

# UCD CSN Technical Instruction #801D

## CSN Data for DART

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

*July 31, 2021  
Version 1.4*

Prepared By: \_\_\_\_\_

Date: \_\_\_\_\_

Reviewed By: \_\_\_\_\_

Date: \_\_\_\_\_

Approved By: \_\_\_\_\_

Date: \_\_\_\_\_

### DOCUMENT HISTORY

<b>Date Modified</b>	<b>Initials</b>	<b>Section/s Modified</b>	<b>Brief Description of Modifications</b>
11/30/18	NJS	1, 2, 3, 7, 8, 9, 10, 11	Rewording for clarity and updating name changes.
7/31/19	DEY, KAG	8	Updated processes and added detail/clarity.
3/6/20	KAG	4-13	Wording changes for clarity.
4/1/21	DEY, AMM	1, 2, 3, 8, 9	Split previous TI document into two documents: DART-specific and AQS-specific; removed all AQS related processes. Added detail for clarity and added new processes.

## 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 .....	5
7. PERSONNEL QUALIFICATIONS, DUTIES, AND TRAINING.....	5
8. PROCEDURAL STEPS .....	5
8.1 Format data for DART Submission .....	5
8.2 Receive, Ingest, and Review Returned Data from DART .....	6
9. EQUIPMENT AND SUPPLIES .....	10
10. QUALITY ASSURANCE AND QUALITY CONTROL.....	10
11. REFERENCES .....	10

## LIST OF FIGURES

Figure 1. Returned DART datasets are ingested into the CSN database .....	7
---	---

## 1. PURPOSE AND APPLICABILITY

The subject of this technical instruction (TI) is the transfer of the resultant data from the Chemical Speciation Network (CSN) to and from the Data Analysis and Reporting Tool (DART). This document describes the procedures for preparing and delivering data to DART and the procedures for ingesting, processing, and reviewing data from DART.

## 2. SUMMARY OF THE METHOD

The University of California, Davis (UCD) analyst prepares a delivery file 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 and then addresses any comments and/or requests from the DART reviewers.

## 3. DEFINITIONS

- **AQS:** EPA's Air Quality System database.
- **Chemical Speciation Network (CSN):** EPA's PM<sub>2.5</sub> 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.
- **State, Local, and Tribal (SLT) Agencies:** Air agencies responsible for operating and maintaining a wide variety of ambient monitoring systems across the United States, including CSN.

## 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 and Reporting Group staff assigned to tasks described in this document have advanced training in data analysis and handling.

## 8. PROCEDURAL STEPS

Data delivery to DART 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 and ingest is performed by the UCD Data and Reporting Group on monthly batches of data (a calendar month of sample start dates). The transfer of CSN data to and from DART occurs in three steps:

1. Format data and submit to DART (Section (8.1)).
2. Receive and ingest returned data from DART (Section 8.2).
3. Review and perform suggested and requested actions from SLT validators (Section 8.2).

### 8.1 Format data for DART Submission

As detailed in TI 801C, results for elements, ions, and carbon are validated using the tools of the *datvalCSN* package. This includes assigning qualifier codes and null codes to specific records. Once validation is completed, the analyst will run the *csn\_calculate\_all* function from the *crocker* package to update data such as concentrations and MDLs where sample volumes or validity of field blanks changed during validation of the original dataset.

```
[month_data] <- crocker::csn_calculate_all([YYYY], [MM], 'production')
```

The analyst will post the data to the database using the following command, including the Results Set ID for the previous valid data set and the Analysis Quality Control (QC) Code that should be assigned to the previous Results Set.

```
[post] <- crocker::csn_post_results([month_data], 'production',  
AnalysisQcCode = 1, comment = ['Initial Posting'], replacingId =  
[previousResultsSetId], replacingQcCode = [PreviousResultSetNewQCcode])
```

The analyst will then run the *csn\_validate* function with the *write.flags* option in the following command set to *TRUE*<sup>1</sup>,

```
[monthData] <- datvalCSN::csn_validate(Month = ['MM'], Year = ['YYYY'], write.flags  
= [FALSE], )
```

---

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

Qualifier codes and null codes are assigned from various checks performed within this function. For records that have both null codes and qualifier codes assigned, only the null code can be delivered to DART. As such, qualifier codes for these records are posted to the UCD CSN database by specifying *write.flags = TRUE*. In addition, all null codes and qualifier codes for certain parameters delivered to AQS but not DART are also posted to the UCD CSN database by specifying *write.flags = TRUE*.

