM	DD0, RR1, P
OPEC Fund for International Development (OFID)	OFID DCSM	DD0, RR1, P
Global Environment Facility (GEF)	GEF DCSM	DD0, RR1, P
Global Fund to Fight Aids, Tuberculosis and Malaria (GFATM)	GFATM DCSM	DD0, RR1, P
World Bank (All Associated Funds)	World Bank DCSM	DD0, RR1, P
International Monetary Fund (IMF)	IMF DCSM	DD0, RR1, P

Table 7, Multilateral Improved Source Collection and Data Archive and Products

Recipient Systems Data Collection Management Plan

Authors:
Scott Stewart (sstewart@aiddata.org)
Brooke Russell (brussell@aiddata.org)

Document History and Version History

Version	Authors	Date Approved	Approved By	Description
1.0	Scott Stewart, Brooke Russell	10/28/2015	Brad Parks, Co-Executive Director	This is the initial public release of the document.

Purpose

The purpose of the **Recipient Systems Data Collection Management Plan** is to address the management of the collection process for Recipient Data. This document is related to the Data Management Plan suite of documents as defined in **Figure 1: Data Management Plan Structure and related documents**.

This document is concerned with the **Data Collection Activities (DCA)** pertaining to **Recipient Systems.** These systems are divided into 4 categories: 1) **Country Aid Information Management Systems (AIMS);** 2) **AIMS Aggregators;** 3) **Country Budget Systems**; and 4) **Budget Aggregators.**

1. Country AIMS Data Collection Activities

AidData systematically engages in data collection for AIMS. Table 1 lists the types of these AIMS for which AidData collects data:

Type	Description
Aid Management Platform (AMP)	Developed by AidData Partner Development Gateway (DG), The Aid Management Program (AMP) combines online software with institutional strengthening activities to help developing country governments that receive official development assistance (ODA) build capacity for aid information management.
Donor Assistance Database (DAD)	The Development Assistance Database (DAD) is an Aid Information Management System (AIMS) developed by Synergy International Systems, Inc for Aid Management, Public Investment and

	National Budgeting.
Custom Recipient Systems	Systems not fitting into these two categories; often developed by in-country parties.

Table 1, AIMS Types

1.1 AIMS Selection Criteria

In order to ensure the highest impact in its AIMS Data Collection, AidData prioritizes these DCAs based upon the following dimensions of the data:

1. **Portfolio/Research Impact,** or how unique, usable, or cornerstone the data is to our users and researcher base.
2. **Data Quality**, or how assured we can be that the data truly is representative and curated. This includes an assessment of the total number of projects, the total number with geocodable project information and an assessment of AidData data model concurrence and variable coverage.
3. **Temporal Coverage**, or the length of time covered by project start dates
4. **Currency**, or how recent the data has been updated, and how likely it is to be updated in the future.
5. **Ease of Integration**, meaning how difficult it will be to gather, normalize and standardize this data to AidData product types.
6. **Contractual or Grant Obligations,** meaning is AidData under an obligation to collect and disseminate the data as part of a contract or grant

Once per quarter, all existing and new Recipient Systems Data Collection Activities are evaluated and prioritized along these dimensions and voted on by the internal **Data Recommendation Working Group (DRWG)**, as part of its charter.

1.2 AIMS Holdings and Collection Methodology

The following table lists all the current and planned holdings for AIMs Data Collection. This table is updated in the first quarter (1Q) of each year.

AIMS are collected by two methods:

1) System Reports. Usually these are available from Aid Management Portal (AMP) systems. System Reports are generally of higher quality because they come directly from the system itself.

2) Portal Mining. This is the systematic scraping of data from a public portal, through publicly available websites, pages and APIs. These are usually Donor Assistance Database (DAD) systems. Table 2 lists the AIMS system that AidData is currently engaged in Data Collection Activities for.

AIMS	Method	Last Update	Planned Update
Uganda AMP	System Report	1Q 2014	--
Nepal AMP	System Report	2Q 2014	--
Senegal AMP	System Report	3Q 2014	--
Timor-Leste AMP	System Report	3Q 2014	--
Honduras AMP	System Report	3Q 2014	--
Haiti AMP	System Report	3Q 2014	--
Malawi AMP	System Report	2Q 2011	--
Nigeria DAD	Portal Mining	2Q 2015	--
Iraq DAD	Portal Mining	2Q 2015	--
Cote D'Ivoire	System Report	--	4Q 2015
Afghanistan DAD	Portal Mining	--	4Q 2015
Colombia DAD	Portal Mining	--	4Q 2015
Somalia DAD	Portal Mining	--	4Q 2015

Table 2, AIMs Collection Schedule

1.3 AIMS Products and Archive

Table 3 lists the **Source Collection and Archive Policy** for each AIMS system, and all the attendant Data Products. AIMS products all use the standard AidData *Source Document and Archive Policy*, which sets guidelines for documenting source information and data collection activities. In terms of data products, AIMS Data has generally been used for the production of Geocoded Level 1 datasets. Quality Assurance and other Value Addition Activities are detailed in the related Data Product Plan. The key for abbreviations for the Data Products are as follows:

- GL0 = AIMS Geocoded Level 0 (Product Definition here)
- GL1 = Geocoded Level 1 (Product Definition here)
- P = Portal Data (Product Definition here)

AIMS	Source Collection & Archive Policy	Data Products
Uganda AMP	Source Document and Archive Standard Policy	GL0, GL1, P
Nepal AMP	Source Document and Archive Standard Policy	GL0, GL1, P
Senegal AMP	Source Document and Archive Standard Policy	GL0, GL1, P

AIMS	Source Collection & Archive Policy	Data Products
Timor-Leste AMP	Source Document and Archive Standard Policy	GL0, GL1, P
Honduras AMP	Source Document and Archive Standard Policy	GL0, GL1, P
Haiti AMP	Source Document and Archive Standard Policy	GL0, GL1, P
Malawi AMP	Source Document and Archive Standard Policy	GL0, GL1, P
Cote D'Ivoire	Source Document and Archive Standard Policy	GL0, GL1, P
Nigeria DAD	Source Document and Archive Standard Policy	GL0, GL1, P
Iraq DAD	Source Document and Archive Standard Policy	GL0, GL1, P
Afghanistan DAD	Source Document and Archive Standard Policy	GL0, GL1, P

Table 3, AIMS Source Archive and Products

2. AIMs Aggregators Data Collection Activities

AidData does not currently engage in gathering data from any secondary publishers or aggregators or AIMS data. If this occurs, this section of the Data Collection Plan will be updated.

3. Country Budget Systems Data Collection Activities

Beginning in 2015, AidData will engage in Data Collection Activities for Country Budget Systems, as detailed below.

3.1 Country Budget Systems Selection Criteria

As with AIMS systems, AidData prioritizes these DCAs based upon the following dimensions of the data:

1. **Temporal Coverage**, or the length of time covered by project start dates
2. **Currency**, or how recent the data has been updated, and how likely it is to be updated in the future.
3. **Data Quality**, or how assured we can be that the data truly is representative and curated. This includes an assessment of the total number of projects, the total number with geocodable project information and an assessment of AidData data model concurrence and variable coverage.
4. **Ease of Integration**, meaning how difficult it will be to gather, normalize and standardize this data to AidData product types.
5. **Portfolio/Research Impact,** or how unique, usable, or cornerstone the data is to our users and researcher base.
6. **Contractual or Grant Obligations,** meaning is AidData under an obligation to collect and disseminate the data as part of a contract or grant

3.2 Country Budget Systems Products and Archives

The following table lists all the current and planned holdings for Country Budget Systems Data Collection. This table is updated in the first quarter (1Q) of each calendar year.

Budget Portal	Method	Last Update	Planned Update
Uganda Budget	Portal Mining	--	3Q 2015

Table 4, Country Budget Collection Schedule

3.3 Country Budget Systems Holdings and Collection Methodology

Table 5 lists the **Source Collection and Archive Policy** for each Budget system, and all the attendant Data Products. All Country Budget Systems use the standard Budget Source Standard Policy, which sets guidelines for documenting source information and data collection activities. In terms of data products, Country Budget Data will generally be used in the production of Geocoded Level 1 datasets. Quality Assurance and other Value Addition Activities are detailed in the related Data Product Plan. The key for abbreviations for the Data Products are as follows:

- GL0 = AIMS Geocoded Level 0 (Product Definition here)
- GL1 = Geocoded Level 1 (Product Definition here)

Budget Portal	Source Data Archive Policy	Data Products
Uganda Budget	Budget Source Standard Policy	GL0, GL1

Table 5, AIMS Source Archive and Products

4. Budget Aggregators Data Collection Activities

AidData does not currently engage in gathering data from any secondary publishers or aggregators or budget data. If this occurs, this section of the Recipient Systems Data Collection Plan will be updated.

Reporting Organization System Data Collection Management Plan

Authors:
Scott Stewart (sstewart@aiddata.org)
Brooke Russell (brussell@aiddata.org)

Document History and Version History

Version	Authors	Date Approved	Approved By	Description
1.0	Scott Stewart, Brooke Russell	10/28/2015	Brad Parks, Co-Executive Director	This is the initial public release of the document.

Purpose

The purpose of the **Reporting Organization System Collection Management Plan** is to address the management of the collection process for Donor Data. This document is related to the Data Management Plan suite of documents as defined in **Figure 1: Data Management Plan Structure and related documents**.

This document is concerned with the **Data Collection Activities (DCA)** pertaining to **Reporting Organization Systems--**system which report data from the donor or reporting organization side, and might act as aggregators for such data. AidData collects project level data from one Reporting Organization System, the OECD Creditor Reporting System (CRS). These donors are often referred to as 'DAC Donors'--from their traditional participation in the OECD's Development Assistance Committee (DAC).

1. OECD CRS

AidData systematically engages in data collection of the OECD CRS data. We update this data from OECD CRS based upon the schedule set forth in the **Holding and Collection Methodology** section of this document.

1.1 OECD CRS Selection Criteria

Being the primary reporting organization system for DAC, AidData has chosen the OECD CRS to be a 'base' data collection activity, upon which other data collection activities are judged for relative merit for data collection. For example, while the World Bank provides project level data to the CRS, AidData supplements this data directly from the World Bank data holdings to increase the quality of the project record. For more details on this process and for which donors this is done, please see the Portal Data Product Definition and the Core Research Release Product Definition.

1.2 OECD CRS Prioritization Criteria

As a fundamental data collection activity, AidData prioritizes the yearly update to the OECD CRS records as the highest priority data collection activity. The CRS data is then used as an important input into the other priority data collection activities, as well as updates to two primary products lines: Portal Data and Core Data Research Releases. For example, if a traditionally non-DAC country begins reporting to the CRS, AidData will evaluate both the Donor Systems Data Collection Management Plan for that country as well as how we represent that shift in our Portal Data and Research Release products. Regardless of that evaluation, all data for all donors reported through the CRS is collected and maintained. This will allow us to keep a full copy of the CRS data for each year. The full CRS data is imported into the portal, and users can choose to search only CRS records in the advanced search function.

1.3 OECD CRS Holdings and Collection Methodology

AidData has OECD CRS data dating back until 1947.   This data has been systematically collected from CDs provided by the OECD (up to and including the 2009 CRS release). From 2010 to present this data is gathered from the OECD Website Bulk download, made available in December of any given year for the prior year's data (e.g. in December 2015, the 2014 OECD CRS will be made available). AidData updates its CRS holdings in the 2Q of each year following the update being made available from the OECD. See Table 2 for the Update Schedule for OECD CRS data.

Years	Method	OECD Update	AidData Planned Update	Notes
1947-2009	OECD CD	2009	--	
2010-present	OECD Website Bulk Download	December of each year	On or before 2Q of following year	Data will be made available on the Portal Data product first, then the Research Release products

Table 2, OECD CRS Update Schedule

AidData employees an accretive model for collecting OECD CRS data. The OECD may revisit previous year's data releases to update its data holdings. AidData however, only imports the new year's records and leaves the previous year's updates unchanged. Over time, this produces a delta between the OECD CRS holdings and AidData representation of same time period. AidData has carefully weighed the benefits and drawbacks of this model and *decided that the accretive model **provides richer data** for the user of our data.* The benefits and drawbacks are:

- Benefits of the Accretive Model
  - Keeps AidData project-level value additions (like Activity Coding) from year to year. We have observed that there is historic difficulty with the stability of the OECD's stable project IDs from one year's release to the next, which makes rejoining our value added data very difficult.
  - Keeps OECD data that may have been deleted as the result of donors moving from non-DAC to DAC. (e.g. Previous Aid Flows). For example, Official Aid

(OA) flows that were committed and disbursed between 1990 and 2004 have been deleted from the official OECD release. AidData has isolated these deleted flows and incorporated them back into its core database.

- Drawbacks to the Accretive Model
  - ○ AidData core holdings will not represent the same data as the OECD CRS portal
  - ○ AidData may miss some project-level changes that represent quality improvements over the data that was imported into the AidData core holdings.

Another important difference in AidData CRS holdings is that traditionally AidData has only imported CRS records with commitment amounts for the project records. This was done because of previous quality and coverage concerns for disbursement data. As the OECD CRS disbursement data has improved, AidData will begin importing CRS records with disbursement only records in 2015 (for the 2013 OECD CRS dataset). To see how this decision affects which data was imported, please see Table 3.

Record Type	Dispensation	Notes
Commitment Only	Imported for all years (1947-present)	
Disbursement Only	Imported from 2013-present	
Commitments and Disbursement	Imported for all years (1947-present)	OECD CRS records have both a Commitment and Disbursement amount attached to the record. if the commitment amount is > 0, the record was imported regardless if there was a disbursement amount.

Table 3, OECD CRS Commitment and Disbursement Import Criteria

1.4 OECD CRS Products and Archive

Table 4 lists the OECD CRS **Source Collection and Archive Policy**, and all the attendant Data Products. In general, OECD CRS Data is used for the production of Level 1 datasets (Research Releases and Portal Data). Quality Assurance and other Value Addition Activities are detailed in the related Data Product Plan. The key for the Data Products are as follows:

- RR1 = Research Release Level 1 (Product Definition here)
- P = Portal Data (Product Definition here)

Source	Source Collection and Data Archive Policy	Data Products
OECD CRS	OECD CRS Collection and Archive Policy	RR1, P

Table 4, OECD CRS Collection and Data Archive and Products

Pure Aggregates Data Collection Management Plan

Authors:
Scott Stewart (sstewart@aiddata.org)
Brooke Russell (brussell@aiddata.org)

Document History and Version History

Version	Authors	Date Approved	Approved By	Description
1.0	Scott Stewart, Brooke Russell	10/28/2015	Brad Parks, Co-Executive Director	This is the initial public release of the document.

Purpose

The purpose of the **Pure Aggregates Data Collection Management Plan** is to address the management of the collection process for pure aggregate data. This document is related to the Data Management Plan suite of documents as defined in **Figure 1: Data Management Plan Structure and related documents**.

This document is concerned with the **Data Collection Activities (DCA)** pertaining to **Pure Aggregates--**such as aggregate data on **remittances**, **FDI flows**, and **US private foundation flow** information. This information comes from various sources and seeks to capture the total resource envelope of development finance. As the name indicates, the details of these flows are only available at the aggregate level (e.g country-year level) and are separated between inflows (flows to the recipient country) and outflows (flows from the donor country).

