

# UCD CSN Technical Instruction #801E

## CSN Data for AQS Delivery

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

*July 31, 2021  
Version 1.0*

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 DART 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.....                                       | 5  |
| 6. INTERFERENCES .....                                 | 5  |
| 7. PERSONNEL QUALIFICATIONS, DUTIES, AND TRAINING..... | 5  |
| 8. PROCEDURAL STEPS .....                              | 5  |
| 8.1 Format data for AQS Submission .....               | 5  |
| 8.2 Upload data to CIA Database.....                   | 10 |
| 8.3 Deliver Data to NPS.....                           | 11 |
| 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. Login screen for the EPA's Exchange Network Services Center website..... | 6  |
| Figure 2. Home screen of the Exchange Network Services Center website.....         | 7  |
| Figure 3. Enter "AQS" into the search bar.....                                     | 7  |
| Figure 4. Search results from "AQS".....   | 8  |
| Figure 5. AQS data submission form.....  | 9  |
| Figure 6. AQRC Status webpage – CSN Timeline.....                                  | 9  |
| Figure 7. CIA Data Upload Portal main page.....                                    | 10 |
| Figure 8. CIA Data Upload Portal upload page.....                                  | 11 |
| Figure 9. Flow diagram of the data delivery process.....                           | 12 |

## 1. PURPOSE AND APPLICABILITY

The subject of this technical instruction (TI) concerns the final delivery of the resultant data from the Chemical Speciation Network (CSN). This document describes the procedure for preparing and delivering data to the Air Quality System (AQS) database, the CSN and IMPROVE Archive Database (CIA), and the National Park Service (NPS).

## 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 reformatted from the Data Analysis and Reporting Tool (DART) format into AQS format for delivery. The final data file is checked for correctness and then submitted to the Environmental Protection Agency's (EPA) AQS database. The Sample data are subsequently ingested into the CIA Database and delivered to NPS.

## 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.
- **CSN and IMPROVE Archive (CIA) Database:** A database of the complete record of CSN and IMPROVE data coupled with a web-based visualization and analysis tool.
- **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.
- **Interagency Monitoring of Protected Visual Environments (IMPROVE):** Federal PM<sub>2.5</sub> and PM<sub>10</sub> sampling network directed by the National Park Service, with sites located principally in remote rural areas.
- **NPS:** National Park Service.
- **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 and Reporting Group staff assigned to tasks described in this document 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 and Reporting Group on monthly batches of data (a calendar month of sample start dates). Final delivery occurs in three steps:

1. Prepare and submit to AQS (Section 8.1).
2. Upload to CIA database (Section 8.2).
3. Deliver Data to NPS (Section 8.3).

### 8.1 Format data for AQS Submission

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])
```

And post using another variation of the *post\_output* function:

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

The analyst will then prepare the final delivery files using the *generate\_aqs\_file* function, once for sample data and once for field blank data, as the formats of the two files are different. The analyst can review the final data before making the file by setting *make.file* to FALSE in the following commands:

```
[aqs] <- generate_aqs_file([dartSets], out.dir = ['U:/CSN/Delivered/AQS/'],  
data.type = ['SAMPLE'], make.file = TRUE)
```