The command form shown here is appropriate for this stage in the process; more options can be included by changing defaults. The default recipient is “DART”, thus does not need to be specified in the command. To find out more information on this command, run *?csn\_validate* in the R environment. Qualifier codes and null codes are assigned when this function is run (see *UCD CSN SOP #801: Processing & Validating Raw Data*). The analyst will then populate the *delivery.DartExport* table in the UCD CSN database using,

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

This function can be used for posting data with different recipients (e.g. DART and AQS) to the UCD CSN database, although the default recipient is ‘DART’. Other options can be included in this command such as whether data are to be overwritten in the database. To find out more information on this command, run *?post\_output* in the R environment.

Next, the *generate\_dart\_file* function collects the necessary data from the UCD 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,  
                                out.dir = [output directory]).
```

Files are delivered to STI (for posting to DART) by UCD via a File Transfer Protocol (FTP) service. The file is prepared in .csv format and named as,  
DART\_YYYY-MM-DD\_to\_YYYY-MM-DD\_YYYYMMDDHHMM.UCDAVIS.csv  
Where the name indicates the sample collection date range for the records, as well as the date and time information for when the file was created at UCD.

## 8.2 Receive, Ingest, and Review Returned Data from DART

The state, local, and tribal (SLT) agencies have 30 days to review their associated data and perform Level 2 validation. Upon completion, the entire dataset is returned to UCD with a change log. UCD receives an automated email from STI with summary information for each batch of data including the percentage reviewed and the percentage with code changes. Three links to .csv files on the FTP server are also included:

- “Data File” for CSNData\_XXXX\_YYYYMMDD.csv (DART return file)
- “Statistics File” for SiteStats\_XXXX\_YYYYMMDD.csv
- “Comments QC Log File” for CSNCommentsQCLog\_XXXX\_YYYYMMDD.csv

Where XXXX denotes a number assigned at STI and YYYYMMDD denotes the date the file was created. If downloading the files using the default web browser (e.g. Google Chrome) is unsuccessful, try using Internet Explorer. The three files are downloaded and saved to the UCD shared drive at U:\CSN\FromSTI\To Be Imported. The DART return file is ingested using the DART import tool in the UCD CSN Management Site web application (csn.aqrc.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).

The DART return file is compared to the data that was delivered from UCD to DART. 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. The DART QC file is the CSV file containing the Comments QC Log that is comprised of comments and actions from validators created in DART.

Figure 1. Returned DART datasets are ingested into the CSN database using the CSN Management Site upload tool.

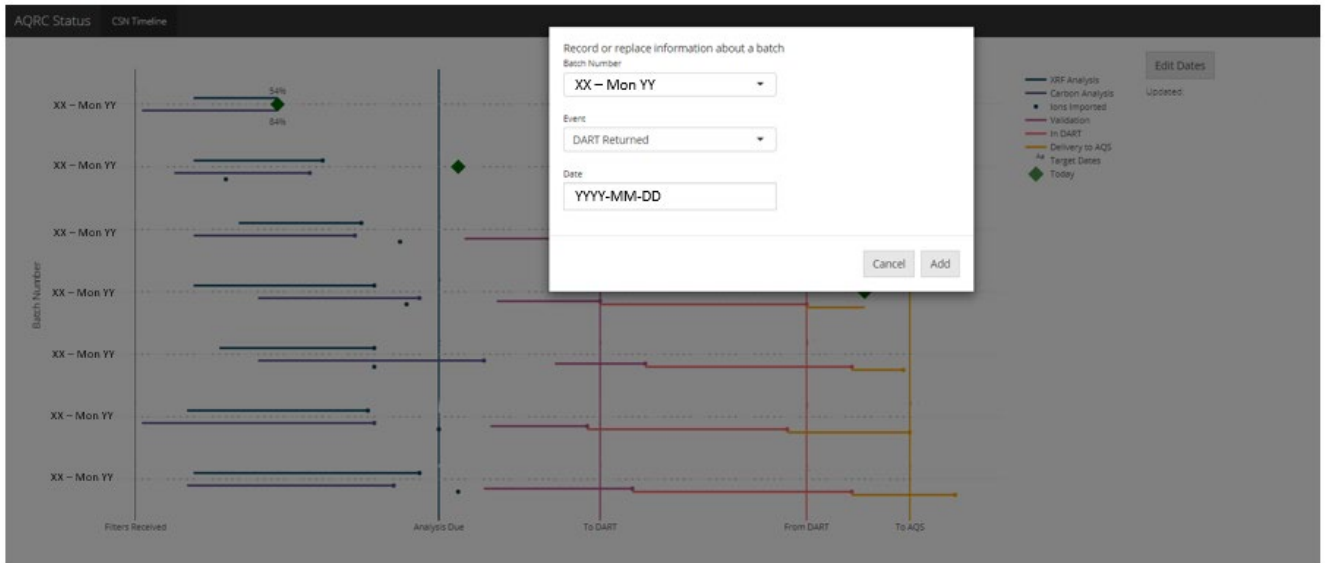
The screenshot shows a web browser window with the URL [csn.aqrc.ucdavis.edu/Import/UploadDartData](http://csn.aqrc.ucdavis.edu/Import/UploadDartData). The page title is "CSN Management Site" and the navigation menu includes "Home", "Analyses", "Import", and "Admin". The "Import" menu is active, and the "DART" option is selected. The main content area is titled "Upload DART data" and contains the following elements:

- A message: "The Export source is either a DartSet in the database or a File. If a File is chosen, the 'Set' field is ignored."
- A warning: "The Dart Ingest does not check for duplicate/existing data already in the database."
- Form fields:
  - Export source: Set**: A dropdown menu with the selected value "2010 - 2020 October (10/1/2020-10/30/2020) [id: 68]".
  - Export source: File**: A text input field with a "Choose File" button and the text "No file chosen".
  - DART return file**: A text input field with a "Choose File" button and the text "No file chosen".
  - DART QC file**: A text input field with a "Choose File" button and the text "No file chosen".
- TestOnly**: A checkbox that is checked, with the text "Runs through the import process but doesn't save the changes to the database." below it.
- A "Go" button.
- A "Messages (0)" section at the bottom.

At the bottom of the page, the footer text reads: "© 2021 - CSN Data Management Application (1.0.7642.18471)".

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. Once the data are ingested successfully to the database, the DART Return date information is to be entered in the CSN Timeline on the AQRC Status webpage (<https://shiny.aqrc.ucdavis.edu/aqrcstatus/>). Using the “Edit Dates” button, select the Batch Number, Event (either *DART Return* or *AQS Delivered*), and the Date (Figure 2). After adding this information, the visual timeline will be updated, indicating success.

Figure 2. AQRC Status webpage – CSN Timeline.



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

