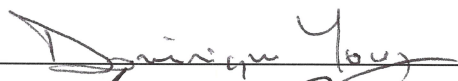


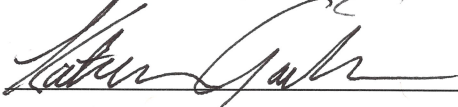
UCD CSN Technical Information #801D

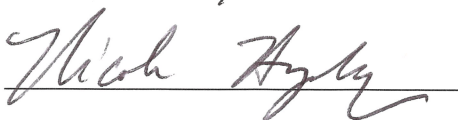
CSN Data Delivery

*Chemical Speciation Network
Air Quality Research Center
University of California, Davis*

*November 30, 2018
Version 1.1*

Prepared By:  Date: 11/28/2018

Reviewed By:  Date: 11/28/2018

Approved By:  Date: 11/28/18

DOCUMENT HISTORY

Date Modified	Initials	Section/s Modified	Brief Description of Modifications
11/30/18	NJS	1,2,3,7,8,9,10,11	Rewording for clarity and updating name changes.

Table of Contents

1. PURPOSE AND APPLICABILITY	4
2. SUMMARY OF THE METHOD.....	4
3. DEFINITIONS	4
4. HEALTH AND SAFETY WARNINGS.....	4
5. CAUTIONS	4
6. INTERFERENCES	4
7. PERSONNEL QUALIFICATIONS, DUTIES, AND TRAINING.....	5
8. PROCEDURAL STEPS	5
8.1 Format data for DART Submission	5
8.2 Receive and Review Returned Data from DART	6
8.3 Prepare and Submit to AQS	7
9. DATA DELIVERY FLOW.....	11
10. EQUIPMENT AND SUPPLIES.....	12
11. QUALITY ASSURANCE AND QUALITY CONTROL.....	12
12. REFERENCES	12

List of Figures

Figure 1. Returned DART datasets can be imported into the CSN database with this custom upload tool.	6
Figure 2. Login screen for the EPA's Exchange Network Services website.	8
Figure 3. Home screen of the Exchange Network Services website.	9
Figure 4. Enter "AQS" into the search bar.....	9
Figure 5. Search results from "AQS". The analyst should choose the Service named "AQS Submit".	10
Figure 6. AQS data submission form.....	11
Figure 7. Flow diagram of the data delivery process.....	11

1. PURPOSE AND APPLICABILITY

The subject of this technical information document (TI) is delivery of the resultant data from the Chemical Speciation Network (CSN). This document describes the procedure for preparing and delivering data to the Data Analysis and Reporting Tool (DART) and the Air Quality System (AQS) database.

2. SUMMARY OF THE METHOD

The UCD analyst prepares delivery files of the validated CSN data sets using custom tools in the *datvalCSN* R package. Data are formatted for delivery to DART, where they are reviewed by state, local, and tribal (SLT) agencies. After the 30 day review period, revised DART results are returned to UCD. The analyst ingests this file into the UCD CSN database, address any comments and/or requests from the DART reviewers, and reformats the dataset into AQS format. The final data file are checked for correctness and then submitted to the EPA's AQS database.

3. DEFINITIONS

- **AQS:** EPA's Air Quality System database.
- **Chemical Speciation Network (CSN):** EPA's PM_{2.5} sampling network, with sites located principally in urban areas.
- **crocker:** A custom software package in the R language that contains the data processing code used to produce, check, and post the final results.
- **CSN database:** A SQL Server database that is the central warehouse of CSN preliminary and final data at UCD.
- **CSV:** a comma-separated value file that is the common format for delivery files.
- **datvalCSN:** A custom software package in the R language that contains the data validation code used to collect, compare, and flag the final results.
- **Data Analysis and Reporting Tool (DART):** A web application for environmental data visualization and validation procedures.
- **STI:** Sonoma Tech, Inc. Contractor developing and operating the DART interface.

4. HEALTH AND SAFETY WARNINGS

Not applicable.

5. CAUTIONS

Not applicable.

6. INTERFERENCES

Not applicable.