```
[aqs] <- generate_aqs_file([dartSets], out.dir = ['U:/CSN/Delivered/AQS/'],  
data.type = ['FIELD'], make.file = TRUE)
```

with one or more appropriate DART data set IDs. For both the *aqsStaging2export* and *generate\_aqs\_file* 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, including confirming that these cases do not exist:

- records with values and null codes;
- records with a null code and qualifiers; and,
- records with no value and no null.

AQS reference documents are located on the shared network drive at U:/CSN/Documentation/Reference. For updated versions, visit <https://www.epa.gov/aqs>.

Once the AQS file is ready for delivery, the analyst will use a web browser to navigate to the EPA's Exchange Network Services website (Figure 1; <https://enservices.epa.gov/login.aspx>).

Figure 1. Login screen for the EPA's Exchange Network Services Center website.

Environmental Information | **exchange Network SERVICES CENTER** | [Help](#) | [Contact Us](#)

### 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:  [Not sure?](#)

[Forgot Username or Password](#)

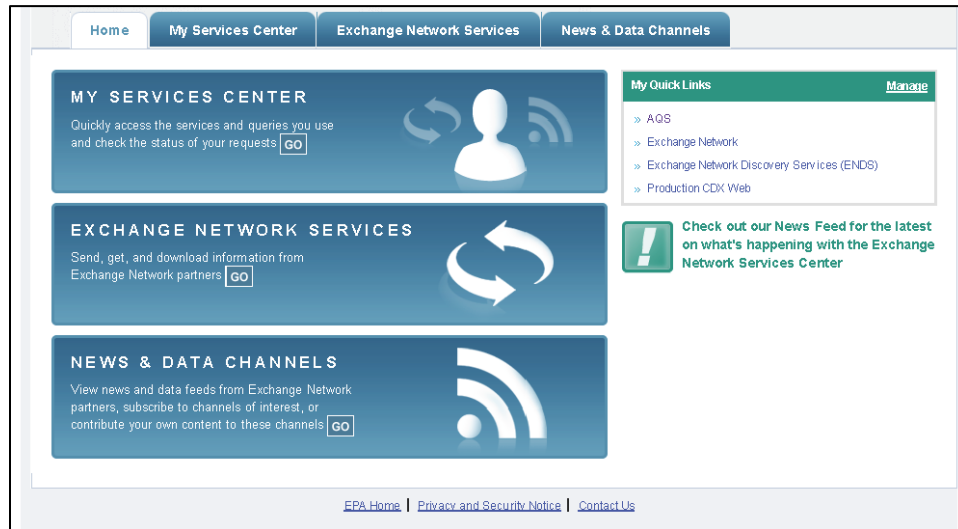
**Warning Notice**

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](#)

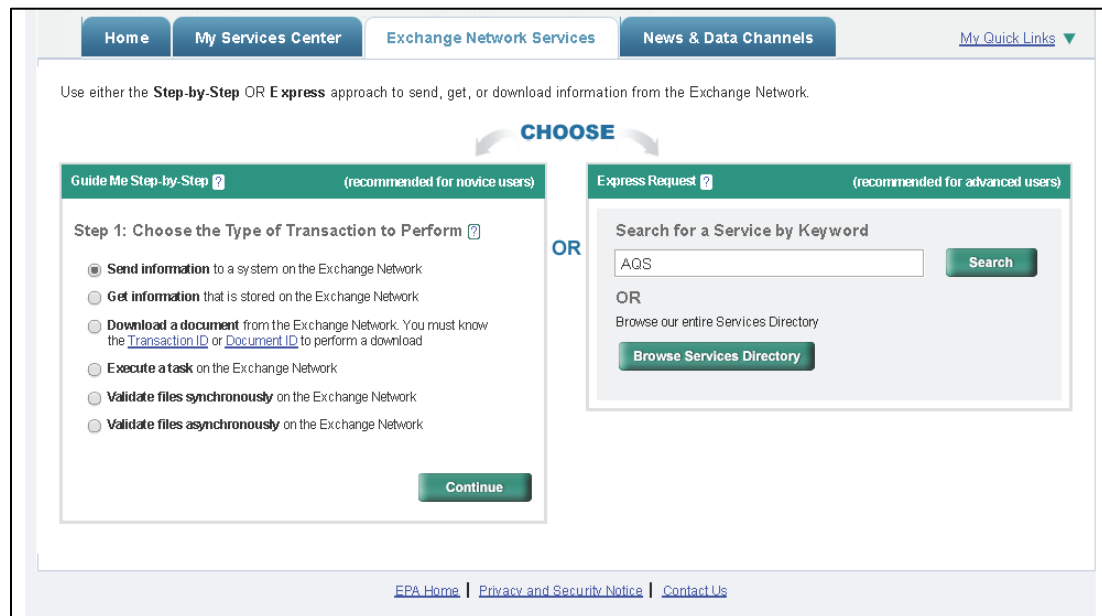
After logging in, the analyst will be presented with the home screen (Figure 2). To be efficient, the analyst should add the AQS service to the “My Quick Links” bar on their home screen. However, it is also possible for the analyst to search for the AQS submission form. To search, the analyst should click on the “Go” button of the Exchange Network Services bar.

Figure 2. Home screen of the Exchange Network Services Center website.