```
[dartChanges] <- datvalCSN::get_changes(dartSetId = [SetId])
```

where the *dartSetId* is a numeric value for the DART data set of interest. The command form shown here is typical for obtaining a dataframe containing change details made between the DART export and DART return datasets; more options can be included by changing defaults. The analyst can also specify a date range instead of the *dartSetId* to obtain the change log. To find out more information on this command, run *?get\_changes* in the R environment.

A comments log file (“Comments QC Log File”), in CSV format, is provided from STI along with the data file. From the comments log file, the analyst can obtain information regarding changes made in DART as well as comments left by the SLT validators (all actions in DART require a comment to be added). The comments log can be filtered to review changes made at a given site and date as well as to view contact details for the SLT validator. The ingest of the DART Return file and DART QC file will also filter these comments and list them, separated by Site IDs, as Ingest Messages.



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

```
[aqs] <- datvalCSN::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 creates a local copy of the data that will be used to populate the *delivery.AqsStaging* and *delivery.AqsStagingQualifierCodes* tables in the UCD CSN database. The following command can be used to post the data created in the previous command to the relevant tables in the UCD CSN database:

```
[aqs.staging.post] <- datvalCSN::post_output([aqs], recipient = 'aqs.staging')
```

Note that this same function was used for posting the DART export data to the database; the recipient has been changed accordingly for this stage in the data pathway. A label and comment are not applicable when posting data to the *delivery.AqsStaging* and *delivery.AqsStagingQualifierCodes* tables as there are no sets for AQS staging data. 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 sample volume value was changed;
- *update\_values*: 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 codes for records that are no longer invalid;
- *invalidate\_aqs\_filter*: applies a null code to a filter (may be parameter specific);
- *swap\_aqs\_records*: depending on the nature of the swap, this function swaps two sets of filter records by swapping dates between two samples, or by swapping the dates for the analytical species parameters and updating the concentrations as necessary;
- *re\_validate\_filter*: remove null code and reapply qualifier codes for a specific filter (may be parameter specific);
- *remove\_specific\_qual*: removes a specific qualifier code for a specific filter, parameter for a given filter, or group of parameters for a given filter;
- *add\_aqs\_qualifiers*: adds a specific qualifier code for a specific filter, parameter for a given filter, or group of parameters for a given filter;
- *change\_start\_dates*: updates the sample start date for a given filter or filter parameters;
- *update\_fromDF*: updates all records for a filter based on the data frame provided when running the function in a command.

When review of the data is complete and requested updates are finalized, the process of delivering the data to the final destinations will begin, as detailed in *UCD CSN TI #801E Data for AQS*.

## **9. EQUIPMENT AND SUPPLIES**

The hardware and software used for CSN data delivery are described in the associated *UCD CSN SOP #801: Processing & Validating Raw Data*.

## **10. 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.

## **11. REFERENCES**

Not applicable.