7. PERSONNEL QUALIFICATIONS, DUTIES, AND TRAINING

The UCD Air Quality Research Center (AQRC) Data & Reporting Group staff assigned to this project all have advanced training in data analysis and handling.

8. PROCEDURAL STEPS

Data delivery is performed using the *datvalCSN* R package, which is developed and maintained by UCD specifically for data processing, monitoring, and validation of the CSN data. Data delivery is performed by the UCD Data & Reporting on monthly batches of data (a calendar month of sample start dates). Delivery occurs in three steps:

1. Format data and submit to DART (Section (8.1)).
2. Receive and reformat returned data from DART (Section 8.2).
3. Prepare and submit to AQS (Section 8.3).

8.1 Format data for DART Submission

In the previous step, results for elements, ions, and carbon are validated using the tools of the *datvalCSN* package. This includes assigning qualifier and null flags to specific records. Once the analyst has completed the validation, the analyst will run the *csn_validate* function with the *write.flags* parameter set to *TRUE*¹,

```
[monthData] <- datvalCSN::csn_validate(Month = ['MM'], Year = ['YYYY'], write.flags = [FALSE], with.mass = TRUE, recipient = "DART", server = 'production')
```

Specifically, the flags assigned when this function is run, as described in *UCD SOP #801 Processing & Validating Raw Data*. The analyst will then populate the *delivery.DartExport* table in the CSN database using

```
[dart.post] <- post_output([allData]@Output.DF, label = [YYYY Month],  
comment = [YYYY Month DART export], server =  
['production']).
```

Next, the *generate_dart_file* function collects the necessary data from the CSN database and formats it for DART. To select specific data, the analyst can enter either one or more unique DartSet IDs or a date range to prepare the file. Additionally, the analyst can generate a local copy of the output file in their R session for final review before generating the delivery file by specifying the *make.file* parameter as *FALSE*. Using DartSet IDs and setting *make.file* to *TRUE* generate the final delivery file for DART in the directory specified for the *out.dir* parameter:

```
[dart.data] <- generate_dart_file([dartSets], make.file = TRUE,
```

¹ Text in [brackets] indicates values that can be changed by the user. Other values should be typed as written.

out.dir = [output directory]).

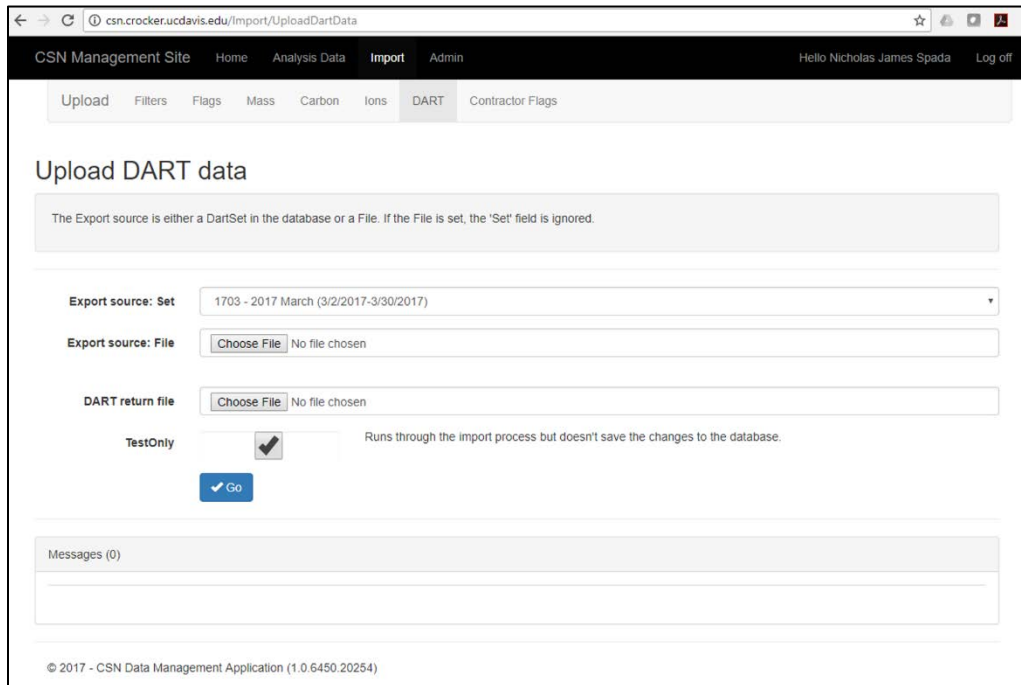
Files are delivered to STI by UCD via FTP service.