1. Pure Aggregate Flows

With the AidData 3.0 portal, AidData currently engages in data collection for three types of aggregate development finance data: 1) remittance data, 2) FDI flows, and 3) US private foundation flows. This data is not stored/available at the project level, but rather contains a single composite value at the country-year level (i.e. 1 remittance inflow amount to Honduras in 2008); this contrasts with aggregating project-level into a single aggregate figure (i.e. adding up all project amounts from the U.S. to Honduras in 2008). Where appropriate, we will routinely update this data and search for additional aggregate data that would help portray a more comprehensive view of development finance flows.

1.1 Pure Aggregate Selection Criteria

For inclusion on the AidData 3.0 portal, AidData focuses on collecting available data aggregate development finance flows under the following criteria:

1. The type of flow represents a development-related flow so that its inclusion in AidData will portray a more comprehensive view of development finance flows for our users.

2. Aggregate data is available at the country-year level.
3. No comprehensive project level data is available with the same coverage.

1.2 Pure Aggregate Prioritization Criteria

In order to ensure the highest impact in its Pure Aggregate Data Collection, AidData prioritizes these DCAs based upon the following considerations and dimensions of the data:

1. **Contractual or Grant Obligations,** meaning is AidData under an obligation to collect and disseminate the data as part of a contract or grant
2. **Portfolio/Research Impact,** or how unique, usable, or cornerstone the data is to our users and researcher base.
3. **Ease of Integration**, meaning how difficult it will be to gather, normalize and standardize this data to AidData product types.
4. **Data Quality**, or how assured we can be that the data truly is representative and curated. This includes an assessment of the total number of projects, the number of fields available, and the quality of the descriptive information in the data.
5. **Financial Information,** meaning that data has financial information such as commitment and disbursement amounts. Data without financial information can still be collected, but data with financial information will generally be prioritized first.
6. **Temporal Coverage**, or the length of time covered by the data source. More recent data is prioritized higher, but increasing AidData's historical coverage is also a priority.

In general these considerations have been presented in order of importance for selecting and prioritizing data collection activities for all Donor System, though specific dimensions may take on greater importance with some donors.

1.3 Pure Aggregates Holdings and Collection Methodology

The following table lists all the current and planned holdings for Pure Aggregates Data Collection. This table is updated in the third quarter (3Q) of each year.

Pure Aggregate data is collected and published using the Pure Aggregates Manager (PAM). PAM is an internal tool built by AidData to upload and publish pure aggregate data using spreadsheets downloaded from each relevant source website (and transformed as described in the relevant Data Collection and Source Methodology document). Table 2 lists the Pure Aggregates for which AidData is currently engaged in Data Collection Activities.

Pure Aggregates	Method	Last Update	Planned Update
Remittance inflows	PAM	Q3 2015	Q2 2016
Remittance outflows	PAM	Q3 2015	Q2 2016
FDI inflows	PAM	Q3 2015	Q2 2016

Pure Aggregates	Method	Last Update	Planned Update
FDI outflows	PAM	Q3 2015	Q2 2016
US private foundations	PAM	Q3 2015	Q2 2016

Table 2, Pure Aggregates Collection Schedule

1.4 Pure Aggregates Products and Archive

Table 3 lists each type of Pure Aggregates along with the **Data Collection and Source Methodologies**, and the entire attendant Data Products. In general, Pure Aggregates data has only been used for the production of Level 1 datasets (specifically Portal Data). Quality Assurance and other Value Addition Activities are detailed in the related DCSM. The key for the Data Products are as follows:

- PAL0 = Pure Aggregates Level 0 (not published)
- P = Portal Data (Portal Product Definition)

Pure Aggregates	Data Collection and Source Methodology Archive	Data Products
Remittance inflows	Remittance Data Collection and Source Methodology	PAL0, P
Remittance outflows	Remittance Data Collection and Source Methodology	PAL0, P
FDI inflows	FDI Data Collection and Source Methodology	PAL0, P
FDI outflows	FDI Data Collection and Source Methodology	PAL0, P
US private foundations	US Foundation Data Collection and Source Methodology	PAL0, P

Table 3, Pure Aggregates Data Collection and Source Methodology Archive and Products

Portal Data Product Definition

Authors:
Scott Stewart (sstewart@aiddata.org)
Brooke Russell (brussell@aiddata.org)

Document History and Version History

Version	Authors	Date Approved	Approved By	Description
1.0	Scott Stewart, Brooke Russell	10/28/2015	Brad Parks, Co-Executive Director	This is the initial public release of the document.

Purpose

The purpose of the **Portal Data Product Management Plan** is to address production of portal data, defining its sources, production, importation and archive. This document is related to the Data Management Plan suite of documents as defined in **Figure 1: Data Management Plan Structure and related documents**.

This document is concerned with detailing the **Data Source Methodology**, **Value Addition Activities (VA)** and **Products and Archives (PA)** pertaining to the three types of portal data: 1) Core Data, 2) Geospatial Data, and 3) Pure Aggregate Data.

1. Core Data
Core Data consists of the official project-level development finance data collected from donor systems or reporting systems. All core project-level data should be systematically incorporated into the Portal. As an organization, AidData seeks to collect all official project-level development finance data from all donors wherever possible. The Portal should be the first place core data is published (except for stand-alone donor datasets).

1.1 Core Data Source Methodology
Portal core data can come from two types of sources: 1) **Donor Systems** data in the form of DD0 datasets and 2) **Reporting Systems** data (for example, OECD CRS data).

1.1.1 Donor Systems Data

1.1.1.1 DD0 Level 0 Datasets: For data coming from donor systems (non-CRS data), the source format should be a level 0 dataset of the original donor data (DD0). Level 0 are maintained only in AidData's drive data repository for that dataset. There is no external component or public release of this data. The entire purpose of Level 0 is to decompose the product inputs into standard table structures and column/field names. Financials are not deflated or converted for Level 0 products, but rather left in the currency and year that they are reported in the source data. For each DD0 dataset there will be a **Data Collection and Source Methodology (DCSM)** document. DD0 datasets should be produced consistently throughout the year as new data is

made available by donors (see Scheduled Data Collection plans in the Donor Systems Data Management Plan). Sometimes multiple sources can be used for one donor, which requires implementing the **Best of Breed** model.

1.1.1.2 Best of Breed: AidData seeks to provide the most comprehensive view of development finance for each donor in both breadth and depth of information, and we have adopted a **Best of Breed** methodology to do so. Some donors have several sources of data we could use, and over different time spans the quality may vary from source to source. As such, AidData's **Best of Breed** model allows us to choose the best source for a donor on a donor-year unit of analysis. Examples include donors that historically have not reported to the CRS but have recently began reporting their data to the CRS. Best of Breed allows us to report the historical data directly from the UAE and then starting with 2012 data, switch to using CRS data for UAE. In a few cases, Best of Breed is also implemented on a donor-financing agency-year unit of analysis.

1.1.1.3 Implementing Best of Breed

Before implementing Best of Breed for any donor, the two (or more) proposed sources need to undergo a thorough analysis to determine the following:

1. **Determine the best source for each year**: The 'best' source should balance coverage, depth of information, and structure. Parameters to consider for coverage include how many financing agencies the source covers, what types of aid might be included or excluded (ODA, OOF, Humanitarian Aid, etc), and the number of fields available from the source. Depth of information includes the quality of descriptions (especially long descriptions). Lastly, structure should also be considered (including how easily the data can be incorporated to AidData's portal and research release data structures).
2. **Analyze any potential overlap:** AidData seeks to reduce duplication and double counting from any sources. One potential problem with a Best of Breed model implemented on a donor-year basis is that data around the years where the source is changed could double-count the same flows. As such, before a source is changed for any specific year, the data must be analyzed to ensure there is no double counting.

1.1.2 Reporting Systems data

1.1.2.1 CRS Data & Accretive Model: For core data coming from Reporting systems (the CRS), the source format should be the raw CRS data. Usually this data will only include the most recent year of data released by the CRS. For details on this data and the **Accretive model** AidData uses, please see the Reporting Organization System Data Collection Management Plan.

1.2. Value Addition Activities

There are five primary value addition activities undertaken in the production of Portal Data: quality assurance and standardization, aligning data structures, currency conversion and deflation, applying geocodes, and assigning activity/purpose codes.

1.2.1 Quality Assurance/Standardization Activities

Quality Assurance/Standardization Value Addition Activities are listed in Table 1 below. Each of the sources for Portal Core Data may have one or more of these applied to it according to requirements and necessity (for example, some data sets might not need to have a text translation performed). For details on these activities for each donor within the Donor Systems, please see their associated Data Collection and Source Methodology documents.

Activity	Abbreviation	Required	Description
Text Translation	TT	Yes	Translation of text to English (usually done through Google Translate)
Deduplication	DD	Yes	Identification (and removal) of possible duplicate records
Removal of Core Multilateral Contributions	MC	Yes	Remove any rows that relate to core contributions to multilateral organizations (to avoid double counting)
Sector/Classification Crosswalk	SC	Yes	Translation of native data systems sectors to AidData vocabulary (will be superseded if a project is assigned AidData activity/purpose codes)
Donor/Recipient/Organization Crosswalk	OC	Yes	Crosswalk (or lookup) of organization information
Date Standardization	DS	Yes	Where necessary, standardize date format to yyyy/mm/dd (including assigning the 1st day of the month for unspecified start dates and last day of the month for unspecified end dates)
Record Linking	RL	Yes	If a project is a multi-year project and has several commitments or disbursements in separate years, those rows should be linked using a Donor Project ID
Schema Validation	SV	Yes	Field-level validation of field contents per the *AidData Data Field Dictionary*
Location Standardization	LS	No	Crosswalk of location data in the native data set to AidData standards.

Table 1, Quality Assurance Value Addition Activities

1.2.2. Aligning Data Structures

Data structures for both the donor systems data and the recipient data structures need to be preserved in a way that AidData can easily transform them into the two structures we use: hierarchical structure and flat-table structure. Generally this means collecting the data in a flat-table structure using donor-recipient-year-project unit of analysis (so multi-year projects are split

over multiple rows), but recording or generating a Donor Project ID that will link the rows together. When importing the data, this Donor Project ID will be used to generate a project view and individual transactions associated with the project.

1.2.3 Currency Conversion and Deflation

Currency conversion and deflation is taken care of in the process of publishing the data on the Portal (through Internal). The process for conversion and deflation is available in the AidData Exchange Rates and Deflation Methodology document. The rates for conversion and deflation can be managed through Internal using the Currency and Deflators manager.

1.2.4 Applying Geocodes

Applying geocodes is a potential value addition activity for core data, but it is currently not prioritized. The main data that is geocoded from the core dataset is World Bank 'Mapping 4 Results' data and some AfDB data (since they have solid project ids). Additional funding would likely be required to prioritize geocoding Core Data.

1.2.5 Assigning Activity/Purpose Codes

Activity coding, AidData's most unique coding operation, is part of a broader effort to create a searchable database that disaggregates aid according to what development finance is actually funding. Coders assign as many activity codes to a given record as needed to capture the full scope of the project and one purpose code that identifies the dominant sector for those activities. Activity coding is double-blind, meaning two researchers code each description independently, and then the codes are compared. Discrepancies across coders are arbitrated by an experienced coder.

The process is managed by AidData's coding manager who prioritizes donors and years to be coded on internal and assigns batches of work to coders in his or her team. Workflows are currently set up donor by donor for the CRS. In Q4 of 2015 we will be able to set up workspaces according to remaining gaps in coverage following the application of the code rules in internal. The immediate goal is to complete AidData's original commitment to its user community for database coverage for activity coding through 2008. Once AidData meets that target, we will reevaluate to identify the next order of priority. Agenda setting for the coming year will likely prioritize the optimization of the auto-coder and improving/sustaining coverage for the SDGs.

1.3 Data Products and Data Archive

1.3.1 Products

Core Portal data is itself a Level 1 Data Product. It is our main Data Product for Core Data. It is also dynamic--meaning it is continually changing with new sources and new activity codes/geocodes being added constantly. As such, the portal data itself is not versioned over specific time periods. The current portal data is referred to as "AidData 3.0", to indicate the change to a Mongo database structure that occurred in 2013.

The main downstream product of the Core Portal Data is the core Research Release (RR). For details on its production and versioning, please see the Core Research Release Product Definition.

Product	Relation	Required	Publish Location
Core Research Release (RR)	Downstream	Yes, at specified intervals	AidData Portal/Archive

1.3.1 Level 1 Product Definition - Core Portal Data

Level 1 are the first public datasets made available. They are published in (3) ways: 1) as static research releases, 2) as a sibling International Aid Transparency Initiative (IATI) product to the IATI registry (this is not a mandatory product, and may not be produced if AidData deems the data to already be contained within registry), and 3) to the main AidData portal. The sibling Level 1 products are defined in sections following the primary (static research release) Level 1 definition. Each product goes through several **Value Addition (VA) Activities**.

In relation to the Portal, Level 1 core Portal Data is prepared for importation into the Portal. Once it is imported into the Portal, it will be published to the AidData website and constitutes a public release of the data. The entire purpose of Level 1 is to prepare the data for publication alongside the existing core Portal data. Financials are deflated and converted to standard currency/base year. If possible, all data should have at least a sectoral classification.

1.3.1.1 Level 1 Product Contents

Table 2 below illustrates the minimum Level 1 product contents. Level 1 products are kept under the "products" folder of the related GitHub repository.

