



USASPENDING.gov
DATA LAB

ANALYTICS USER GUIDE

To be used on conjunction with:

DATA LAB ALPHA RELEASE

Last Updated: September 19th, 2017

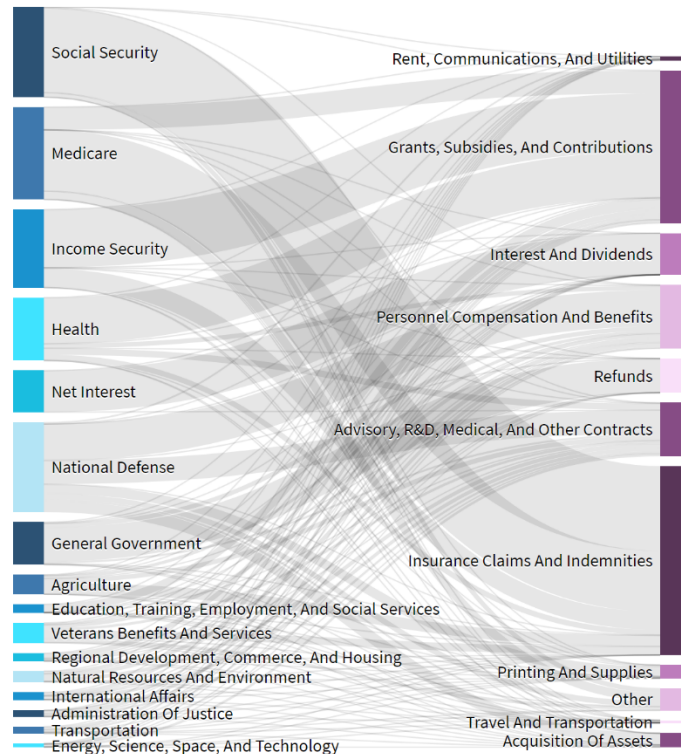
Contents

Data Tools Development Notes	3
Data Dictionary.....	6
API Guide.....	7
API Endpoints	7
Merging Endpoints	8

Data Tools Developer Notes

The tools and visualizations that were published as part of Data Lab Alpha Release were built by our team using the following methodology.

1. BUDGET OBJECT CLASS SANKEY



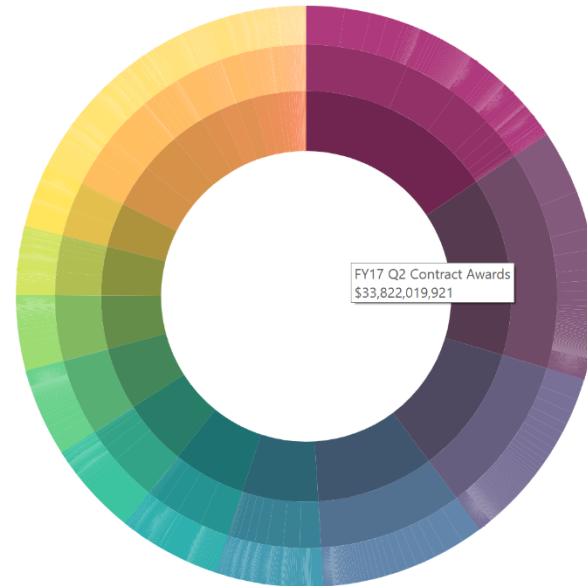
- Endpoint used: **TAS Categories**
- We grouped and aggregated using the following fields:
 - obligations_incurred_by_program_object_class_cpe,
 - object_class.major_object_class_name, and
 - treasury_account.budget_function_title
- The twenty official congressional budget functions were rolled up into sixteen budget function categories. We combined budget functions which had similar infrastructure needs and/or similar missions.
- The five major object classes were broken down into eleven more specific classes, each of which focuses on highlighting specific budget item categories. These categories were created to be more descriptive and meaningful than the original five major object class categories, and less overwhelming than the thirty-five official object class categories.