8.2 Receive and Review Returned Data from DART

The SLT agencies have 30 days to review their associated data and perform validation. Upon completion, the entire dataset is returned to UCD with a change log. The data are ingested using the DART import tool in the UCD CSN Management Site web application (csn.crocker.ucdavis.edu). The analyst will navigate to the “Import” tools (top menu bar), then to the “DART” upload tool (second from right option near top of screen) (see Figure 1).

To aid in document review, the DART return file is compared to the data that was delivered from UCD. The analyst can either specify the data set in the database using the top drop down menu labeled “Export source: Set” or by providing the path to the delivered file using the “Choose File” button next to the label “Export source: File”. The “DART return file” is the comma-separated value (CSV) file containing the dataset that was validated in DART.

Figure 1. Returned DART datasets can be imported into the CSN database with this custom upload tool.



The screenshot displays the 'Upload DART data' interface within the CSN Management Site. The browser address bar shows the URL csn.crocker.ucdavis.edu/Import/UploadDartData. The navigation menu includes 'Upload', 'Filters', 'Flags', 'Mass', 'Carbon', 'Ions', 'DART', and 'Contractor Flags'. The 'DART' menu item is currently selected. The user is identified as 'Hello Nicholas James Spada' and has a 'Log off' option. The main content area features a heading 'Upload DART data' and a note: 'The Export source is either a DartSet in the database or a File. If the File is set, the 'Set' field is ignored.' Below this, there are three input fields: 'Export source: Set' with a dropdown menu showing '1703 - 2017 March (3/2/2017-3/30/2017)', 'Export source: File' with a 'Choose File' button and 'No file chosen' text, and 'DART return file' with a 'Choose File' button and 'No file chosen' text. A 'TestOnly' checkbox is checked, with a note: 'Runs through the import process but doesn't save the changes to the database.' A blue 'Go' button is positioned below the 'TestOnly' checkbox. At the bottom, there is a 'Messages (0)' section and a footer: '© 2017 - CSN Data Management Application (1.0.6450.20254)'.

As with the other data upload processes, use the “Go” button to begin ingestion. It is advisable to first run the import with the “TestOnly” box checked and review any error or warning messages. The ingested data appear in the *analysis.DartReturn* table.

The analyst will review the change log from STI, including comments, and verify that all changes are consistent with the DART users' comments. To collect and review the changes made during DART validation, the analyst can run

```
[dartChanges] <- datvalCSN::get_changes([startdate], [enddate])
```

where both the *startdate* and the *enddate* parameters are entered as 'YYYY-MM-DD'. Typically, this process is being performed on the most recent dataset. If this is the case, only the start date needs to be entered.

The analyst will then convert the data set from DART to AQS format using the function,

```
[aqs] <- dart2aqs([dartSetId])
```

with the appropriate DART data set ID. The analyst can instead provide start and end dates as in the *generate_dart_file* function. The *dart2aqs* function populates the *delivery.AqsStaging* and *delivery.AqsStagingQualifierCodes* tables in the UCD CSN database.