File	Location	Required	Description
README.pdf	\	Yes	A README document describing the dataset and license and attribution information.
projects.csv	\products\Level_1	Yes	The projects file as a comma separated file. Please see the AidData Data Dictionary for field names
locations.csv	\products\Level_1	No	Project locations as a comma separated file. Please see the AidData Data Dictionary for field names
transactions.csv	\products\Level_0	Yes	The transactions.csv as a comma separated file. Please see the AidData Data Dictionary for field names

Table 2, Level 1 Product Contents

1.3.1.2 Level 1 Product Versioning and Review Process

All Level 1 core data ready for importation into the Portal is reviewed for release by the **Data Recommendation Working Group.** Per the Data Recommendation Working Group (DRWG) charter, a majority vote is required to release any Level 1 product for public release. Figure 2 below details the semantic versioning process flow for Level 1 (and beyond) data releases.
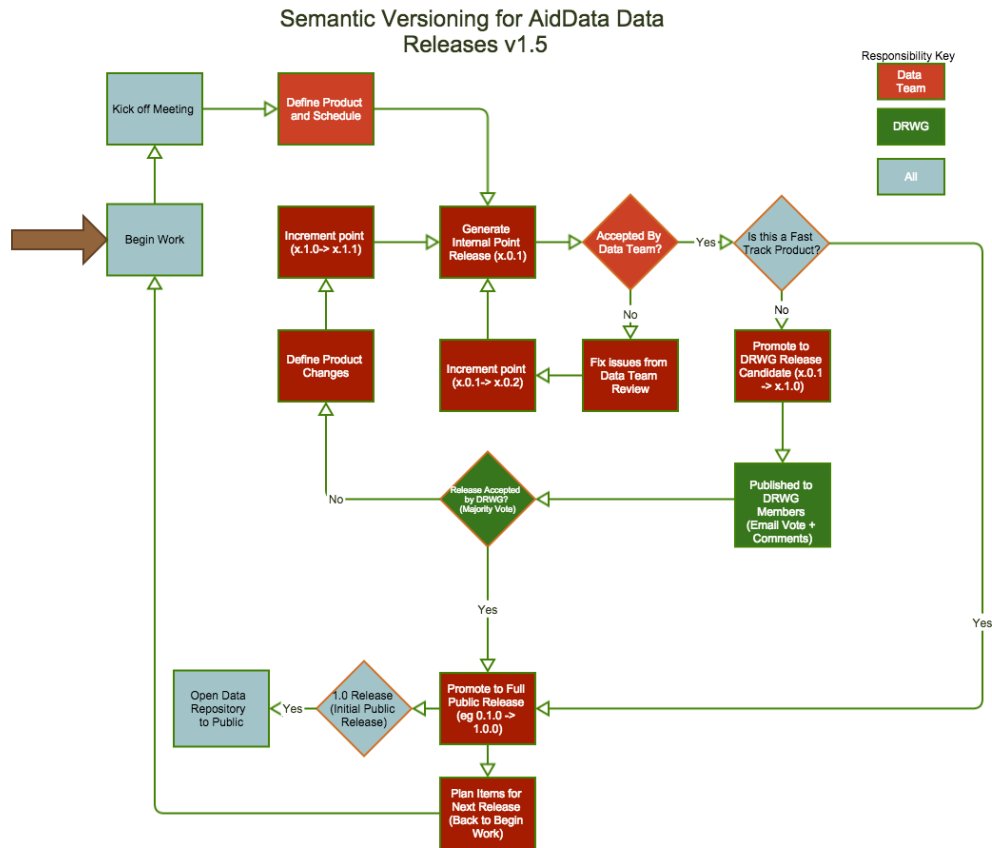

Figure 2, Versioning and Review Process

1.3.2 Level 1 Final Product Archive

For these products, AidData only keep the final and current products are hosted on the AidData portal. As new versions are generated, older versions are archived to the AidData GitHub Dataset Repository, as binary releases (zip files) only. Please contact data@aiddata.org if you require access to this repository.

2. GeoSpacial Data

As recipient systems data is geocoded and released by the data team, the research releases are posted on the Portal's website for 'Donor Datasets'. Currently these datasets are not regularly pushed through the Portal publishing and advanced search side of the website. However, there are a few select donors that are searchable on the Portal at aiddata.org/maps.

3. Pure Aggregate Data

In addition to the project-level data provided by Core and Geospatial Data, the portal also holds aggregate data for flows that are not available at the project-level. Currently this data includes FDI inflows and outflows, Remittance inflows and outflows, US private foundation data, and aggregate (from the project-level) international aid flows. See the ***Pure Aggregates Data Collection Management Plan*** for more details. As an organization, AidData seeks to collect relevant aggregate flows when available. The Portal is the only place to publish such data.

3.1 Pure Aggregate Source Methodology
Portal Pure Aggregate data comes from 2 types of sources: 1) PAL0 Dataset as collected and transformed as described in the relevant DCSM. The data for FDI, remittances, and US private foundations come through PAL0s. 2) Aggregated Core Flows. The data for International Aid Flows are aggregated donor-recipient-year flows from the core data. When a user clicks on an individual aggregate flow, the portal will switch the user into the core data portion of the portal.

3.1.1 PAL0 Dataset:

For pure aggregate data coming from outside sources, the source format should be a level 0 level dataset of the original donor data (PAL0). Level 0 are maintained only in AidData's Google Drive data repository for that dataset. There is no external component or public release of this data. The entire purpose of Level 0 is to decompose the product inputs into standard table structures and column/field names. Financials need to be reported in USD nominal amounts. For each PAL0 dataset there will be a **Data Collection and Source Methodology (DCSM)** document. PAL0 datasets should be produced once a year (likely in Q2).

Once the PAL0 dataset is ready, it needs to be imported to the portal through the Pure Aggregates Manager. The exact structure that the data needs to be in is described in the relevant DCSMs.

3.2 Value Addition Activities
There are three primary value addition activities undertaken in the production of Aggregate Portal Data: 1) Quality assurance and standardization, 2) aligning data structures, and 3) currency deflation.

3.2.1 Quality Assurance/Standardization Activities

Quality Assurance/Standardization Value Addition Activities are listed in Table 1 below. Each of the sources for Portal Aggregate Data may have one or more of these applied to it according to requirements and necessity. For details on these activities for each type of aggregate data, please see their associated Data Collection and Source Methodology (DCSM) document.

Activity	Abbreviation	Required	Description
Origin/Destination Country Crosswalk	CC	Yes	Crosswalk (or lookup) of country information (ISO 2 code is used in final format)
Date Standardization	DS	Yes	Where necessary, standardize year format to "1/1/XXXX" (only for "Year" field).

Activity	Abbreviation	Required	Description
Schema Validation	SV	Yes	Field-level validation of field contents per the AidData Data Dictionary

Table 1, Quality Assurance Value Addition Activities

3.2.2. Aligning Data Structures

The PAL0 dataset needs to be in an origin/destination/year unit of analysis format. For specific details on making this transformation, see the **Data Structure Methodology** document for each pure aggregate data type (Remittances, FDI, US Private Foundations).

3.2.3 Currency Deflation

All source information for the Pure Aggregates should come in USD, so currency conversion is not built into the import module for pure aggregates. Deflation is necessary though, and that is executed inside the backend of the Portal during publishing, so there are no steps to complete the deflation. Details on AidData's deflation methodology are available in the AidData Exchange Rates and Deflation Methodology document.

3.3 Pure Aggregate Data Products and Data Archive

3.3.1 Products

Pure Aggregate Portal data is itself a Level 1 Data Product. The portal is the only data product that through which the pure aggregate data is published--there are no downstream products. The portal data itself is not versioned over specific time periods. The current portal data is referred to as "AidData 3.0" though to indicate the change to a Mongo database structure that occurred in 2013.

3.3.2 Level 1 Product Versioning and Review Process

See section 1.3.1.2.

Core Research Release Product Definition

Authors:
Scott Stewart (sstewart@aiddata.org)
Brooke Russell (brussell@aiddata.org)

Document History and Version History

Version	Authors	Date Approved	Approved By	Description
1.0	Scott Stewart, Brooke Russell	10/28/2015	Brad Parks, Co-Executive Director	This is the initial public release of the document.

Purpose

The purpose of the **Core Research Release Product Definition Plan** is to address production of the Core Research Release, defining its sources, production, presentation and archive. This document is related to the Data Management Plan suite of documents as defined in **Figure 1: Data Management Plan Structure and related documents**.

This document is concerned with the **Value Addition Activities (VA)** and **Products and Archives (PA)** pertaining to the Core Research Release. All core data collected by AidData will be released as Core Research Releases at established intervals. The base of the Research Release is all core data exported from the Portal. The exported data is then transformed and improved with the Value Addition Activities (VA) specified below. As such, the Research Release serves as 1) a time-stamped version of AidData's core data, which allows users and researchers to replicate their research results, and 2) a version of the data that is provided in a flat-table structure, allowing for users and researchers to analyze annual flows.

The Core Research Release Product is a downstream product of the other core data production processes. The data is published through the following steps outlined in Figure 2.
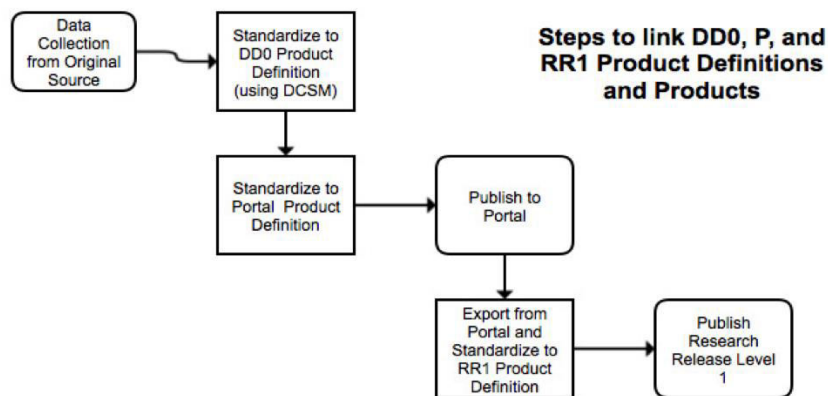
As Figure 2 shows, the first step towards producing a Core Research Release is to export all core data from the Portal. All data should first be imported to the Portal to be included in the Research Release.

1. Value Addition Activities

There are two primary value addition activities undertaken in the production of core data research releases after exporting the portal data: 1) flattening the structure and 2) quality assurance/data standardization.

1.1 Creating a Flat-Table Structure

The first necessary VA activity is to take the hierarchical structure (one project to multiple transactions) and transform it into a flat-table structure where the unit of analysis is donor-recipient-project-year. To do this, use the RR-Transform Code. At its basics, the code will do the following transformations:

- At the transaction level, each type of transaction is currently on a separate row. So move each transaction type and associated amount to a separate column (and currency types to their own separate columns)—so have one column for commitment amount, one column for disbursement amount, one column for repayment amount, and one column for co-financing amount. Concatenate multiple co-financing agencies into one cell (delimited) and co-financing amounts into one cell (delimited). This should leave us with a table that has one row for each year the project had a transaction flow.
- Map project-level information (such as long description, start/end dates, AidData purpose and activity codes, etc) onto the transformed transaction-level rows.
- The final result should create a donor-project-year table.

NOTE: The exporter is currently due to be updated so that there is an option to export the data in a flat file.

1.2 Quality Assurance/Data Standardization

Additional Field Standardization should be carried out on the following fields. Specifics are documented internally. Cofinancing agencies, CRS Amounts, Source, Source Detail, Finance Type, Flow Type, Imputed Flow Type, Type of Assistance, Funding Agency Type, Borrowing Agency Name, Beneficiary Name, Cofinancing Agency Name, Cofinancing Amounts, Loan Term, Grace Period, Number of Repayments, Repayment Type, First Repayment Date, Interest Rate, Grant Element.

2. Data Source, Data Products and Data Archive

2.1 Source Data, Document Archive and Data Repository
Core Research Release Data may come from any data source as collected by a **Data Collection Activity** and imported into the Portal. For each reporting system data product there will be **Data Collection and Source Methodology** (DCSM) and one more data products, attached to specific processing Levels for the product. These products are defined in section 2.2, below.

2.2 Products

Core research releases follow the processing Levels as defined in the AidData Data Management Master Document. As per the standard, the Level 1 product is first public release of the data to the portal. The Core Research Release represents a Level 1 product. This document will be updated to include those data products when the data sets are produced.

2.2.1 Level 1.1 Product Contents - Core Research Release

The final core research release should include the flat table dataset as well as a readme file with the relevant data information. Table 4 below lists the primary data contents of the Level 1.1 dataset.

File	Location	Required	Description
README.pdf		Yes	A README document describing the dataset and license and attribution information and donor-specific details.
projects_flat.csv		Yes	the projects_flat file as a CSV separated file. Please see the AidData Data Field Dictionary for field names

Table 1, Level 1.1 Product Contents

2.2.2 Level 1 Product Versioning and Review Process

All Level 1.1 core research releases are reviewed for release by the **Data Recommendation Working Group.** Per the Data Recommendation Working Group (DRWG) charter, a majority vote is required to release any Level 1.1 product for public release. Figure 3 below details the semantic versioning process flow for Level 1 (and beyond) data releases.
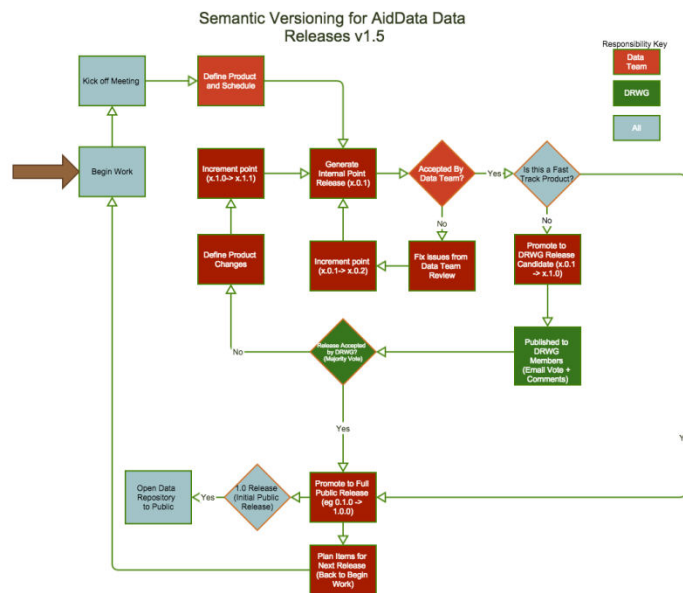

Figure 3, Versioning and Review Process

2.2.3 Level 1 Product Naming Conventions and File Types

Core research releases all follow the AidData Naming Convention for file names. Namely:

[DataSetPreamble]_[DataSetType]_[ProcessingLevel]_[Version].zip
So for example, the Level 1, version 1.1 of the Core Research Release would be named:

AidDataCore_ResearchRelease_Level1_v3.1.zip

All Level 1.1 core research release datasets are zip files, with the contents defined in the previous section.

2.2.4 Level 1.1 Downstream and Sibling Products

There are currently 6 downstream or sibling products from the Core Research Release product.

Product	Relation	Required	Publish Location	Description
Core Research Release - Thin	Sibling	Yes	aiddata.org/aiddata-research-releases	A copy of the Core Research Release with only 6 fields (AidData ID, donor, recipient, year, Purpose Name)
Core Research Release – Since 1996	Sibling	Yes	aiddata.org/aiddata-research-releases	A copy of the Core Research Release with only projects with transactions from 1996 +
Core Research Release – non-OECD	Sibling	Yes	aiddata.org/aiddata-research-releases	A copy of the Core Research Release with only data collected from sources outside the OECD CRS
Core Research Release – Aggregate Donor, Recipient, Year	Sibling	Yes	aiddata.org/aiddata-research-releases	Data from the Core Research Release aggregated to the donor, recipient, year level.
Core Research Release – Aggregate Donor, Recipient, Year, Purpose Name	Sibling	Yes	aiddata.org/aiddata-research-releases	Data from the Core Research Release aggregated to the donor, recipient, year, purpose name level.

Table 4, Level 1.1+ Core Research Sibling and Downstream Products

2.2.5 Level 1 Final Product Archive

For these products, AidData keeps the final current products as well as the older versions hosted on the AidData portal. Research releases are an important product for our research users who need replicable results from our data. All Core Research Releases versions should be available on the Portal at all times (through the Research Release web page).

2.2.6 Level 2 Product Definition

There are currently no formal Level 2 Products. When AidData standardizes its Level 2 processing for Core Research Release Data, this section of the document will be updated.

3. Data Holdings and Production Schedule

Table 6 below defines all the current Core research releases, value additions and sibling products.