2. CONTRACT EXPLORER SUNBURST

FY17 Q2 Contract Awards

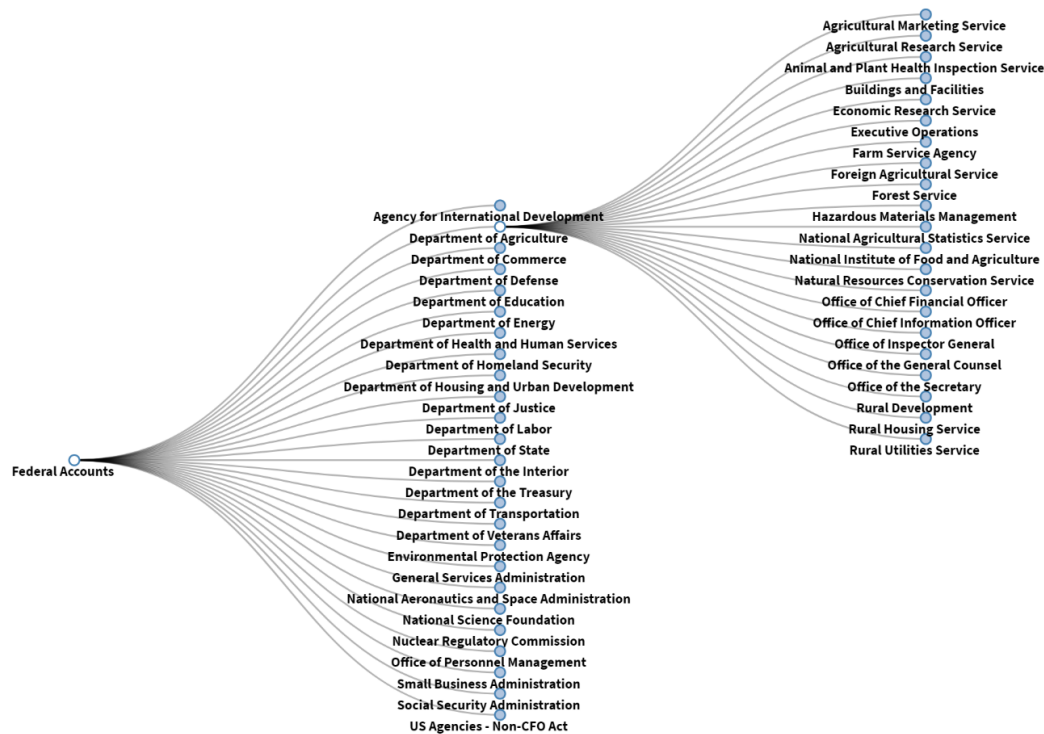
\$33,822,019,921

\$5,265,758,204 Department Of Health And Human Services
\$4,749,502,022 Department Of Veterans Affairs
\$3,441,701,574 National Aeronautics And Space Administration
\$3,135,630,956 Department Of Energy
\$2,024,444,602 Department Of Homeland Security
\$1,933,066,714 Department Of Defense
\$1,757,437,804 Department Of Commerce
\$1,663,594,281 Department Of State
\$1,419,041,593 Agency For International Development
\$1,297,637,658 Department Of Justice



- Endpoints used: **Awards, Transactions**
- These endpoints were merged on piid. The piid field is available using a POST request on the Transactions endpoint.
- To create the visualization, the Awards endpoint data was subset by selecting records that correspond to contracts, designated by being of type: purchase order, delivery order, or definitive contract. From these records, we grouped and aggregated on:
 - funding_agency.toptier_agency.name
 - funding_agency.subtier_agency.name
 - recipient.recipient_name, and
 - total_obligation
- To create the information panel, the merged data set was grouped and aggregated on:
 - funding_agency.toptier_agency.name
 - funding_agency.subtier_agency.name
 - recipient.recipient_name
 - contract_data.product_or_service_code, and
 - total_obligation.
- There are 1,731 unique product service codes currently in the data, these product service codes were rolled up into 188 parent product service codes that reflect the general category of the product or service being provided.