The analyst will then be presented with a choice between a Step-by-Step guide and a search bar (Figure 3). The analyst should type “AQS” into the search bar.

Figure 3. Enter "AQS" into the search bar.



The search results will show all available processes associated with the AQS system (Figure 4). To access the AQS submission form, the analyst should choose the service that has “AQS Submit” specified in the “Service Name” field (usually the third option listed).

Figure 4. 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)" and a dropdown menu showing "AQS". There are "Filter" and "Clear" buttons. Below the search bar, there is a table with 6 columns: Service Transaction, Dataflow, Service Name, Service Description, Node, and Service Provider. The table contains 5 rows of results. The second row, "AQS Submit", is the target service. At the bottom of the page, there are links for "EPA Home", "Privacy and Security Notice", and "Contact Us".

| Service Transaction       | Dataflow | Service Name              | Service Description  | Node                 | Service Provider                                   |
|---------------------------|----------|---------------------------|--|----------------------|--|
| <a href="#">Get Info</a>  | 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  |
| <a href="#">Send Info</a> | AQS      | ProcessAQSDoc             | Air Quality System Document Submissions  | NetNode2             | U.S. Environmental Protection Agency               |
| <a href="#">Send Info</a> | AQS      | AQS Submit                | AQS Submit: Send files to the Air Quality System (AQS).  | NGNProd2.0           | U.S. Environmental Protection Agency               |
| <a href="#">Get Info</a>  | AQS      | GetAQSRawDataInsertByDate | AQS - GetAQSRawDataInsertByDate Service  | NV                   | Nevada Division of Environmental Protection (NDEP) |