RR Version	Release Date	Planned Release Date
Plaid 1.9	March 2010	-
AidData 2.0	November 2011	-
AidData 3.0	-	Q4 2015
AidData 3.1	-	Q1 2016

Table 5, Core Research Release Product

Geocoded Data Research Release Product Management Plan

Authors:
Scott Stewart (sstewart@aiddata.org)
Brooke Russell (brussell@aiddata.org)

Document History and Version History

Version	Authors	Date Approved	Approved By	Description
1.0	Scott Stewart, Brooke Russell	10/28/2015	Brad Parks, Co-Executive Director	This is the initial public release of the document.

Purpose

The purpose of the **Geocoded Data Research Release Product Management Plan** is to address production of geocoded data, defining its sources, production, presentation and archive. This document is related to the Data Management Plan suite of documents as defined in **Figure 1: Data Management Plan Structure and related documents**.

This document is concerned with the **Value Addition Activities (VA)** and **Products and Archives (PA)** pertaining to all geocoded data research releases. All data collected by AidData may be geocoded. AidData does not systematically geocode the entire data holdings, but rather engages in geocoding either through contractual or grant obligations or by strategically engaging in such activities for the benefit of the organization or the user community.

1. Value Addition Activities

There are two primary value addition activities undertaken in the production of geocoded data research releases: 1) geocoding and 2) quality assurance/data standardization.

1.1 Table of Quality Assurance Activities

Quality Assurance Value Addition Activities are listed in Table 1 below. Each of the Data Products may have one or more of these applied to it according to requirements and necessity (for example, some data sets might not need to have a currency exchange calculation performed).

Activity	Abbreviation	Required	Description
Geocoding	GC	Yes	Geocoding project documents per the AidData Geocoding Methodology
Text Translation	TT	Yes	Translation of text to English (or other language)
Sector/Classification Crosswalk	SC	Yes	Translation of native data systems sectors to AidData vocabulary; or translation of AidData vocabulary to

Activity	Abbreviation	Required	Description
			another sector vocabulary (e.g. OECD)
Currency Conversion	CC	Yes	Conversion of currency to from US Dollars (USD)
Financials Deflation	FD	No	Deflation of financials to a base year.
Deduplication	DD	No	Identification (and removal) of possible duplicate records
Location Standardization	LS	No	Crosswalk of location data in the native data set to AidData standards.
Donor/Recipient/Org anization Crosswalk	OC	No	Crosswalk (or lookup) of of organization information
Record Verification/Triangu lation	RV	No	Independent record verification of data from a second source (or third source)
Record Linking	RL	No	Linking of project records to existing records in the core (or other) database
Schema Validation	SV	No	Field-level validation of field contents per AidData Data Field Dictionary

Table 1, Quality Assurance Value Addition Activities

2. Data Source, Data Products and Data Archive

Geocoded Data may come from any data source as collected by a **Data Collection Activity**. For each static geocoded research release product there will be a **Source Data Document and Archive Policy** and one more data products, attached to specific processing Levels for the product. These products are defined in section 2.2, below.

2.1 Source Data, Document Archive and Data Repository

Static geocoded data research releases follow the standard **Source Document and Archive Policy** as defined by each upstream **Data Collection Activity (DCA)** for the product. These documents are kept in the GitHub repository for each data set. Please contact the data@aiddata.org for access to these documents.

2.2 Products

Static geocoded research releases follow the processing Levels as defined in the AidData Data Management Master Document. As per the standard, the Level 1 product is first public release of the data to the portal. There are currently no products beyond Level 1 for static geocoded

research releases. This document will be updated to include those data products when the data sets are produced.

2.2.1 Level 0 Product Definition

Level 0 are maintained only in the GitHub Data repository for that dataset. Within the repository there is a top level structure. There is no external component or public release of this data. The entire purpose of Level 0 is to decompose the product inputs into standard table structures and column/field names. Financials are not deflated or converted for Level 0 products, but rather left in the currency and year that they are reported in the source data. Sectoral (or other classification methods) are converted to YAML, in order to preserve sector hierarchies, if present.

2.2.1.1 Level 0 Product Contents

Table 2 below illustrates the minimum Level 0 product contents. Level 0 products are kept under the "products" folder of the related GitHub repository.

File	Location	Required	Description
projects.csv	\products\Level_0	Yes	The projects file as a Comma separated file. Please see the AidData Data Field Dictionary for field names
locations.csv	\products\Level_0	Yes	Project locations as a Comma separated file. Please see the AidData Data Dictionary for field names
transactions.csv	\products\Level_0	Yes	The transactions.csv as a Comma separated file. Please see the AidData Data Dictionary for field names
projects_ ancillary.csv	\products\Level_0	No	Comma separated ancillary (usually non-QAed) data related to the projects table. Field names may or may not be in the AidData Data Dictionary
locations _ancillary.csv	\products\Level_0	No	Comma separated ancillary (usually non-QAed) data related to the locations table. Field names may or may not be in the AidData Data Field Dictionary
transactions _ancillary.csv	\products\Level_0	No	Comma separated ancillary (usually non-QAed) data related to the transactions table. Field names may or may not be in the AidData Data Field Dictionary

Table 2, Level 0 Product Contents

2.2.2 Level 0 Product Versioning and Review Process

All Level 0 static geocoded research releases are reviewed by the Data Team only. These products are internal products only, and will not be used by the user community until Level 1 processing has occurred (and gone through the Quality Assurance/Standardization for Level 1).

Versioning is maintained in the GitHub data repository, with each new product becoming a label on the repository itself.

2.2.3 Level 0 Product Naming Conventions and File Types

As mentioned, Level 0 are maintained only in the GitHub Data repository for that dataset. There is no external facing component or dataset, and hence no naming convention for files. The repository itself follows the format:

[DataSetPreamble]-geocoded-dataset

All AidData geocoded dataset repositories are housed in the Github organization, and are all currently private repositories.

2.2.4 Level 0 Downstream and Sibling Products

Table 3 below lists downstream and sibling products for the Level 0 data.

Product	Relation	Required	Publish Location
AidData Level 1	Downstream	No	AidData Portal/Archive

Table 3, Level 0 Geocoded Research Sibling and Downstream Products

2.2.6 Level 1 Product Definition

Level 1 are the first public datasets made available. They are published in (3) ways: 1) as static research releases, 2) as a sibling International Aid Transparency Initiative (IATI) product to the IATI registry (this is not a mandatory product, and may not be produced if AidData deems the data to already be contained within registry) and 3) to the main AidData portal. The sibling Level 1 products are defined in sections following the primary (static research release) Level 1 definition. Each product goes through several **Value Addition (VA) Activities**. Please see section 1.1 for the definition of these activities and Section 3 for a list of all data holdings (as well as what VAs were done on each dataset.

2.2.6.1 Level 1 Product Contents

Each dataset can have varying contents. However, there are minimum requirements for any dataset which must be met to be considered a complete (or releasable) data set. Table 4 below lists the minimum contents for a Level 1 Geocoded Data Research Release. Individual datasets might have different data files as part of the release. The datapackage.json file will always serve

as the complete manifest file for the data release. Table 4 below lists the primary data contents of the Level 1 dataset.

File	Location	Required	Description
README.pdf	\	Yes	A README document describing the dataset and license and attribution information.
datapackage.json	\	Yes	an OKFN Data Package file for the dataset: http://dataprotocols.org/data-packages/
projects_ ancillary.csv	\data	Yes	The projects file as a Comma separated file. Please see the AidData Data Dictionary for field names
locations_ ancillary.csv	\data	Yes	Project locations as a Comma separated file. Please see the AidData Data Field Dictionary for field names
transactions_ ancillary.csv	\data	Yes	The transactions.csv as a Comma separated file. Please see the *AidData Data Field Dictionary* for field names
level_1a.csv	\data	Yes	The Level 1A geocoded data table (sibling product) as a Comma separated file. This is a pre-joined version of the projects and locations table suitable for visualizations.

Table 4, Level 1 Product Contents

2.2.7 Level 1 Product Versioning and Review Process

All Level 1 static geocoded research releases are reviewed for release by the **Data Recommendation Working Group.** Per the Data Recommendation Working Group (DRWG) charter, a majority vote is required to release any Level 1 product for public release. Figure 2 below details the semantic versioning process flow for Level 1 (and beyond) data releases.
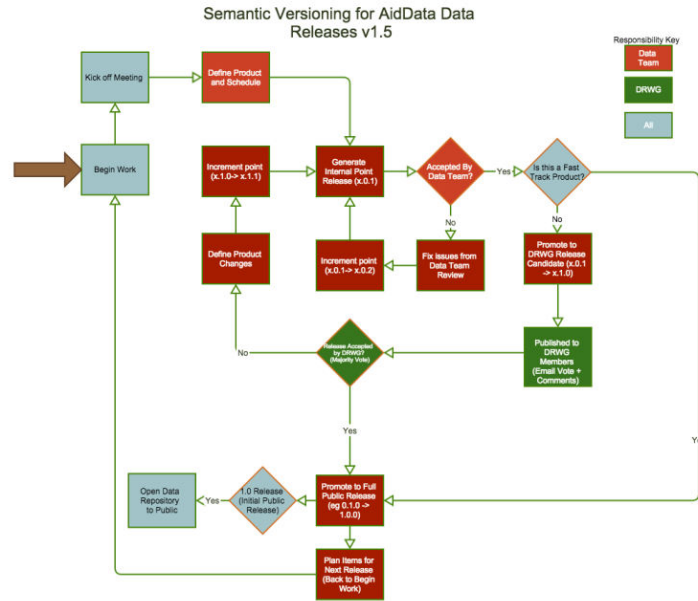
Figure 2, Versioning and Review Process

2.2.8 Level 1 Product Naming Conventions and File Types

Static geocoded research releases all follow the AidData Data Field Dictionary for file names. Namely:

[DataSetPreamble]_[DataSetType]_[ProcessingLevel]_[Version].zip

So for example, the Level 1, version 1.1 of the Uganda Aid Management Platform Geocoded Dataset is named:

UgandaAMP_GeocodedResearchRelease_Level1_v1.1.zip

All Level 1 static geocoded datasets are zip files, with the contents defined in the following section.

2.2.9 Level 1 Downstream and Sibling Products

Static geocoded research public releases (e.g. Level 1) have 3 downstream/sibling Data Products as defined in Table 5. Sibling releases contain all data in the Level 1 data product, but transformed to a new format (e.g. IATI). Downstream products are further transformations of the Level 1 Product. No downstream products currently exist for Level 1 Geocoded Data. This document will be updated as those products are defined.

Product	Relation	Required	Publish Location

IATI Release	Sibling	No	IATI Registry
AidData Portal	Sibling	No	AidData Portal
AidData Level 1A	Sibling	Yes	AidData Portal/Archive
AidData Level 2	Downstream	No	AidData Portal/Archive
AidData Level 2a	Downstream	No	AidData Portal/Archive
AidData Level 2b	Downstream	No	AidData Portal/Archive

Table 5, Level 1 Geocoded Research Sibling and Downstream Products

2.2.9.1 Level 1 Sibling Product - IATI Release

The Level 1 IATI Release is a sibling product to the Level 1 Research Release. With each update to the Level 1 product, an IATI file is also created and registered with the IATI registry. There is no difference in content between the Level 1 Research Release and the IATI file of the same data set. The file is name with the following naming convention:

iati_[IATIVersion]_[DataSetPreamble]_[DataSetType].xml

So for example, IATI 1.04 version of the Uganda Aid Management Platform Geocoded Dataset IATI file is named:

iati_1.04_UgandaAMP_GeocodedResearchRelease.xml

For these products, AidData only keep the final and current products are hosted on the AidData portal. As new versions are generated, older versions are archived to the AidData GitHub Dataset Repository.

2.2.9.2 Level 1 Sibling Product - Level 1A

The Level 1A Geocoded Research Release is a sibling product to the Level 1 Research Release. It is included in the Level 1 Static Research Release Data Product. This file is a pre-joined version of the Projects and Locations tables of the release, specifically for quick visualizations and data journalism.

2.2.10 Level 1 Final Product Archive

For these products, AidData only keeps the final and current products are hosted on the AidData portal. As new versions are generated, older versions are archived to the AidData GitHub Dataset Repository, as binary releases (zip files) only. Please contact data@aiddata.org if you require access to this repository.

2.2.11 Level 2 Product Definition

There are currently no formal Level 2 Products. When AidData standardizes its Level 2 processing for Geocoded Data, this section of the document will be updated.

3. Data Holdings and Products

All Geocoded Research Releases are available via the AidData Portal at the following URL:

http://aiddata.org/geocoded-datasets

IATI Product Definition

Authors:
Scott Stewart (sstewart@aiddata.org)
Brooke Russell (brussell@aiddata.org)

Document History and Version History

Version	Authors	Date Approved	Approved By	Description
1.0	Scott Stewart, Brooke Russell	10/28/2015	Brad Parks, Co-Executive Director	This is the initial public release of the document.

Purpose

The purpose of the **IATI Product Management Plan** is to address production of IATI data, defining its sources, production, presentation and archive. This document is related to the Data Management Plan suite of documents as defined in **Figure 1: Data Management Plan Structure and related documents**.
.
This document is concerned with the **Value Addition Activities (VA)** and **Products and Archives (PA)** pertaining to all IATI data research releases. Generally speaking IATI products are "companion" data products to other AidData products, containing the same information as its companion data product, but differing only in format and presentation (namely as an IATI xml file).

All data collected by AidData may be produced as IATI data. AidData does not systematically produce IATI data for its entire data holdings, but rather engages in production either through contractual or grant obligations or by strategically engaging in such activities for the benefit of the organization or the user community. AidData also takes into consideration the uniqueness of the dataset to the IATI/development community. To that end, there may be some datasets that AidData produces IATI data or research releases for that are not published to the IATI registry (if for example, the data is already available on the IATI registry, AidData will not duplicate the records). These decisions are made by the Data Manager on a case by case basis.

1. Value Addition Activities

There are three primary value addition activities undertaken in the production of IATI data research releases: 1) activity coding, 2) geocoding, and 3) quality assurance/data standardization.

1.1 Table of Value Addition Activities

Quality Assurance Value Addition Activities are listed in Table 1 below. Each of the IATI Data Products may have one or more of these applied to it according to requirements and necessity (for example, some data sets might not need to have a currency exchange calculation performed).

Activity	Abbreviation	Required	Description
Geocoding	GC	No	Geocoding project documents per the AidData Geocoding Methodology
Activity Coding	AC	No	Activity Coding project data per the AidData Activity Coding Methodology
IATI Product Generation	IP	Yes	Generation of the Dataset as an IATI xml file.
Text Translation	TT	Yes	Translation of text to English (or other language), if needed
Sector/Classification Crosswalk	SC	YES	Translation of native data systems sectors to AidData vocabulary; or translation of AidData vocabulary to another sector vocabulary (e.g. OECD)
Currency Conversion	CC	No	Conversion of currency to US Dollars (USD)
Financials Deflation	FD	Yes	Deflation of financials to a base year.
Deduplication	DD	No	Identification (and removal) of possible duplicate records
Location Standardization	LS	No	Crosswalk of location data in the native data set to AidData standards.
Donor/Recipient/ Organization Crosswalk	OC	No	Crosswalk (or lookup) of of organization information
Record Verification/ Triangulation	RV	No	Independent record verification of data from a second source (or third source)
Record Linking	RL	No	Linking of project records to existing records in the core (or other) database

Activity	Abbreviation	Required	Description
Schema Validation	SV	Yes	Field-level validation of field contents per the *AidData Data Field Dictionary*; validation of IATI files against IATI schemas

Table 1, Quality Assurance Value Addition Activities

2. Data Source, Data Products and Data Archive

As IATI Product Data may come from any data product produced by AidData, as defined by its individual product plan (and associated methods for data collection, quality assurance and source archival).