3. FEDERAL ACCOUNTS DENDROGRAM



- Endpoint used: Awards
- We used the fields:
 - `treasury_account_identifier.agency_id`
 - `treasury_account_identifier.federal_account.main_account_code`
 - `treasury_account_identifier.funding_toptier_agency.fpds_code`
 - `obligations_incurred_total_by_tas_cpe`
 - `treasury_account_identifier.federal_account.id`, and
 - `treasury_account_identifier.federal_account.account_title`
- These fields were grouped and aggregated the obligations data based on the Treasury Account Number (Agency ID plus main account code), then merged the endpoint data with the OMB hierarchy of Agency and Bureau Titles and corresponding Treasury Account Numbers for all active accounts. For agencies where the agency id was more than 2 digits long, we used the first 2 digits of the FPDS code for the agency id instead. This merge allowed us to add obligation amounts to the hierarchy of federal accounts and then display this information through the dendrogram's tree structure.
- We then made a few changes to make the information easier to navigate, including grouping all non-CFO act agencies together and pulling out or grouping together offices when the number of levels within an agency made it difficult to display through the dendrogram's set structure. In cases where a level within the agency would have been excluded, we added it at the end of the account name to be more descriptive. The federal account nodes link to the federal account pages on USAspending.gov based on the account Id.

Data Dictionary

Federal financial reporting elements are not always intuitive. For the official definitions of each data element you will find

The latest [Data Dictionary](#) is linked on the “Data Model” section of the DATA Act GitHub page.

API Guide

Purpose: To provide users of the USASpending API with information to facilitate use of data accessible from the API and lower the barriers to entry for new users.

Audience for the product: Internal DATA Act analytics team analysts and the public.

API Endpoints

The API provides several endpoints, a summary is provided for 8 key endpoint datasets that are available to export using a get request.

- A. Accounts Awards:** Financial account data by TAS account includes program activity, award id and object class. Each observation will include the relevant identifiers for PIID, FAIN, URI, treasury account data, awarding and funding agency, and federal account budget accounts data.
- B. Awards:** List of award records. Each observation identified by “id”, includes type of award, Procurement Instrument Identifier (PIID), Federal Account identification Number (FAIN), Unique Record Identifier (URI)¹, amount, awarding and funding agency details, as well as recipient details and place of performance information.
- C. Federal Accounts:** List of Federal Accounts. Observations are identified by “id”, and includes the following: agency identifier, main account code, account title, and the concatenated federal account code.
- D. Treasury Appropriations Accounts (TAS):** List of treasury appropriation accounts by Treasury Account Symbol (TAS). Observations are identified by the treasury account identifier, includes federal account information, reporting, funding, and awarding agency information as well as budget function codes.
- E. TAS Balances:** List of treasury appropriation account balances by quarter and by TAS. Observations identified by appropriation account balance id, includes Current Period End (CPE) and Fiscal Year End (FYE) obligations and outlay data, as well as treasury account, budget function codes, federal account, awarding and funding agency information.
- F. TAS Categories:** List of treasury appropriation account balances by fiscal year, TAS, program activity, and object class. Observations are identified by field named “financial account by program activity object class id”, include obligations and outlay information by CPE and FYE, program activity, submission, object class, treasury account identifiers, budget function, and funding and awarding agency identifiers.

A note on “TAS” ...

TAS is a complex concept, so let’s break this down each component for clarity. Each TAS is composed of:

- **Main Account Code (MAC):** 4 digits, Forms the bases of the federal account
- **Sub Account Code (SAC):** 3 digits, Rare, used in giant agencies that need to split up their accounts. Default is 000
- **Agency ID:** Agency Identifier, assigned by congress. Same role as CGAC
- **Allocation Transfer Agency ID (ATA):** Allocation Transfer Agency ID (usually not included)
- **Beginning period of availability (BPOA):** year when obligations are s available **AND**
- **Ending Period of availability (EPOA):** when they can no longer give obligations
- OR**
- **ATC:** Instead of specific years, an “X” represents an indefinite account: it can be used until deauthorized

¹ URI is only used where the combination of an assistance observation is not unique to a FAIN and Modification number. URI is defined by the agency.