The analyst will need to review the DART validator changes and comments and make any requested updates in the aforementioned tables. The *datvalCSN* package includes a number of functions to aid in this process, including:

- *update_conc*: corrects concentration values after a flow value was changed
- *update_value*: applies a scalar multiplication to analytical results for a specified filter ID
- *update_single_value*: replaces a result value for a specified parameter and filter ID
- *resurrect_flags*: reapplies qualifier flags for records that are no longer invalid
- *invalidate_aqs_filter*: applies a null code to a filter (may be parameter specific)
- *swap_aqs_records*: swaps dates between two samples
- *re_validate_filter*: remove null code and reapply qualifier flags for a specific filter (may be parameter specific)

8.3 Prepare and Submit to AQS

Once all changes have been made in the *delivery.AqsStaging* and *delivery.AqsStagingQualifierCodes* tables, the analyst will create the final dataset in the *delivery.AqsExport* table using the *aqsStaging2export* function,

```
[aqs.export] <- aqsStaging2export([dartSets])
```

The analyst will then prepare the final delivery file using the *generate_aqs_file* function,

```
[aqs] <- generate_aqs_file([dartSets], out.dir = ['U:/CSN/Delivered/AQS/'])
```

with one or more appropriate DART data set IDs. For both of these functions, the analyst can instead provide start and end dates. The analyst should review the resulting output dataset for accuracy and consistency with AQS formatting rules. AQS reference documents are located on the shared network drive at U:/CSN/Documentation/Reference.

Once the AQS file is ready for delivery, the analyst will use a web browser of their choice and navigate to the EPA's Exchange Network Services website (<https://enservices.epa.gov/login.aspx>).

Figure 2. Login screen for the EPA's Exchange Network Services website.

The screenshot shows the login page for the EPA's Exchange Network Services Center. The page has a header with the "Exchange Network" logo and "SERVICES CENTER" text. Below the header, there is a "SERVICES CENTER" section with a description and a "Request an Account" link. To the right is a "Login" form with fields for "Username:" and "Password:", a "Domain:" dropdown menu, and a "Login" button. Below the form is a "Warning Notice" section with a disclaimer. At the bottom of the page are links for "EPA Home", "Privacy and Security Notice", and "Contact Us".

Environmental Information | [Help](#) | [Contact Us](#)

Exchange Network | **SERVICES CENTER**

SERVICES CENTER

The Exchange Network Services Center is a web-based tool designed to allow Exchange Network users to easily send, get, and download information from other partners on the network.

Note: to access this tool, you must already have an Exchange Network user account assigned to you.

[Request an Account](#)

Login

Username:

* Username is required.

Password:

* Password is required.

Domain: default [Not sure?](#)

[Forgot Username or Password](#)

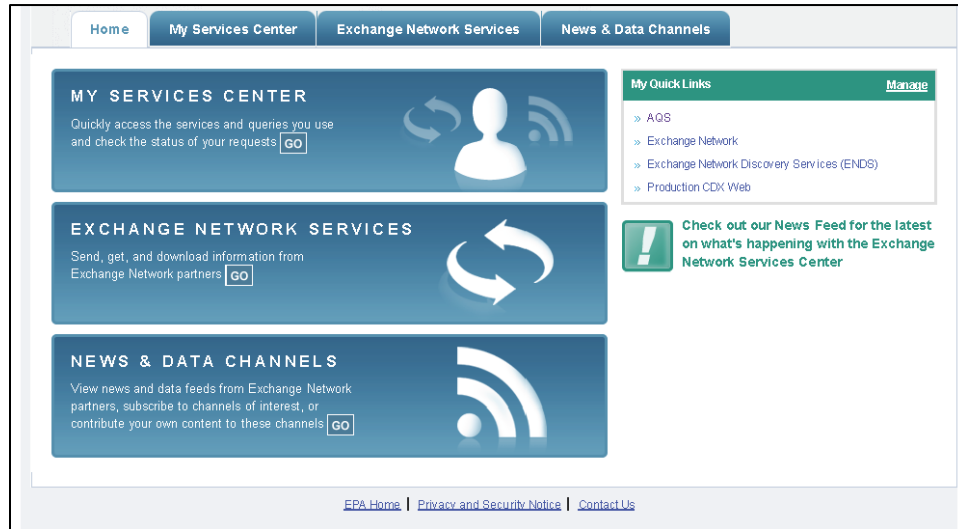
Warning Notice

This application is part of a United States Environmental Protection Agency (EPA) computer system, which is for authorized use only. Unauthorized access or use of this computer system may subject violators to criminal, civil, and/or administrative action. All information on this computer system may be monitored, recorded, read, copied, and disclosed by and to authorized personnel for official purposes, including law enforcement. Access or use of this computer system by any person, whether authorized or unauthorized, constitutes consent to these terms.