2.1 Source Data, Document Archive and Data Repository

IATI data products follow the standard *Source Document and Archive Policy* as defined by its companion data product and upstream **Data Collection Activity (DCA)** for the companion product. These documents are kept in the GitHub repository for each data set. Please contact the data@aiddata.org for access to these documents.

2.2 Products

IATI products releases follow the processing Levels as defined in the *AidData Data Management Master Document*, and per their companion data product. Generally, IATI products are companion products to Level 1 products. There are currently no IATI products for data beyond Level 1.

2.2.1 Level 0 Product Definition

Level 0 are maintained only in the GitHub Data repository for that datasets' companion upstream Level 0. There is no separate level 0 for the IATI product. Please see the Level 0 definition of the companion product for definitions of contents and fields.

2.2.6 Level 1 Product Definition

Level 1 are the first public datasets made available which are eligible to have IATI sibling/companion products. The definition of the product is dictated through the companion product (e.g. a geocoded research release) and the IATI standard itself. AidData currently produces only IATI 1.04 compliant datasets.

2.2.6.1 Level 1 Product Contents

Each IATI dataset is a single IATI xml file. The file itself will always be distributed with its companion research release, alongside the tabular data. As mentioned in the preamble, the Data Manager may also decide to publish the file itself to the AidData section of the IATI registry, based upon the uniqueness of the dataset to the registry, or the addition of specific value to existing data which make it useful to the broader community.

2.2.7 IATI Product Versioning and Review Process

All IATI releases are reviewed for release by the **Data Recommendation Working Group.** Per the Data Recommendation Working Group (DRWG) charter, a majority vote is required to release any IATI product for public release. Versions of IATI products are defined by their companion data product (e.g. Level 1, v 1.1). Figure 2 below details the semantic versioning process flow for Level 1 (and beyond) data releases.
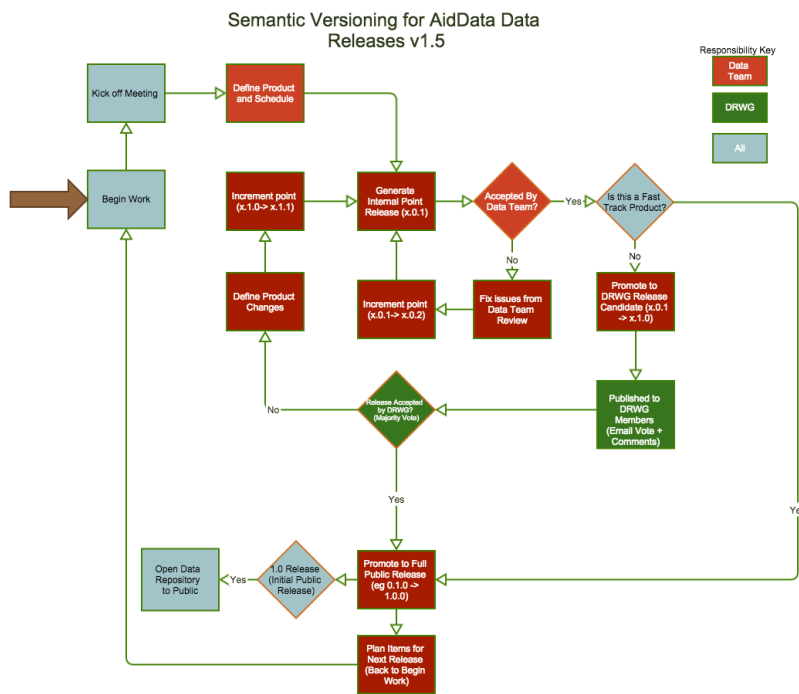

Figure 2, Versioning and Review Process

2.2.8 IATI Product Naming Conventions and File Types

IATI products all follow the *AidData Naming Convention* for file names. Namely:

iati_[IATIVersion]_[DataSetPreamble]_[DataSetType].xml

So for example, the 1.04 IATI file for the Level 1 Senegal AMP Geocoded Research Release Dataset is named:

iati_1.04_Senegal_AMP_Geocoded_Research_Release.xml

Generally, these are included in the research release package, under the /data folder, but please refer the appropriate product plan for further details.

2.2.9 IATI Product Downstream and Sibling Products

IATI data products have 3 downstream/sibling Data Products as defined in Table 2. Sibling releases contain all data in the Level 1 data product, but transformed to a new format (e.g. IATI). Downstream products are further transformations of the Level 1 Product. No downstream products currently exist for Level 1 Geocoded Data. This document will be updated as those products are defined.

Product	Relation	Required	Publish Location
IATI Release	Sibling	No	IATI Registry
AidData Portal	Sibling	No	AidData Portal
AidData Level 1A	Sibling	Yes	AidData Portal/Archive
AidData Level 2	Sibling/Downstream	No	AidData Portal/Archive
AidData Level 2a	Sibling/Downstream	No	AidData Portal/Archive
AidData Level 2b	Sibling/Downstream	No	AidData Portal/Archive

Table 2, IATI Product Sibling and Downstream Products

2.2.9.1 IATI Sibling Product - Level 1 Product

The IATI product can be a sibling product to a Level 1 Research Release. With each update to the Level 1 product, an IATI file might also be created and registered with the IATI registry, at

the discretion of the Data Manager. There is no difference in content between the Level 1 Research Release and the IATI file of the same data set.

For these products, AidData only keep the final and current products are hosted on the AidData portal. As new versions are generated, older versions are archived to the AidData GitHub Dataset Repository.

2.2.9.2 IATI Sibling Product - Level 1A

A Level 1A Research Release product might also be a sibling product to an IATI Product when the IATI product is also a sibling of the related Level 1 product. These products would contain the same data as the companion Level 1 and Level 1A products.

2.2.10 IATI Product Archive

For these products, AidData only keeps the final and current products are hosted on the AidData portal. As new versions are generated, older versions are archived to the AidData GitHub Dataset Repository, as binary releases (zip files) only. Please contact data@aiddata.org if you require access to this repository.

2.2.11 Level 2 Product Definition

There are currently no formal Level 2 Products. When AidData standardizes its Level 2 processing data products, this section of the document will be updated, if there are attendant Level 2 IATI products.

3. Data Holdings and Products

All IATI Products are available via the AidData Portal at the following URL and (if published to the registry) at the AidData IATI registry:

http://aiddata.org/geocoded-datasets

and

http://dashboard.iatistandard.org/publisher/aiddata.html

AidData Data Policy Documents

Source Document and Archive Policy
AidData Naming Convention
AidData Exchange Rates and Deflators

The AidData Data Policy Documents provide an overview of standardized procedures used in all data products and operations.

Source Document and Archive Policy

Authors:
Scott Stewart (sstewart@aiddata.org)
Brooke Russell (brussell@aiddata.org)

Document History and Version History

Version	Authors	Date Approved	Approved By	Description
1.0	Scott Stewart, Brooke Russell	10/28/2015	Brad Parks, Co-Executive Director	This is the initial public release of the document.

Purpose

The purpose of the **Source Document and Data Archive Policy** is to clearly define standards which govern the archive of source documents and data products. This document is a member of the AidData Data Policy Documents family of documents and is related to the Data Management Plan as described in **Figure 1: Data Management Plan Structure and related documents**.

1. Source Document Definition

Simply, source document collection is the 'cradle' of the cradle to grave data management. Source Documents are any document for which AidData creates downstream data products. The Documents themselves may be tabular data or project documents (PDFs). AidData keeps these documents for the purposes of replicability of datasets, and to use them to correct any issues with the data that might be identified at a later time. This data is considered "raw" and must be kept as close to the original representation of the data as possible at the time of collection as possible. Below is a list of some of the sources for which Source Documents are collected and maintained.

Source	Definition & Notes
OECD CRS	Yearly txt files from the OECD
Non-DAC/ Donor Collected	Data gathered by the Non-DAC data collection efforts, or possibly web scraped data from a donor website or API (e.g. World Bank)
TUFF	Phase I Factiva articles and links and Phase II web searches
AIMs	Exports from the AIMs system itself, possible web scraped
Other	Any data gathered for the purposes of data production generation

2. Archive Policy

All source documents for all data products must be kept for a minimum of 3 years. When beginning a new data product, the responsible party must provide all source documents to the Data Team for archive. The Data Team will then create a source document folder in the Source Document Repository for the data. Large external datasets (satellite data or raster data) will handled on a case by case basis, but usually will not be archived in this manner.

2.1 Stable URLS

For project level data, a stable URL to the source document for each source project record will be created and attached to each project record. This URL should also be used if the data is imported into AidData internal. No data will be imported into internal without stable URLs to individual source documents.

3.0 Exceptions to this policy

This policy may be overridden if a separate (and accepted) source document and policy is created and accepted for any new data product as part of its data management plan.

AidData Naming Convention

Authors:
Scott Stewart (sstewart@aiddata.org)
Brooke Russell (brussell@aiddata.org)

Document History and Version History

Version	Authors	Date Approved	Approved By	Description
1.0	Scott Stewart, Brooke Russell	10/28/2015	Brad Parks, Co-Executive Director	This is the initial public release of the document.

Purpose

The purpose of the **AidData Naming Convention** is to clearly define standards which govern the naming of fields and datasets. This document is a member of the AidData Data Dictionaries family of documents as defined in **Figure 1: Data Management Plan Structure and related documents**.

1. Filename and Composition Standards for Datasets

This document seeks to provide a guide to naming files for use by internal and external users of our data.

All datasets produced by AidData should be fit the following form:

[DataSetPreamble]_[DataSetType]_[ProcessingLevel]_[Version].zip

Where each bracketed section is defined in the following table, and each section is separated by an underscore ("_"). Each dataset set should be compressed using a standard zip compression.

Section	Definition & Notes
DataSetPreamble	The is the user-defined or data set specific preamble used to describe the dataset. It can be as descriptive as necessary, and may not contain underscores (e.g. UgandaAMP). The preamble is title-cased.
DataSetType	This section defines what type of dataset the file is. There are currently 2 types of data sets: **ResearchRelease** and **GeocodedResearchRelease.** Additional types will be added here.
ProcessingLevel	Processing levels are one of the processing levels defined in the Master Data Management Plan, and title cased (e.g. Level0, Level1, Level2a)
Version	Version string as "v[SemanticVersion]m where SemanticVersion is defined by the Semantic Version of the release

Example: The Uganda Aid Management Platform Level 1 Geocoded Research Release, version 1.2 would be named:

DataSetPreamble = UgandaAMP
DataSetType = GeocodedResearchRelease
ProcessingLevel = Level1
Version = v1.2

or,

UgandaAMP_GeocodedResearchRelease_Level1_v1.2.zip

When unzipped, each dataset should unzip into a folder with the same name as the zip file (e.g. UgandaAMP_GeocodedResearchRelease_Level1_v1.2). At the root level of the folder there should be Data Package JSON manifest file for the dataset, which describes the contents of the dataset.

2. Field Naming Standards

Fields in AidData Data Products should:

1. Be lower case (e.g. aid_flows) and;
2. Individual components of a name must either be separated by an underscore ("_"), and;
3. The field component separator must not be mixed within fields in the dataset and;
4. Fields definitions from the AidData Data Field Dictionary must be used first, before the creation of a new field.

Examples:

LEGAL:
project_id
donor_year

ILLEGAL:
AidData Project
nominal usd_committments
project id

AidData Exchange Rates and Deflation Methodology

Authors:
Scott Stewart (sstewart@aiddata.org)
Brooke Russell (brussell@aiddata.org)

Document History and Version History

Version	Authors	Date Approved	Approved By	Description
1.0	Scott Stewart, Brooke Russell	10/28/2015	Brad Parks, Co-Executive Director	This is the initial public release of the document.

Purpose

The purpose of the **AidData Exchange Rates and Deflation Methodology** is to clearly explain the exchange rate and deflation sources, formulas, and standards which govern any currency exchange and deflation calculations undertaken by AidData in its data products. This document is a member of the AidData Data Dictionaries family of documents as defined in **Figure 1: Data Management Plan Structure and related documents**.

1 Currency Conversion and Deflation Purpose

Financial values collected as part of AidData's Data Collection Activities (DCAs) must be converted and deflated so that they are comparable across currencies and years. AidData's methodology follows after the OECD, whose documentation is available here:
http://www.oecd.org/dac/aidstatistics/informationnoteonthedacdeflators.htm.

The full methodology involves two steps: 1) Exchange Rates and 2) Deflation.

2 Exchange Rates

2.1 Exchange Rate Methodology

Before deflation, all values must first be expressed in nominal (current) U.S. dollars (USD). This is done with an LCU per USD exchange rate, applied by:

(original value) / (LCU per USD) = (new value)

For example:

100 EU / .7 = 142.57 USD

2.2 Exchange Rate Sources

OECD countries: OECD –
http://stats.oecd.org/Index.aspx?DatasetCode=SNA_TABLE4

Taiwan: Central Bank of the Republic of China (Taiwan) --
http://www.cbc.gov.tw/content.asp?CuItem=1879

All others: World Bank -- http://data.worldbank.org/indicator/PA.NUS.FCRF/countries

3 Deflators

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

> *Inflation * Change in Buying Power = Deflator*

3.1 Inflation

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

3.1.1 Inflation Methodology

Inflation is measured as relative to a given base year. AidData 2.0 used the base year of 2011. Percentages were then generated using the following formula:

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

For example, in 2011, Colombia's GDP inflation was 4.2%. Taking 2011 as the base year, the percentage for 2011 is 100%. So, to calculate the percentage for 2008, using 2011 as the start year:

> $$\underset{P_{year}}{100} \;=\; P_{previous} + (P_{previous} * \underset{I_{year}}{.04})$$

This yields 96% as Colombia's percentage for 2008. (Decimals have been rounded for this example, but were not rounded for AidData's deflator table.) In 2008, Colombia's GDP inflation was 8%. Then, to calculate 2007, 2008 is the start year:

> $$\underset{P_{year}}{96} \;=\; P_{previous} + (P_{previous} * \underset{I_{year}}{.08})$$

This yields 89% as Colombia's percentage for 2007.

3.1.2 Inflation Sources

The following sources are used to compile the inflation values.
> **OECD countries:** OECD -- http://stats.oecd.org/Index.aspx?QueryId=27220
> **Taiwan:** Council for Economic Planning and Development --
> > http://www.cepd.gov.tw/encontent/m1.aspx?sNo=0001453

Kuwait: For some years, World Bank data was not available. For 1990 to 1990, AidData used World Bank "High income: non OECD" GDP inflation and for 2008 to 2010, AidData used Kuwait's CPI inflation from the World Economic Outlook.
All others: World Bank GDP Inflation --
http://data.worldbank.org/indicator/NY.GDP.DEFL.KD.ZG

3.2 Change in Buying Power

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

3.2.1 Buying Power Methodology

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

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

For example, the Korean Won to USD rate was 1273.9 in 2011 and 804.4 in 1996. The subsequent change in buying power is:

$$1273.9 / 804.4 = 1.58$$

Note that this methodology yields a ratio of 1 for all currencies pegged to the USD.

3.2.2 Buying Power Sources

The data used for the buying power formula are generated from the historical exchange rates described above. See the Source section 2.2.

3.3 Finalized Deflators

3.3.1 Deflators Methodology

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

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

3.3.2 Deflators Sources

These figures are built on inflation and change in buying power, described above. See the source section there.

4 Quality Assurance Checks

To be sure that the methodology, sources and implementation were sound, AidData used DAC countries as reference. AidData generated deflators for DAC countries using World Bank data and then compared those deflators to the ones provide by OECD. The results of those comparisons are found in sheet "E2 (OECD-AidData % diff)". This sheet is also linked, so changes to other sheets will change this sheet and clicking on a cell will reveal the formula used to generate the value.

The average percent difference between AidData-generated deflator and OECD-provided deflator is less than .05%. Because the process was so highly accurate for DAC countries, it is assumed that the process is also accurate for non-DAC countries.

5 Examples on Using GDP Deflators

Amounts in LCU should be converted to nominal USD, using the LCU per USD exchange rates found in sheet A1. Then, the values should be *divided by* the percentages in sheet "E1". For example, in 1975, Kuwait funded an electrification project in Bangladesh worth 6,400,000 *KD1975* (AidData ID 2427051). To convert this amount to USD 2011, first, convert it to USD 1975:

$$6{,}400{,}000 \;_{KD\ 1975} / (.29003 \;_{KD/USD\ 1975}) = 22{,}066{,}505.30 \;_{USD\ 1975}$$

Next, divide it by the AidData deflator:

$$22{,}066{,}505.30 \;_{USD\ 1975} / 20.83\% = 105{,}936{,}175.20 \;_{USD\ 2011}$$

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

AidData Dictionaries

AidData Data Field Dictionary
AidData Term Dictionary

The AidData Dictionaries provide customized definitions and explanations for terms common to AidData products and operations.

AidData Data Field Dictionary

Authors:
Scott Stewart (sstewart@aiddata.org)
Brooke Russell (brussell@aiddata.org)

Document History and Version History

Version	Authors	Date Approved	Approved By	Description
1.0	Scott Stewart, Brooke Russell	10/28/2015	Brad Parks, Co-Executive Director	This is the initial public release of the document.

Purpose

The purpose of the **AidData Data Field Dictionary** is to clearly define the data fields across all the AidData data products. This document is related to the Data Management Plan suite of documents as defined in **Figure 1: Data Management Plan Structure and related documents**.

This document seeks to provide a comprehensive list of definitions for the data fields used throughout the various AidData Data Products. Whenever possible, AidData seeks to align our definitions of fields and terms with the definitions used by the OECD-CRS, IATI, and the general development sphere.

Product Key:

PE = Portal Export
L1G = Level 1 Geocoded Data
L1C = Level 1 Core Data
PAE = Pure Aggregate Export
L1T = Level 1 TUFF Data*

*TUFF fields & definitions are provided in Table 2

Table 1: Data Field Definitions

Field Name	Products	Definition & Notes
AD Sector Codes	PE, L1G	The AidData sector code based on the AD sector name. Pipe delimited.
AD Sector Names	PE, L1G	The AidData sector based on a crosswalk from the AMP-assigned sector. Pipe delimited.
Additional Info	PE, L1C	Any additional information on the project that does not fit elsewhere in the data scheme.
AidData Activity Code/Name	PE, L1C	Identifies the full list of activities involved in each development project. This field is assigned to a project using the AidData Activity Coding Methodology (including a double-blind and arbitration system). Each project can have as many activities assigned as necessary. All activities (codes and names) are provided in the same field in a tab-delimited list.

Field Name	Products	Definition & Notes
AidData ID/Project ID	PE, L1C	Unique identifier for each project/activity published on the AidData Portal.
AidData Purpose Code/Name	PE, L1C	The dominant activity that the project is meant to foster. This field is usually assigned to a project using the AidData Activity Coding Methodology (including a double-blind and arbitration system). Before a project is activity coded though, this field can be generated from the CRS purpose code. Each project can only have one purpose code.
AidData Sector Code/Name	PE, L1C	The AidData Sector (3-digit) that the project is meant to foster. This field is usually generated from the purpose code assigned to the project through the AidData Activity Coding methodology. Before a project is activity coded though, this field can be generated from the CRS purpose code, a donor-provided sector, or the title of the project (as a temporary way to make sure each project has at least a sector classification).
Beneficiary	PE, L1C	The organization or group that is meant to benefit from the project. This field is not normalized.
Borrower	PE,L1C	For loans this is the organization actually borrowing money. If the donor does not explicitly list a borrower, this field is left blank to avoid incorrect assumptions.
Cofinancers	PE,L1C	Organizations (or countries) that are also providing funding to the specified project. This field is only entered when explicitly provided by the donor.
Commitment Amount	PE,L1C	The loan or grant amount for a project or program as given by the donor. Conceptually, we use the same definition as the OECD-DAC: "A firm obligation, expressed in writing and backed by the necessary funds, undertaken by an official donor to provide specified assistance to a recipient country or a multilateral organisation. Bilateral commitments are recorded in the full amount of expected transfer, irrespective of the time required for the completion of disbursements. " If the donor reports the amount in both USD and another currency we enter the original currency amount (to standardize our currency exchange rates).
Commitment Date	PE,L1C	This date is pulled first from the date of board approval. When such is unavailable we will use the date of loan signature or grant approval.
Contact/Role Of Contact	PE,L1C	Name and role of the person associated with a project as provided by the donor. We do not gather email or phone contacts at this time.
CRS Markers	PE, L1C	This is a set of fields that the CRS codes to indicate whether a project focuses on specific topics, including biodiversity, climate, desertification, or the environment.
Destination	PAE	The destination country for the specified aggregate flow from all other countries. For example, if the destination is Nigeria for FDI, then that row represents the total FDI flow into Nigeria from all other countries.
Disbursement Amount	PE,L1C	The sum of all disbursements for a given project per year. As we are trying to adhere to the one line per project paradigm as much as possible we do not enter separate disbursements as separate rows but rather enter cumulative disbursements on each project.
Donor ISO3	PE, L1C, L1G	Donor's ISO-3 code. With geocoded data, this can have multiple ISO3 codes when there are multiple donors listed, in which case the list will be pipe delimited.
Donor Project Id	PE,L1C	This is the provided donor's ID for tracking the project. We seek to capture meaningful IDs where available (meaning they are used internally by the donor to identify a specific project). If not available, this field may report a database ID if provided by the donor. This field is used to identify multi-year projects, so in some cases (marked clearly in the readme files) we generate a unique donor project ID when a donor did not supply one. The general format should be "[Donor ISO]-[Recipient ISO]-[Generated Number]". For example, a project from Romania to Moldova, the generated donor project ID could be ROU-MDA-101.
Donor Secondary Project Id	PE, L1C	A secondary project ID used by the donor or the original source. For core data where the source is the OECD CRS, this field is the CRSID as reported in the original CRS data.
Donor/Funder	PE,L1C,	Donor name. For bilaterals this is the name of the sovereign government (i.e. France,

Field Name	Products	Definition & Notes
	L1G	Saudi Arabia). For multilaterals this gets slightly more confusing – it is the name of the donor, and soft/hard lending windows may be treated as separate donors (i.e. IDA & IBRD are listed as separate donors, AFDB & AFDF are listed as separate donors). We attempt to follow CRS practice in this manner. With the geocoded datasets, this field can have multiple donor names, which are pipe delimited.
End Actual Type	PE, L1G	Type of the project end (example: end-actual)
End Date (Actual)	PE,L1C, L1G	Date project ended. A full date is required for this field. However, donor data will often report incomplete dates (such as month-year combos or only the year). In that case, the default date will be the the last day of the specified month or year. (i.e. if the documentation says the project ended in December 2011, the start date in the Portal would be 2011-12-31). Some data sources, including IATI data, draw a distinction between "planned end date" and "actual end date." In the AidData portal, the End Date corresponds to the "actual end date." If a donor source only provided one end date and did not specify whether it was a planned or actual end date, the default is to assume it is the actual end date, so it is mapped to the End Date field. This field is recorded as an ISO date format (YYYY-MM-DD).
Even Split Commitments	L1G	Commitment and disbursement amounts for a project are not location-specific. As a way to estimate the amount of aid flowing to a specific geocoded location, this field takes the total commitment amount for a project and splits it evenly across all project locations.
Finance Type	PE, L1C	Indicates what form the assistance came in. Field values include grant, loan (concessional), loan (non-concessional), loan (unknown), grant-like, equity, and so forth.
Financing Agency/Provider	PE,L1C	The sub agency within the sovereign government (i.e. Ministry of Foreign Affairs, USAID, etc) that financed the project. This is generally only applicable for bilateral donors.
Financing Type	PE, L1C	The project's type of international assistance flow, including the following: **Loan (Non-concessional)**: A project where the funds provided must be paid back using market interest rates or the terms less than 25% concessional. **Loan (unknown)**: A project where the funds provided must be paid back, but the concessionality of the loan terms are unknown. It is likely many of these loans are concessional, but without sufficient information to certify a project's concessionality, the flow type is assigned to this. **Loan (concessional)**: A project where a portion of the funds provided must be paid back, but the loan uses concessional terms. To qualify for this category, the concessionality must meet the OECD-DAC's requirements including having a grant element larger than 25%. **Grant**: A project where flows given do not need to be paid back. This can include the monetary value of in-kind contributions provided by donors. According to CRS directives, grants are transfers in cash or in kind for which no legal debt is incurred by the recipient. For DAC/CRS reporting purposes, it also includes debt forgiveness, which does not entail new transfers; support to non-governmental organisations; and certain costs incurred in the implementation of aid programmes. **Loan & Grant**: A project that includes a combination of loan and grant. This is only used if the grant proportion is unknown or falls under the 25% threshold. Otherwise, the project should be categorized as a concessional loan. **Equity Investment:** Some development financiers invest in a recipient country's institutions/companies. They purchase shares/equity in the company. The CRS identifies equity investment as "comprises direct financing of enterprises in a developing country which does not (as opposed to direct investment) imply a lasting interest in the enterprise. **Grant-Like:** A project where loan repayments directly benefit the recipient

Field Name	*Products*	*Definition & Notes*
		country. This category is only used when specified by the reporting organization (the CRS). According to CRS directives, this category comprises a) loans for which the service payments are to be made into an account in the borrowing country and used in the borrowing country for its own benefit, and b) provision of commodities for sale in the recipient's currency the proceeds of which are used in the recipient country for its own benefit. ● **Export Credits:** Loans for the purpose of trade and which are not represented by a negotiable instrument. They may be extended by the official or the private sector. If extended by the private sector, they may be supported by official guarantees. ● **Not Reported:** The donor did not report enough information to determine the financing type.
Flow Type	PE,L1C	The project's type of international assistance flow, including the following: ● **ODA:** The project counts as Official Development Assistance as categorized by the OECD-DAC. (See ODA definition). ● **OOF:** According to the CRS, this category includes official sector transactions which do not meet the ODA criteria. This category is only used when specified by the reporting organization (the CRS). (See OOF definition). ● **ODA-like Flows:** Flows of official financing administered by a non-DAC donor which meets the standards of official development assistance; the promotion of the economic development and welfare of developing countries as its main objective, and a perceived to be concessional in character with a grant element of at least 25 percent (using a fixed 10 percent rate of discount). ● **OOF-like Flows:** Financing to developing countries for representational or essentially commercial purposes provided by a non-DAC donor; ii.) Official bilateral transactions intended to promote development but having a grant element of less than 25 per cent; iii.) Official bilateral transactions, whatever their grant element, that are primarily export-facilitating in purpose. This category includes by definition export credits extended directly to an aid recipient by an official agency or institution ("official direct export credits"); iv.) The net acquisition by governments and central monetary institutions of securities issued by multilateral development banks at market terms; v.) Subsidies (grants) to the private sector to soften its credits to developing countries; vi.) Funds in support of private investment.
Flow Type	PAE	The type of aggregate flow, including FDI inflow or outflow, remittance inflow or outflow, international aid, or international US private foundation flows.
Fund	PE,L1C	The fund financing a project within a multilateral agency. For example, the NADB is a donor but it has several funds that it can pull money from to fund projects. This field allows specifies what fund that commitment was pulled from. Sometimes it'll just be OCR, sometimes special funds, sometimes left blank. There can be a many-to-one relationship between specified funds and one project, so if a project is pulling money from more than one fund (under a single donor), it is captured here.
Funding Agency Type	PE, L1C	Indicates whether a donor is a bilateral or multilateral agency.
Funding Region	PE, L1C	The region where the donor/funder is located.
Geoname ID	PE, L1G	The geoname ID retrieved from http://geonames.org
Grace Period	PE,L1C	The time from the date of loan signature to the date the first repayment is due. The unit for this field is years. So a value of '2' indicates the grace period is 2 years. This is only filled in if explicitly specified by the donor.
Grant Element	PE, L1C	The grant element reflects the financial terms of a loan transaction: interest rate, maturity (interval to final repayment) and grace period (interval to first repayment of

Field Name	Products	Definition & Notes
		capital). It is a measure of the concessionality (softness) of a loan. It is calculated as the difference between the face value of a loan and the discounted present value of the service payments the borrower will make over the lifetime of the loan, expressed as a percentage of the face value. To qualify as Official Development Assistance (pre 2018), the grant element needs to be above 25%.
Guarantor	PE,L1C	For loans this is the organization guaranteeing the loan. Only available for donors that explicitly make this information available.
Implementing Agency	PE,L1C	The organization that is responsible for implementing the project and is the channel of delivery (in OECD-DAC terminology). This usually ranges from a government agency (from the donor or partner country), an intergovernmental organization (such as UNDP), to private national or international NGOs. This field is not normalized. When collecting data directly from the donor, we standardize implementing agency names as much as possible with reporting the full agency name and then its acronym, e.g. "United Nations Development Programme (UNDP)." Sometimes implementing agency is listed as executing agency, or in the CRS, it is reported in the field "Channel Reported Name".
Implementing Agency	PE, L1C	The agency responsible for executing the project.
Imputed Flow Type	PE, L1C	Indicates whether a project is likely ODA, OOF, or other. While data originally from the OECD CRS reports the flow type for all flows, data from other sources does not always clearly report this information. This classification is often important for users. To help make the data more useful, AidData uses what information is available to estimate whether the flow is likely to be ODA or OOF. Estimated flows are reported as "ODA-like" and "OOF-like", whereas records not requiring estimation (because the source classified the flow) is classified using "ODA" and "OOF" classifications.
Interest Rate	PE,L1C	The interest rate provided on the loan term. The unit is %-ages. So an interest rate of four and one half percent is entered as 4.5. This is only filled in if explicitly specified by the donor.
Is Geocoded	PE, L1G	Flag to indicate whether a project was geocoded
Latitude	PE, L1G	Latitude of a location in decimal degrees
Loan Term	PE,L1C	The length of the entire loan. This is often referred to as loan maturity. The unit for this field is years. So a value of '30' indicates the loan term is 30 years. This is only filled in if explicitly specified by the donor.
Location	PE,L1C	Reported geographic location (from donor or reporting organization).
Location Type Code	PE, L1G	One of the location type codes from http://geonames.org/export/codes.html
Long Description	PE,L1C	A long description of the project activities and objectives. The length of the description, and the resulting depth of information, varies by donor.
Longitude	PE, L1G	Longitude of a location in decimal degrees
Origin	PAE	The origin country for the specified aggregate flow to all other countries. For example, if the origin for FDI flows is listed as the US, then that row represents the total FDI flow from the US to all other countries.
Place Name	PE, L1G	The name of the place from geonames
Planned End Date	PE,L1C	For the sources that make this information available, this is the date that the project was planned to end. When available, this date (theoretically) can be compared to the actual end date provided in the End Date field.
Planned Start Date	PE,L1C	For the sources that make this information available, this is the date that the project was planned to start. When available, this date (theoretically) can be compared to the actual start date provided in the Start Date field.
Precision Code	PE, L1G	The precision code of the assigned location during the geocoding process. 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 these varying levels of precision. Precision code 1 = exact location, 2 =

Field Name	*Products*	*Definition & Notes*
		"near" a specific location, 3 = district/city (ADM2), 4 = state/province (ADM1), 5 - location lies between populated places, along rivers/road/borders, or spans an area larger than a state/province, 6 = country, 7 = sub-country information is not available, 8 = local or national capital.
Project Location Id	PE, L1G	Unique project location identifier
Recipient	PE,L1C	The sovereign country that is the recipient of the project. This is often referred to as the Partner country in recent aid literature. We have sought to match our country names to those used by the CRS. Similarly we have tried to match our regional groupings to theirs. At times flows are directed at NGOs – in that case we put the recipient as the country that is benefitting from the project. In some cases the donor identifies a region as a recipient instead of a specific country. In these cases, we have mapped the donor-specified region to the corresponding region in our dataset. Some donors will report one project to multiple countries or regions but only one commitment amount for the project as a whole. In those cases, we seek to retain as much information as possible by identifying the smallest regional category that encompasses all recipients, and that is reported in the recipient field. When we know specific country names but have to use a regional recipient we retain the information on the individual countries in the additional info field. If the commitment amount is disaggregated for each recipient, then each recipient and corresponding amount will be recorded in separate rows (in a flat table structure)
Sector Name Trans	PE, L1G	The translated, pipe-delimited list of AMP sectors
Short Description	PE,L1C	A short description of the project activities. This field is only filled in for data sourced from the CRS. For data collected directly from the donor, all descriptive information should be mapped to either the title or the long description.
Source	PE,L1C	Source type for the project records. The current source types include OECD CRS, Donor Documents, AMP, and IATI.
Source Detail	PE,L1C	URL providing source information. Our goal is to provide a pdf version of the original source for each project. Currently this is complete for recently-collected data, and we hope to expand this to all data soon.
Start Actual Type	PE, L1G	Type of the project start (example: start-actual)
Start Date (Actual)	PE,L1C, L1G	Date project began. A full date is required for this field. However, donor data will often report incomplete dates (such as month-year combos or only the year). In that case, the default date will be the the first day of the specified month or year. (i.e. if the documentation says the project started in March 2011, the start date in the Portal would be 3/1/2011). Some data sources, including IATI data, draw a distinction between "planned start date" and "actual start date." In the AidData portal, the Start Date corresponds to the "actual start date." If a donor source only provided one start date and did not specify whether it was a planned or actual start date, the default is to assume it is the actual start date, so it is mapped to the Start Date. This field is recorded as an ISO date format (YYYY-MM-DD).
Status	PE, L1G	Status of the project (ex. Completed)
Title	PE,L1C, L1G	The name of a project or program specified by the donor or reporting organization (such as the CRS). Of all the descriptive fields (including title, short description, and long description), title should be prioritized over the other fields. If limited descriptive information exists for a project when collecting data from a source, this should be the first field to be filled in--so if a donor only provides a single sentence or sentence fragment describing the project, that information should become the title of the project.
Total Commitments	PE, L1G	Sum of the total commitments from the transactions for a specific project. These are not location-specific and should not be summed across multiple locations for a

Field Name	Products	Definition & Notes
		project.
Total Disbursements	PE, L1G	Sum of the total disbursements from the transactions for a specific project. These are not location-specific and should not be summed across multiple locations for a project.
Total Project Cost	PE,L1C	The total cost of the project from all funding sources. This usually includes co-financing amounts from other donors as well as the partner country itself. This field is only filled in when the donor provides the information explicitly. Any totals using this field across projects is likely to produce double counting.
Transaction Currency	PE, L1G	Type of currency (example: USD, EUR)
Transaction Date	PE, L1G	The date of the transaction in ISO format (YYYY-MM-DD)
Transaction Value	PE, L1G	Amount in currency of transaction
Transaction Value Code	PE, L1G	Type of transaction (example: C = Commitment, D = Disbursement)
Transaction Year	PE, L1G	The year of the transaction
Umbrella	PE, L1C	This is only applicable for multilateral donors. It helps users group associated donors together. For example, World Bank associated trust funds fall under the World Bank umbrella, and AFDB and AFDF fall under the AFDB Group umbrella.
Year	PE, L1C, PAE	This is the year of the recorded transaction. For most cases this represents the year of the annual flows for the project transactions (so it is a transaction-level field). For some data collected from donor reporting systems, this represents the year of the project commitment (if annual transaction flows were not available or were undesirable for some reason).

Table 2: TUFF Field Definitions

Field Name	Products	Definition
Active	L1T	Dichotomous variable to isolate duplicate or false projects. Each record is either "active" or "inactive". Note: project IDs are initially coded as "active" and remain such until labeled "inactive" by a researcher. Project IDs found to be duplicates of preexisting project IDs, as well as those that are not found to constitute a project, are labeled as "inactive." Inactive projects are later reviewed for potential data grabs and are subsequently deleted from the database. Prior to data analysis, researchers should exclude all inactive projects (marked "f").
Active Code	L1T	Dummy variable for "active" variable. 1=active; 2=not inactive.
All Recipients	L1T	Countries and/or regions receiving development assistance from donor entity. Followed with comma by sub-national location, if specified in reports. Each recipient entity is followed by the percentage of funds being disbursed to that entity.
Amount	L1T	Monetary amount committed by the donor entity for the completion of a project, in the most original currency reported.
Cofinanced	L1T	Dummy variable to indicate the project has multiple donors (including China), where the financial amount provided by each donor separately could not distinguished.
CRS Sector Code	L1T	CRS Sector: Three-digit code for primary project sector. Codes are taken from the OECD's CRS sector codes ("DAC 5" codes). See http://www.oecd.org/dac/aidstatistics/purposecodessectorclassification.htm. Sheet 4 in GREY contains all CRS sector codes used in the dataset
CRS Sector Name	L1T	Full descriptions of the 3-digit CRS sector code.
Currency	L1T	Currency associated with the monetary amount for a project as it was originally reported.
Debt Uncertain	L1T	Dummy variable to indicate a debt reduction agreement in which the amount of debt reduced in the final deal cannot be ascertained.
Deflator Used	L1T	Deflator used to convert project amount in original currency to 2009 U.S. Dollars value
Description	L1T	1-2 paragraph comprehensive summary of the activities supported by the project. See the methodology for a detailed explanation of how to record this variable.
Donor	L1T	Country providing the finance for the project.
Donor Agency	L1T	List of all organizations involved which are based in the donor entity and could be identified.
Donor Agency Count	L1T	Number of donor organizations involved in the planning and implementation of the project that could be identified.
End Actual	L1T	The actual date on which a project was completed.
End Planned	L1T	The announced completion date for a project.
Exchange Rate Used	L1T	Exchange rate used to convert project amount in original amount to 2009 U.S. Dollars value
Factiva Sources	L1T	URL(s) for all documentation found through Factiva during Stage One searches.
Field Completeness Score	L1T	This variable rates a TUFF project based on the number of critical fields that have been accurately population. Projects are penalized 1 point for missing information in the following fields: amount, commitment year, sector, funding agency, flow, and implementing agency. Projects are also penalized 1 point if they have the status Pipeline: Pledge or the flow class Vague OF. Project receive 4 bonus points if they have been verified by in-country enumerators. The theoretical maximum score a project can receive is 13.
Flow	L1T	Describes the way finances, goods or services were transferred from the donor to recipient in the project. See the methodology for precise classifications. Flow types include: Debt Forgiveness, Debt Rescheduling, Export Credits, Foreign Direct Investment, Freestanding Technical Assistance, Monetary Grant, In-Kind Grant,

Field Name	Products	Definition
		Grant (Vague), Loan Guarantee, Loan, Joint Venture with Recipient, Scholarships/Training in Donor Country, Strategic/Supplier Credit
Flow Class	L1T	Classification of financial flow by degree of concessionality and level of state involvement. Categories include: **-ODA-like:** Flows of official financing administered by a non-DAC donor which meets the standards of official development assistance; the promotion of the economic development and welfare of developing countries as its main objective, and a perceived to be concessional in character with a grant element of at least 25 percent (using a fixed 10 percent rate of discount). **-OOF-like:** Financing to developing countries for representational or essentially commercial purposes provided by a non-DAC donor; ii.) Official bilateral transactions intended to promote development but having a grant element of less than 25 per cent; iii.) Official bilateral transactions, whatever their grant element, that are primarily export-facilitating in purpose. This category includes by definition export credits extended directly to an aid recipient by an official agency or institution ("official direct export credits"); iv.) The net acquisition by governments and central monetary institutions of securities issued by multilateral development banks at market terms; v.) Subsidies (grants) to the private sector to soften its credits to developing countries; vi.) Funds in support of private investment. **-Vague Official Finance:** TUFF projects financed by an official agency but which lack sufficient information on the intent or concessionality of an agreement to accurately sort into the ODA-like, OOF-like or Official Investment categories. **-Official Investment:** An international investment by a donor state agency in an enterprise resident in another country's economy. The donor agency must itself purchase a stake in the recipient enterprise, with the expectation of seeing a return on this investment for the donor government.
Flow Class Code	L1T	Coded "flow class" variable. 2 = ODA-like; 3 = OOF-like; 5 = Vague (Official Finance); 14 = Official Investment.
Funding Agency	L1T	List of all organizations involved in funding the project that could be identified. Includes each organization's type.
Grace Period	L1T	Reported grace period of the loan, in years
Grant Element	L1T	Grant element of the loan, in percent. Uses reported grant element of the loan, unless could be calculated using OECD grant element calculator (assumes 10 percent rate of discount, 2 payments per annum, and equity principal payment).
Ground-truthed	L1T	Dummy variable to indicate the projects that have been visited by professional enumerators to confirm project data
Implementing Agency	L1T	List of all organizations involved in project implementation that could be identified. Includes each organization's type.
Intent	L1T	Perceived intent of the finance provider. Categories covered by "donor intent" include: Development; Commercial; Representational; Mixed (some development); Mixed (no development)
Intent Code	L1T	Coded "intent" variable. 1 = Development, 2 = Commercial, 3 = Representational, 4 = Mixed, 5 = Mixed (No Development), 6 = Mixed (Some Development)
Interest Rate	L1T	Reported interest rate of the project loan, in percent
Is Official Finance	L1T	Coded variable denoting Official Finance designation: 1 = is official finance, 0 = not official finance.
Line of Credit	L1T	Dummy variable to indicate the project involves a line of credit. Lines of credit may or may not be used by the recipient.
Loan Type	L1T	Categorical variable to represent the general nature of the loan. Categories covered include: interest-free; concessional; non-concessional; no information.
Maturity	L1T	Reported duration of the project loan, in years
Project ID	L1T	Unique identification number assigned to every project record created in the database.

Field Name	Products	Definition
Recipient Agency	L1T	List of all organizations involved which are based in the recipient entitity that could be identified.
Recipient Agency Count	L1T	Number of recipient organizations involved in the planning and implementation of the project that could be identified.
Recipient Condensed	L1T	Same as the "recipients" variable, but projects with multiple recipients receive "Africa, regional" here.
Recipient Count	L1T	Number of countries or regions receiving funding through the project.
Recipient COW Code	L1T	Recipient country, as coded by Correlates of War data. http://www.correlatesofwar.org/datasets.htm
Recipient IMF Code	L1T	Three-digit name of recipient country as reported by the IMF. http://www.imf.org/external/np/sta/cdis/pdf/060508g.pdf
Recipient ISO2 Code	L1T	2-letter code for recipient country in ISO 3166-1 alpha-2 (published by the International Organization for Standardization) http://en.wikipedia.org/wiki/ISO_3166-1_alpha-2
Recipient ISO3 Code	L1T	3-letter code for recipient country in ISO 3166-1 alpha-3 (published by the International Organization for Standardization). http://en.wikipedia.org/wiki/ISO_3166-1_alpha-3
Recipient OECD Code	L1T	Truncated 3-letter name of recipient country from the OECD codebook. www.oecd.org/dataoecd/31/15/34107835.xls
Recipient OECD Name	L1T	Name of recipient country as coded by the OECD. www.oecd.org/dataoecd/31/15/34107835.xls
Recipient UN Code	L1T	Three-digit code of recipient country as reported by the UN. http://unstats.un.org/unsd/methods/m49/m49.htm
Sector Comment	L1T	Details regarding the specific sector nature of the project. For projects coded as "Multisector," coders can list each of the known sectors in "Sector Comment."
Source Triangulation Score	L1T	This variable rates a TUFF project based on the number and diversity of information sources supporting it. Projects are awarded points in the following way. 1 point for each media report supporting a record (a project can receive at most 4 points from media reports. 3 points from official sources including donor, recipient and multilateral agencies. 2 points from academic articles and other academic conference or working papers. 2 points for implementing agency sources. 1 point from NGO sources and social media. The theoretical maximum score a project can receive is 15.
Sources	L1T	String of text on source documentation, following this pattern: "[Document URL], [source document type], [platform used to access source document];"
Sources Count	L1T	The number of documents used to create the project record.
Start Actual	L1T	The actual date on which project implementation began.
Start Planned	L1T	The announced start date for a project.
Status	L1T	This TUFF field captures the current stage of a project based on available open-source documentation. The label takes the following values: Pipeline: Pledge, a non-legally binding announcement of aid which should not be counted toward a donor's aggregate assistance; Pipeline: Commitment, a legally binding agreement to provide financing; Implementation, used when we have found evidence that disbursements of money, goods or services has begun; Completion, when disbursements end or the donor hands a project over to a recipient; Suspended and Cancelled.
Status Code	L1T	Coded interval for the "status" variable. 0 = Cancelled; 11 = Pipeline: Commitment; 12 = Pipeline: Pledge; 2 = Implementation; 3 = Completion; 4 = Suspended.
Title	L1T	Short phrase describing the nature of the project.
USD 2009	L1T	Deflated monetary equivalent of reported monetary amount in reported currency to 2009 U.S. Dollars.
Verified	L1T	Status of the project record within the database. "Checked" IDs have received both Stage One and Stage Two treatment. "Raw" IDs were created in Stage One

Field Name	Products	Definition
		searches, but have not been checked through Stage Two treatment. "Suspicious" IDs are sourced in dubious reports and require further investigation. Excluding project IDs marked "Suspicious" and "Raw" from data analysis is highly recommended.
Verified Code	L1T	Coalesced binary variable for the "verified" variable. 0 = Raw and Suspicious; 1 = Checked.
Year	L1T	Year in which a formal commitment was made between a supplier of development finance and a recipient. For projects with no report of a formal commitment, the year in which an informal pledge was announced is used for the year variable.
Year Uncertain	L1T	Dummy variable for projects without any sources reporting a specific commitment year. In these cases, the year of the earliest media report serves as the "commitment year".

AidData Term Dictionary

Authors:
Scott Stewart (sstewart@aiddata.org)
Brooke Russell (brussell@aiddata.org)

Document History And Version History

Version	Authors	Date Approved	Approved By	Description
1.0	Scott Stewart, Brooke Russell	10/28/2015	Brad Parks, Co-Executive Director	This is the initial public release of the document.

Purpose

The purpose of the **AidData Term Dictionary** is to clearly define key concepts and terms related to AidData's work and the development sphere in general. This document is related to the Data Management Plan suite of documents as defined in **Figure 1: Data Management Plan Structure and related documents**.

This document seeks to provide a comprehensive list of definitions for the terms used throughout AidData's Data Management Plan and various working teams.

Term	*Definition & Notes*
Aid Management Platform (AMP)	A software application developed by Development Gateway that allows recipient countries to track and manage incoming development flows.
AidData Activity Codes	AidData's coding scheme pioneers a unique approach to categorizing projects where each project not only gets one overall purpose code (much like the CRS purpose code), but it is also tagged with activity codes to indicate all the separate activities included in the project. The AidData coding scheme includes over 550 separate activity codes loosely based on the OECD CRS purpose code structure. Each project can be assigned as many activity codes as necessary (including across multiple sectors).
Bilateral Donor	A sovereign country providing official development finance to another country. AidData seeks to collect information on all official development finance from all bilateral donors.
Bilateral Flows	Flows provided directly by a donor country to an aid recipient country. The AidData core database seeks to collect data on all bilateral flows with a development purpose from all bilateral donors.
BRICS	An acronym for an association of five major emerging economies (and consequently emerging donors): Brazil, Russia, India, China, South Africa. These donors are all considered non-DAC/non-traditional donors.
Conditionality	A measure of the "softness" of a credit reflecting the benefit to the borrower compared to a loan at market rate. (cf. Grant Element). Technically, it is calculated as the difference between the nominal value of a tied aid credit and the present value of the debt service as of the date of disbursement, calculated at a discount rate applicable to the currency of the transaction and expressed as a percentage of the nominal value.
Constant Dollars	For financial information that is available in a time series, it is important to account for inflation/deflation across those years. When financial information is provided in constant dollars, all financial amounts have been deflated to a base year to make the amounts comparable. This is in contrast to current or nominal dollars, where inflation has not

Term	*Definition & Notes*
	been accounted for (so time series analysis should not use nominal or current amounts).
Creditor Reporting System (CRS)	The OECD DAC's Creditor Reporting System (CRS). Members of the OECD are required to report their official development financing activities to the OECD DAC in two forms: 1) DAC Aggregates and 2) CRS Project-Level Data. The DAC aggregates represent total official flows (both bilateral and multilateral) from the specified donor. Many non-DAC donors will voluntarily report aggregate information to the DAC. The details on these flows are limited (because they are aggregates). The CRS data in contrast is provided at the project-level, which provides full details on where the development flows are going, when, and to what effect. Many multilateral donors also report to the CRS (voluntarily) along with 2 non-DAC bilateral donors (Kuwait and United Arab Emirates). The majority of AidData's core dataset comes from CRS data (especially DAC Bilateral Flows). But AidData supplements CRS data with additional data from other donors or sources to provide the most comprehensive view of development finance.
Data Management Plan (DMP)	A series of living documents that address the management of data produced by AidData, from the time of their data collection, to their entry into permanent archive. The Master document has many support documents which go into greater detail about each of these processes and documents data sources.
Debt Relief	In the context of international development, debt relief is when a donor forgives the debt of developing countries resulting from concessional loans. The Heavily Indebted Poor Countries (HIPC) initiative was ultimately launched to provide systematic debt relief for the poorest countries, whilst trying to ensure the money would be spent on poverty reduction.
Development Assistance Committee (DAC)	The committee of the OECD which deals with development co-operation matters. Currently there are 29 members of the DAC: Australia, Austria, Belgium, Canada, the Czech Republic, Denmark, Finland, France, Germany, Greece, Iceland, Ireland, Italy, Japan, Korea, Luxembourg, The Netherlands, New Zealand, Norway, Poland, Portugal, Slovak Republic, Slovenia, Spain, Sweden, Switzerland, the United Kingdom, the United States and the European Union. The UAE is also a DAC participant, but not a full voting member.
Donor System	This term is used in the AidData Data Management Plan to indicate the donors that we get data for direct from donor sources (annual reports, donor websites, donor documents, etc). The Donor Systems Data Management Plan details the data collection activities associated with this group of donors.
Double Counting	The possibility of reporting a development finance flow more than once within the AidData core database. The main potential source for double counting comes from the cross between reporting bilateral core contributions or earmarked contributions to multilateral agencies and then reporting on the flows from multilateral agencies to recipients. We seek to reduce this possibility as much as possible by following OECD CRS reporting directives and only tracking the following flows: 1) Pure bilateral flows, 2) Multi-bilateral flows reported by the bilateral donor, and 3) Multilateral flows from core resources. So the AidData database specifically excludes bilateral contributions to multilateral agencies in the form of core contributions as well as multilateral flows using earmarked contributions reported by the multilateral agency. Other sources for double counting include 1) using multiple sources for one donor (including in the same year or across years) and 2) duplicated projects from the same source. AidData seeks to eliminate all double counting wherever possible using rigorous quality assurance methods.
Export Credits	Loans for the purpose of trade and which are not represented by a negotiable instrument. They may be extended by the official or the private sector. If extended by the private sector, they may be supported by official guarantees.
Fragmentation of Aid	Describes aid that comes in too many small slices from too many donors, creating unnecessary and wasteful administrative costs and making it difficult to target aid where it is needed most.

Term	Definition & Notes
Fungibility of Aid	Fungibility of aid refers to the situation where receipt of funds from a donor for a specific purpose (e.g. education) allows the recipient country's government to reduce its own expenditure in that area and to spend its money on something else. This problem is most notorious when a recipient government is able to switch its expenditure to military equipment or status symbols (e.g. a presidential jet or a new palace) but it can also have tragic consequences if a government becomes dependent on donors to support a sector and that support is later withdrawn for political reasons. Addressing the problem of fungibility is a key concern for all donors.
Geocoding	The process of assigning a specific latitude and longitude to development project locations. AidData specializes in geocoding development projects from both the donor perspective (development project information originating from the donor to all recipients) as well as recipient perspective (project information originating from the recipient from all donors).
GIS	A geographic information system (*GIS*) is a system designed to capture, store, manipulate, analyze, manage, and present all types of spatial or geographical data. The geocoded data that AidData produces can be loaded into GIS software to visualize the information to analyze the information in different ways.
Grant Element	Reflects the financial terms of a commitment: interest rate, maturity and grace period. It measures the concessionality of a loan, expressed as the percentage by which the present value of the expected stream of repayments falls short of the repayments that would have been generated at market rates. The "market rate" the DAC uses is 10% (though this was a largely arbitrary threshold). Thus, the grant element is zero for a loan carrying an interest rate of 10 percent; it is 100 per cent for a grant; and it lies between these two limits for a soft loan. If the face value of a loan is multiplied by its grant element, the result is referred to as the grant equivalent of that loan (also Concessionality Level) [Note: the grant element concept is not applied to the market-based lending operations of the multilateral development banks. Instead, these are classified as concessional if they include a subsidy ("soft window" operations) and non-concessional if they are unsubsidised ("hard window" operations)].
Heavily Indebted Poor Countries (HIPC)	A group of 38 developing countries with high levels of poverty and debt overhang which are eligible for special assistance from the International Monetary Fund (IMF) and the World Bank.
Humanitarian Aid/Emergency Aid	Humanitarian aid is material or logistical assistance provided for humanitarian purposes, typically in response to humanitarian crises including natural disasters and man-made disaster. The primary objective of humanitarian aid is to save lives, alleviate suffering, and maintain human dignity. In contrast, development aid has a longer perspective.
International Aid Transparency Initiative (IATI)	IATI developed an international standard to make information about aid spending easier to find, use, and compare. It's a framework for publishing information on development cooperation activities in a timely, comprehensive and forward-looking manner that is universally comparable. Donors and reporting organizations can publish their data using the IATI standard and then post it on the IATI registry. AidData is a member of the technical working group for IATI. The IATI standard uses hierarchical data (similar to the portal data for AidData).
Military Aid	Aid intended to bolster a country's defense sector. This type of aid is not included in AidData or the OECD CRS.
Multi-bi Flows	Projects implemented by multilateral organizations on behalf of donor countries. In this case, a bilateral donor gives a multilateral agency a contribution, but that contribution is earmarked in some way (either for a specific project, thematic area, or recipient). These flows are reported as bilateral flows by the bilateral donor government. As such, AidData seeks to track all multi-bi flows where the donor makes the information available.
Multilateral Contributions	Monetary contributions by a bilateral donor government to a multilateral agency. These can take two forms: Core contributions or earmarked contributions. Core contributions consist of money that becomes pooled with other bilateral contributions to become part of the multilateral agency's regular budget. With core contributions, the bilateral donor

Term	Definition & Notes
	loses all control over how the money is spent. Capital subscriptions or assessed fees count as core contributions. With earmarked contributions, the bilateral donor specifies a particular purpose for the money--either at the recipient, sector, or project level. As such, the multilateral agency essentially becomes the implementing/executing agency for the project, but the bilateral donor retains some amount of control over how the money is allocated. Earmarked contributions are reported by the bilateral donor as a multi-bi flow and are included in the AidData core database. In contrast, core contributions to multilateral agencies are not included in AidData to reduce the risk of double counting.
Multilateral Donor/Organization	International institutions with governmental membership which conduct all or a significant part of their activities in favor of development and aid recipient countries. They include multilateral development banks (e.g. World Bank, regional development banks), United Nations agencies, and regional groupings (e.g. certain European Union and Arab agencies). AidData seeks to collect information on all official development finance from all multilateral donors.
Multilateral Flows	Flows provided by a multilateral development agency to a recipient country funded through the multilateral agency's regular budgets. AidData seeks to track all data on multilateral flows funded through core, pooled funding sources (see multilateral contributions).
NGOs	Non-Governmental Organizations. Funding originating from NGOs is not considered official development finance, so AidData does not currently track this information.
Non-DAC Donors	Donors that are not members of the OECD-DAC. AidData uses this term to indicate donors where we collect data direct from the donor instead of through the OECD CRS. Emerging donors, non-traditional donors, and south-south cooperation donors are often used as synonyms of non-DAC donors.
Non-Traditional Donors	Donors that are not members of the OECD-DAC. AidData uses this term to indicate donors where we collect data direct from the donor instead of through the OECD CRS. Emerging donors, non-DAC donors, and south-south cooperation donors are often used as synonyms of non-DAC donors.
ODA-Eligible Country	A country that, according to the OECD DAC, is eligible to receive ODA. These consist of all low and middle income countries based on gross national income (GNI) per capita as published by the World Bank, with the exception of G8 members, EU members, and countries with a firm date for entry into the EU. The list also includes all of the Least Developed Countries (LDCs) as defined by the United Nations (UN). If a concessional loan or grant goes to a non-ODA eligible country, then the flow is categorized as OOF instead of ODA.
Official Aid (OA)	This is an outdated category of aid flows that the OECD DAC no longer uses. It was used after 1991 - 2009 for development flows targeted towards countries that had higher income levels or were created as a result of the breakup of the USSR. Unfortunately the OECD CRS decided to delete these flows from its records as of 2010. To better capture a more complete picture of historical flows, AidData has preserved these flows by pulling them off of the OECD CRS 2009 CD and included them in our core data (where the source = "Previous Aid Flows").
Official Development Assistance (ODA)	A measure used by the OECD DAC to classify international flows focused on development in countries most in need compared to flows that may be to richer countries or with less than developmental purposes. To qualify as ODA, a flow must meet 4 criterion: 1) Come from an official source (government agency etc), 2) be focused on development (as opposed to having commercial or representational purposes), 3) have concessional terms -- if it's a loan, the grant element must be at least 25%, and 4) be to an ODA-eligible country. Flows for military purposes are excluded.
Official Development Financing	A categorization used by the OECD DAC to indicate all international flows that qualify as ODA or OOF.
Other Official Flows (OOF)	A measure used by the OECD DAC to classify international flows that are beyond the scope of ODA--either because they were not aimed at development or because they were not concessional (had a grant element of less than 25%).

Term	Definition & Notes
Randomistas	Researchers who believe that it is vital to collect evidence on the impact of possible new development initiatives, that this evidence should be collected using randomised control trials, and that (in most cases) aid should be confined to projects that such evidence supports.
Remittances	Remittances are monies sent from one individual or household to another. International remittances are those sent by migrant workers who left their home country. Domestic remittances are those sent by migrant workers who left their home village or town to work elsewhere in their home country (e.g. rural-urban migration; sometimes also referred to as national remittances). AidData incorporates remittance data into its pure aggregates portion of the portal. The data comes from the World Bank.
Reporting System	This term is used in the AidData Data Management Plan to indicate the data we get from aggregating data systems. Currently the only reporting system incorporated into AidData data products is the data from the OECD CRS on select bilateral and multilateral donors.
South-South Cooperation	In the context of international development, South-South Cooperation (SSC) involves the exchange of resources, technology, and knowledge between developing countries, also known as countries of the global South. Broadly speaking, South-South Cooperation involves development activities initiated by developing countries for other developing countries. These activities usually focus on knowledge sharing and capacity building, allowing developing countries to share their experiences and knowledge with other developing countries.
Sustainable Development Goals (SDGs)	A proposed set of targets relating to future international development. They are to replace the Millennium Development Goals once they expire at the end of 2015. They currently include a series of 17 far-reaching goals with 169 separate indicators to measure their progress through 2030.
Technical Cooperation	Technical co-operation is the provision of advice and/or skills, in the form of specialist personnel, training and scholarship, grants for research and associated costs.
Tied Aid	Official grants or loans where procurement of the goods or services involved is limited to the donor country or to a group of countries which does not include substantially all aid recipient countries. Tied aid loans, credits and associated financing packages are subject to certain disciplines concerning their concessionality levels, the countries to which they may be directed, and their developmental relevance so as to avoid using aid funds on projects that would be commercially viable with market finance, and to ensure that recipient countries receive good value.
Trilateral/Triangular Cooperation	A type of development cooperation involving three partners--typically one South-South cooperation partner, a traditional donor (bilateral or multilateral), and a beneficiary country in the developing world. Usually it facilitates knowledge sharing and training across all partnering organizations. An example would be Honduras partnering with Germany to implement a capacity building project in El Salvador. Information on these transactions is usually limited (due to inadequate reporting mechanisms). To the extent possible, AidData seeks to include information on trilateral cooperation.

Some definitions have been pulled from International Development Matters at
http://idml.co.uk/resources/glossary/ and the OECD DAC Glossary.