- G. Subawards:** List of subaward records. Observations identified by “id”, include subaward number; award data; recipient details; North American Industry Classification System (NAICS) codes; submission identifiers; awarding, funding, and subtier awarding agency identifiers.
- H. Transactions:** List of transactions including contracts, grants, loans, etc. Observations are identified by “id”, includes transaction description, amount and date, as well as awarding, funding, and subtier agency engaging in the transaction. Transactions also include detailed recipient, contract, and place of performance data.

A note on “Obligations” v “Outlays”

Obligations are promises to pay for goods or services at a later point in time. This does not mean money spent as agencies still have this money. Outlays, on the other hand, is money spent and gone.

Table 1: Endpoint Unique Identifiers includes a summary of each endpoint and the unique identifiers that can be used to identify unique observations within each dataset.

Table 1: Endpoint Unique Identifiers

Endpoint Title	Unique Identifier
Accounts Awards	<ul style="list-style-type: none"> <i>financial_accounts_by_awards_id</i>
Awards	<ul style="list-style-type: none"> <i>id</i>
Federal Accounts	
Treasury Appropriation Accounts (TAS)	<ul style="list-style-type: none"> <i>treasury_account_identifier</i> <i>tas_rendering_label</i>
TAS Balances	<ul style="list-style-type: none"> <i>appropriation_account_balances_id</i>; or Combine: <i>treasury_account_identifier.treasury_account_identifier</i>² <i>treasury_account_identifier.tas_rendering_label</i>³
TAS Categories	<ul style="list-style-type: none"> <i>financial_accounts_by_program_activity_object_class</i>; or Combine: <i>treasury_account_identifier.treasury_account_identifier</i>; <i>program_activity.id</i>; <i>object_class.object_class</i>
Subawards	
Transactions	<ul style="list-style-type: none"> <i>id</i>