| <a href="#">Get Info</a>  | AQS      | AQDEMonitorData           | AQS - AQDEMonitorData Service  | WA                   | Washington State Department of                     |

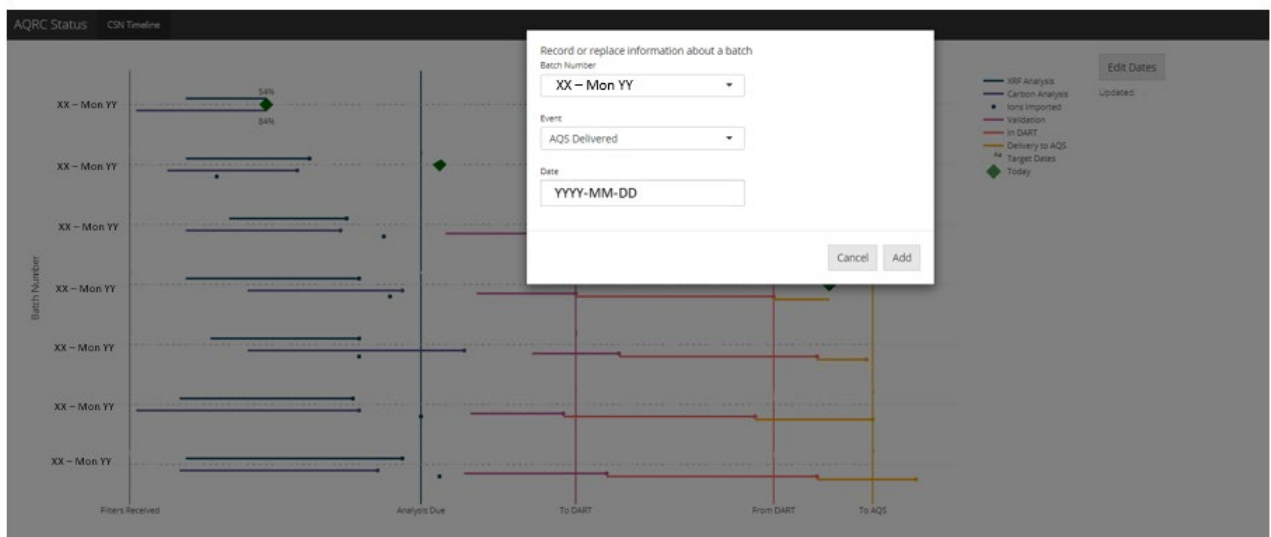
Finally, the analyst will fill out the submission form with their email address, AQS user ID, screening group (PM2.5 Speciation), the file type (FLAT), the final processing step (LOAD), and whether or not to stop on errors (NO). See Figure 5 for an example. Use the "Choose File" button to select the sample data 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. The analyst should repeat this process to submit the field blank data file generated in the previous step.



Figure 5. AQS data submission form.

Once the data are delivered successfully to AQS, the delivery 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 6). After adding this information, the visual timeline will be updated, indicating success.

Figure 6. AQRC Status webpage – CSN Timeline.



Requests may be made by SLT agencies for data to be updated in AQS. Requests made within six months of the initial AQS delivery date are to be accommodated. First, update the data in the *delivery.AqsStaging* table using the functions from *datvalCSN* mentioned in Section 8.2 of *UCD CSN TI #801D CSN Data for DART* to aid the process. Next, run the *aqsStaging2export* function from Section 8.1 above, and set *action* to “update”. Within the R Studio environment, use the *dplyr::filter* function to reduce the file to only the records to be updated in AQS. Then, post this selection of data to *delivery.AqsExport* same as above. The analyst creates a file of only the data to be updated in AQS by setting the *action* to “U”:

```
[aqs] <- generate_aqs_file([dartSets], out.dir = ['U:/CSN/Delivered/AQS/'],  
data.type = ['SAMPLE'], action = "U", make.file = TRUE)
```

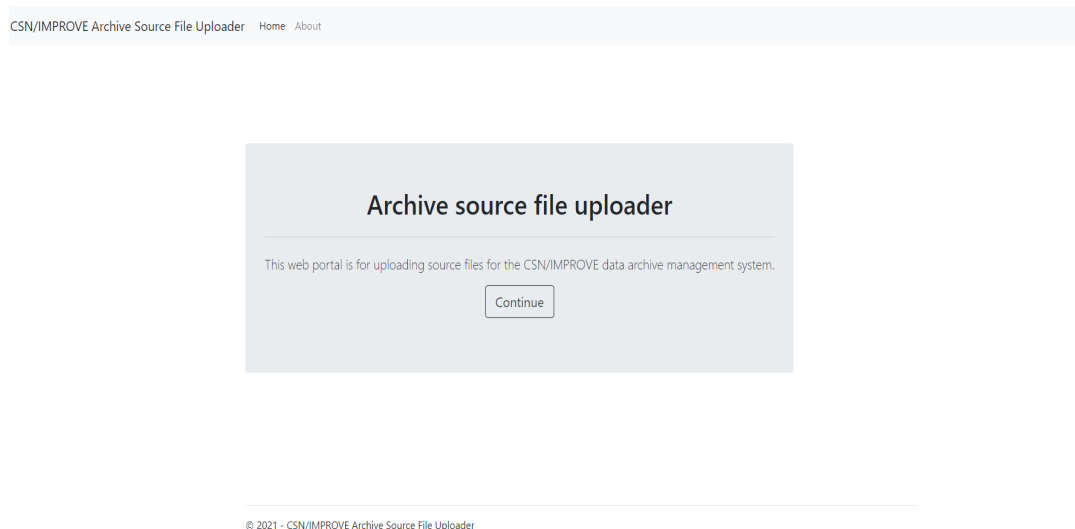
```
[aqs] <- generate_aqs_file([dartSets], out.dir = ['U:/CSN/Delivered/AQS/'],  
data.type = ['FIELD'], action = "U", make.file = TRUE)
```

The analyst can follow the same steps described in this section to deliver this data file to AQS.

## 8.2 Upload data to CIA Database

Data that has been successfully uploaded to AQS should be uploaded to the CIA database. Only sample data is to be uploaded; field blank data is not delivered to the CIA database. The analyst will use a web browser to navigate to the CIA Upload portal (Figure 7; <https://cia-uploadportal.azurewebsites.net/>).

Figure 7. CIA Data Upload Portal main page.



After continuing to the next page (Figure 8) by clicking on ‘Continue’, the analyst will select the CSN Network option and click on ‘Browse’ under ‘Select Source File’ to navigate to the sample data file to upload. The analyst will click on the ‘SUBMIT’ button

to upload the data. After successful submission, a message is displayed to note the file has passed Level 1 validation.

Figure 8. CIA Data Upload Portal upload page.

CSN/IMPROVE Archive Source File Uploader Home About

### Please specify network and select source file

Select Network:

CSN

IMPROVE

Select Source File:

Browse...

Click continue to submit and validate source file

|

© 2021 - CSN/IMPROVE Archive Source File Uploader

Whenever a new site is added, the sites table in the CIA database needs to be updated. The request should be directed to the Software Group. If data is delivered prior to the site being added to the CIA database, a message will be displayed to note a site is not in the database after submitting the file. The Software Group is to be informed. The data will stage but will not migrate. No further action is required from the analyst. Once the relevant sampler information is added to the CIA database, the staged data will successfully be posted to the CIA database when the SQL query to migrate data is next run (typically every night).

### 8.3 Deliver Data to NPS

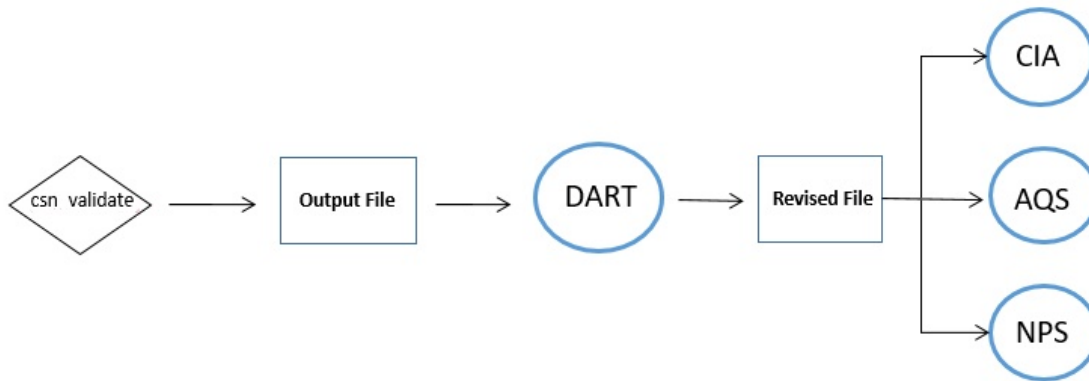
For the last step in the delivery process, the analyst will email the Sample Data file to NPS. The delivery happens by emailing Bret Schichtel and Shawn McClure and CCing Scott Copeland and members of the UCD AQRC group including the Associate Director of Quality Research, the Associate Director of Data & Software, the Program Manager, and the Lead QA Officer.

## 9. DATA DELIVERY FLOW

This section describes the data flow through the data delivery process used to execute all CSN validation checks. Figure 9 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 Section 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 *dart2aqs* 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 *aqsStaging2export* followed by the *generate\_aqs\_file* function. The data is reviewed again and submitted to AQS through the Exchange Network Services Center web portal (see Section 8.3). Source code for the functions shown in Figure 9 is stored in the *crocker* source repository.

Figure 9. 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 hardware and software used for CSN data delivery are described in the associated *UCD CSN SOP #801: Processing & Validating Raw Data*.

## 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.