[EPA Home](#) | [Privacy and Security Notice](#) | [Contact Us](#)

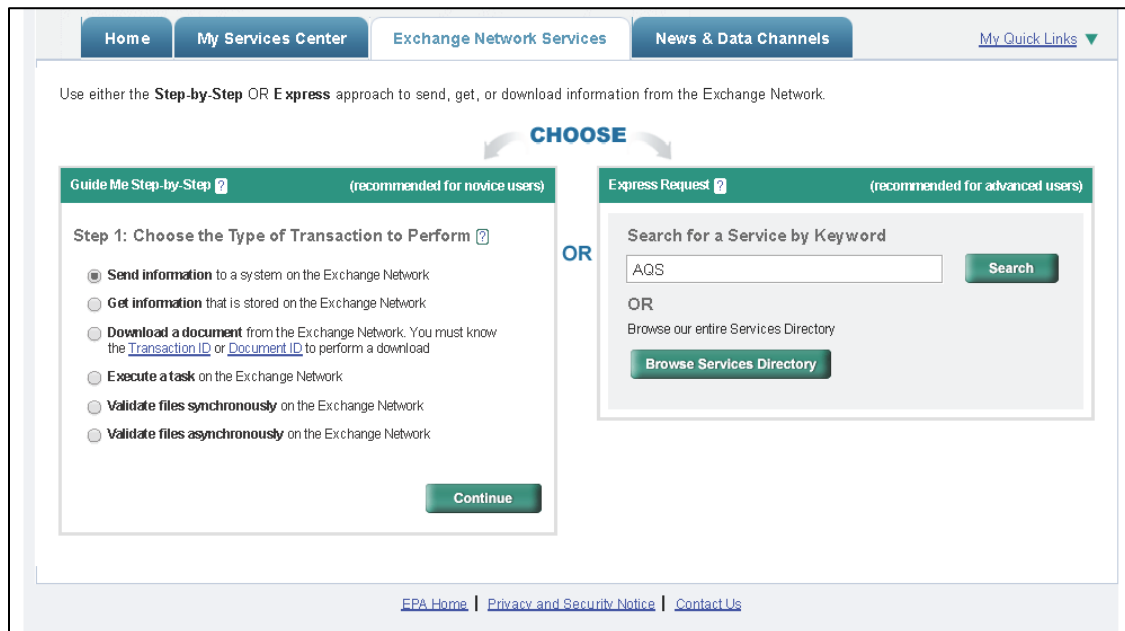
After logging in, the analyst will be presented with the home screen (Figure 3). The AQS service can be added to the analyst's home screen in the "My Quick Links" bar, or the analyst can search for the AQS submission form. To search, the analyst will need to click on the "Go" button of the Exchange Network Services bar.

Figure 3. Home screen of the Exchange Network Services website.



The analyst will then be presented with a choice between a Step-by-Step guide and a search bar. The analyst may type "AQS" into the search bar as shown in 4.

Figure 4. Enter "AQS" into the search bar.



The search results will show all available processes associated with the AQS system. The analyst should choose the Service that has "AQS Submit" specified in the "Service Name" field. This is typically the third choice in the list, as in 5, but may vary. This will take the analyst to the AQS submission form. Note that it is more efficient to add this service to the analyst's quick links on the home screen.

Figure 5. Search results from "AQS". The analyst should choose the Service named "AQS Submit".

The screenshot shows a web interface for a Services Directory. At the top, there are navigation tabs: Home, My Services Center, Exchange Network Services, and News & Data Channels. Below the tabs is a search bar with the text 'Filter By: Keyword(s) AQS' and buttons for 'Filter' and 'Clear'. The search results are displayed in a table with the following columns: Service Transaction, Dataflow, Service Name, Service Description, Node, and Service Provider. The table contains five rows of results, with the second row, 'AQS Submit', being the target service.

Service Transaction	Dataflow	Service Name	Service Description	Node	Service Provider
Get Info	AQDE	AQDERawData	Queries or Solicits the Raw Data for the AQDE Flow. The return is an XML file that conforms to the AQS Version 2.0 Schema.	NewJerseyNodeV1_Prod	NJDEP
Send Info	AQS	ProcessAQSDoc	Air Quality System Document Submissions	.NetNode2	U.S. Environmental Protection Agency
Send Info	AQS	AQS Submit	AQS Submit: Send files to the Air Quality System (AQS).	NGNProd2.0	U.S. Environmental Protection Agency
Get Info	AQS	GetAQSRawDataInsertByDate	AQS - GetAQSRawDataInsertByDate Service	NV	Nevada Division of Environmental Protection (NDEP)
Get Info	AQS	AQDEMonitorData	AQS - AQDEMonitorData Service	WA	Washington State Department of

Finally, the analyst will fill out the submission form with the analyst's email address, AQS user ID, screening group (PM2.5 Speciation), the file type (FLAT), the final processing step (LOAD), and whether or not to stop on errors (NO). See Figure 6 for an example. Use the "Choose File" button to select the file generated from the previous step. Press the "SEND DATA" button to submit the form. The progress of the data submission can be monitored through the same web portal.

Figure 6. AQS data submission form.

The screenshot shows a web interface for submitting AQS data. The main form area is titled "Express Request: AQS Submit" and contains several input fields:

- "Select a Document to Upload (max. size 1 GB):" with a "Choose File" button and "No file chosen" text.
- "Enter Sender's Email Address to Notify of Transaction Status Changes:" with an empty text box.
- "AQS User ID:" with an empty text box.
- "Additional Data Flow Specific Information:" section containing:
 - "Screening Group:" dropdown menu with "PM25 Speciation" selected.
 - "File Type:" dropdown menu with "FLAT" selected.
 - "Final Processing Step:" dropdown menu with "Post" selected.
 - "Stop On Error:" dropdown menu with "No" selected.
- A link: "Provide information (metadata) about this Document (recommended)".
- "Cancel" and "SEND DATA" buttons at the bottom.

The right sidebar, titled "You are currently using the following Service:", displays:

- Service Name: AQS Submit
- Description: AQS Submit: Send files to the Air Quality System (AQS).
- Transaction Type: Submit
- Dataflow: AQS
- Node: NGNProd2.0
- Publisher: U.S. Environmental Protection Agency
- A link: "Click here for Additional service help information"
- A link: "Select a different Service"

9. DATA DELIVERY FLOW

This section describes the data flow through the data delivery process used to execute all CSN validation checks. Figure 7 outlines the flow of data from the UCD validation results to final delivery to AQS. The function *generate_dart_file* is executed by the analyst to generate the initial data set for DART review (see 8.1). Then upon receipt of the validated data from DART, the data is re-ingested in the UCD CSN database into the *DARTreturn* table. The analyst then executes the *dart2aq*s function to reformat the data into AQS format and populate the *delivery.AqsStaging* and *delivery.AqsStagingQualifierCodes* tables. Any requested changes are made by the analyst in these two tables. The final dataset is prepared by executing the *aq*sStaging2export followed by the *generate_aqs_file* function. The data is reviewed again and submitted to AQS through the Exchange Network Services web portal (see 8.3). Source code for the functions shown in Figure 7 is stored in the *crocker* source repository.

Figure 7. Flow diagram of the data delivery process. Rectangles represent data files, diamonds represent R functions, circles represent databases, and lines represent inputs and outputs.



10. EQUIPMENT AND SUPPLIES

The associated hardware and software used for CSN data delivery are described in the associated UCD SOP #801.

11. QUALITY ASSURANCE AND QUALITY CONTROL

Software bugs and data management issues are tracked through JIRA tracking software. All users have access to our internal JIRA website and can submit, track, and comment on bug reports.

12. REFERENCES

Not Applicable.