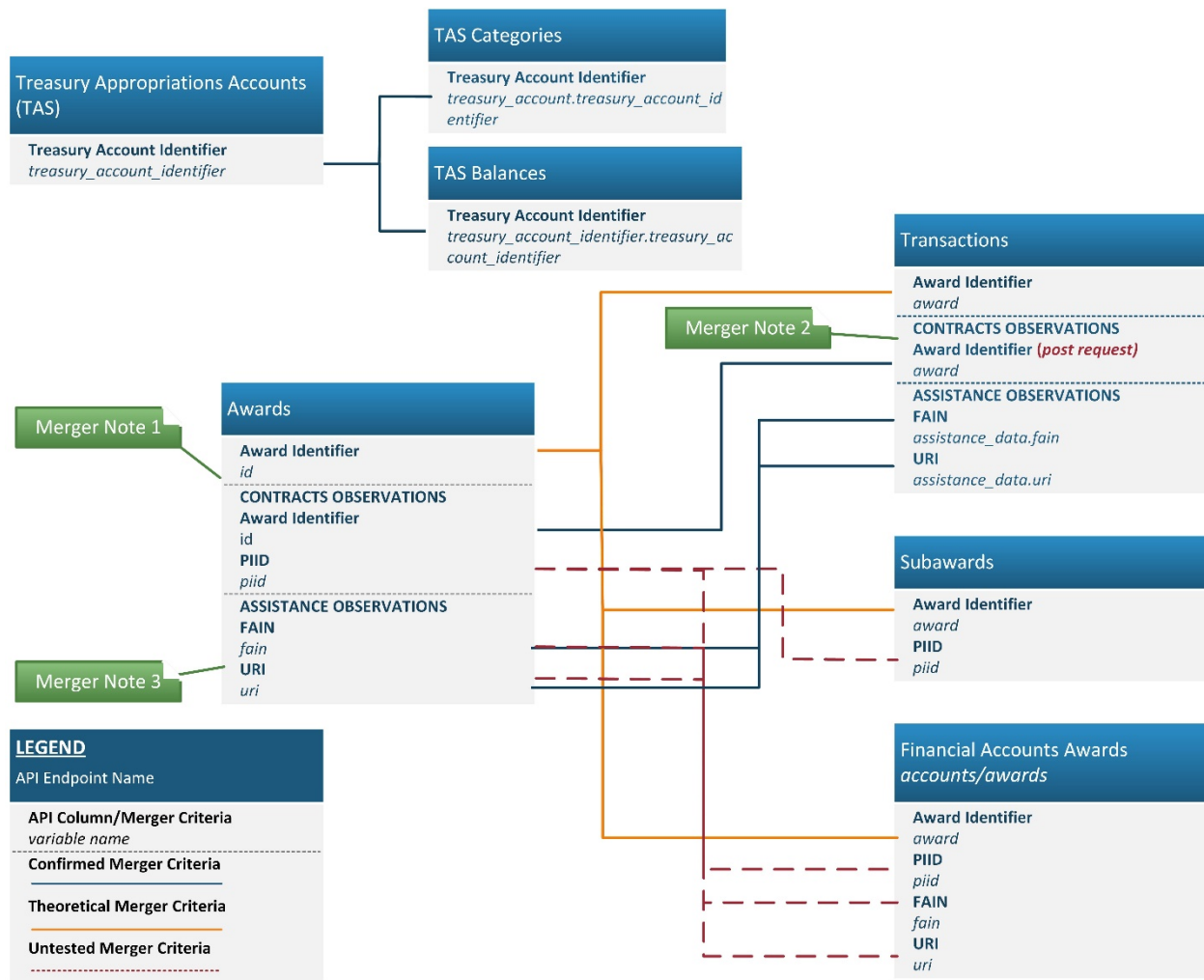
Merging Endpoints

For the endpoints listed above, **Error! Not a valid bookmark self-reference.** visualizes criteria that can be used to connect relevant datasets. This is a living document and will be updated as merger criteria is tested and confirmed.

² This will require some cleaning due to duplication of data submitted more than once for the same account but should be unique when duplicates are identified and removed.

³ This will require some cleaning due to duplication of data submitted more than once for the same account but should be unique when duplicates are identified and removed.

Figure 1: Endpoint Merging Map (Draft)



1. **Merger Note 1:** The Awards dataset generated by the get request can only be merged by separating observations into two groups, Contracts and Assistance observations, and merging by the criteria outlined above for each respective group.
2. **Merger Note 2:** The Awards dataset and transactions dataset theoretically should be able to be merged on id and award, respectively. However, this will require the use of the transactions post request to include the award column, which is not available in the get request.
3. **Merger Note 3:** Matching Award assistance observations suffers from data quality issues, therefore merging all assistance entries first on FAIN, then creating a second merge dataset merged on URI results in the greatest number of retained records. Note, to combine into one assistance dataset, each merged file will need to have the same columns (remove the unmerged duplicate FAIN/URI column from each), appended the datasets, then remove duplicates from the final complete assistance dataset.

Table 2: Endpoint Merger Criteria

Dataset A	Data Set and Merge Variables
Awards	Transactions: <ul style="list-style-type: none"> Contract Entries: <ul style="list-style-type: none"> <i>id</i> – award (only available in the transactions post request) Assistance Entries: <ul style="list-style-type: none"> <i>uri</i> – assistance_data.uri <i>fain</i> – assistance_data.fain Subawards <ul style="list-style-type: none"> TBD Accounts Awards <ul style="list-style-type: none"> Theoretically a complete match should be available using: <ul style="list-style-type: none"> <i>award to id</i> (creates approx.. 30% match on awards) <u>Three-part merger:</u> <ul style="list-style-type: none"> First match all observations, and drop all unmatched observations: <ul style="list-style-type: none"> <i>id to award</i> For Contract Entries, merge using: <ul style="list-style-type: none"> <i>piid</i> Assistance Entries (direct payments, loans, insurance, grants), merge using: <ul style="list-style-type: none"> <i>Fain and uri</i> Retain only matched for each merger, and then rbind the matched observations. Lastly delete all duplicated entries.
Federal Accounts	Awards: <ul style="list-style-type: none"> TBD
Treasury Appropriation Accounts (TAS)	TAS Balances:
TAS Balances	TAS Categories:
TAS Categories	TAS and TAS Balances: