

UCD CSN Technical Instruction #801A

CSN Data Ingest

*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

| Date Modified | Initials | Section/s Modified | Brief Description of Modifications |
|----------------------|-----------------|---------------------------|--|
| 11/30/18 | NJS | 1,2,7,8,9,10 | Rewording for clarity and updating name changes. |
| 7/30/19 | YN, KAG | 1,8 | Changed process for carbon and ion analysis pathways. Wording changes for clarity. |
| 3/5/20 | KAG | 5-11 | Wording changes for clarity. |
| 3/31/21 | DEY, AMM | 8 | Added steps for lab blank data and replicate ions analysis data. Added information on revised ions data. |
| | | | |
| | | | |

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 | 5 |
| 8. Procedural Steps..... | 5 |
| 8.1 Filter Records, Sample Operational Data, Lab Blanks, Null and Validity Flags..... | 5 |
| 8.2 Ion Analysis Results..... | 7 |
| 8.2.1 Ions..... | 8 |
| 8.3 Elemental and Carbon Analysis Results | 9 |
| 8.3.1 Elemental Analysis | 9 |
| 8.3.2 Carbon Analysis..... | 9 |
| 8.4 Mass Data..... | 9 |
| 8.5 Ingesting Revised Files | 10 |
| 9. Equipment and Supplies | 11 |
| 10. Quality Assurance and Quality Control..... | 11 |
| 11. References..... | 11 |

LIST OF FIGURES

| | |
|---|----|
| Figure 1. Filter data upload page from the UCD CSN Data Management website..... | 6 |
| Figure 2. Null code and validity code upload page. | 7 |
| Figure 3. Ion analysis results upload page. | 8 |
| Figure 4. Mass analysis results upload page..... | 10 |
| Figure 5. Mass analysis results upload page..... | 11 |

LIST OF TABLES

| | |
|---|----|
| Table 1. Automated validity checks performed by the UCD CSN Data Management website during the filter data upload process. | 6 |
| Table 2. Automated validity checks performed during the null code and validity flag upload process..... | 7 |
| Table 3. Automated validity checks performed during the ions analysis results upload..... | 9 |
| Table 4. Automated validity checks performed during the mass analysis results upload. | 10 |

1. PURPOSE AND APPLICABILITY

The subject of this technical instruction (TI) is handling electronic filter and laboratory records from samples collected in the Chemical Speciation Network (CSN). This document is intended to guide users on the receiving and validating of CSN filter and laboratory records and ingestion to the UCD CSN database. These include sample operational data and filter records from Wood PLC (Wood), ion analysis results from RTI International (RTI), and elemental and carbon analysis results from the University of California, Davis (UCD).

2. SUMMARY OF THE METHOD

Filter records from the Sample Handling Laboratory (Wood) and ion analysis results from RTI are received in data files, typically delivered as .csv files, and ingested into the UCD CSN database using the UCD CSN Data Management website. Elemental and carbon analysis results from UCD are automatically ingested.

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.
- **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.
- **Ion Chromatography (IC):** An analytical technique used to determine the concentration of ions.
- **Thermal Optical Analysis (TOA):** An analytical technique used to determine the concentration of carbon.
- **Energy Dispersive X-Ray Fluorescence (EDXRF):** An analytical technique used to determine the concentration of elements.

4. HEALTH AND SAFETY WARNINGS

Not applicable.

5. CAUTIONS

Not applicable.

6. INTERFERENCES

Not applicable.

7. PERSONNEL QUALIFICATIONS

The UCD Air Quality Research Center (AQRC) Data & Reporting Group staff assigned to tasks described in this document have advanced training in database programming and database management.

8. PROCEDURAL STEPS

Four data ingest steps are required prior to data processing and validation.

1. Filter records, including sample operational data, lab blank records, null and validity flags, from Wood. (Section 8.1)
2. Ion routine analysis results and replicate analysis results from RTI (Section 8.2).
3. Elemental and carbon analysis results from UCD (Section 8.3).
4. Mass analysis results from Wood (Section 8.4). Mass results are typically only available for specially designated sites.

These four steps are outlined below.

8.1 Filter Records, Sample Operational Data, Lab Blanks, Null and Validity Flags

Filter records are sent from Wood to UCD via email, typically on the same day as the shipment of corresponding physical filters. Filter records are delivered as four files:

1. FilterDataTransfer_[xxx].csv,
2. FilterDataNullFlags_[xxx].csv
3. FilterDataValidFlags_[xxx].csv,
4. LabBlanksDataTransfer_[xxx].csv

where [xxx] represents a number corresponding to the delivery batch. FilterDataTransfer contains a single record for each filter, including sample operational data (for example flow rate and ambient temperature). FilterDataNullFlags and FilterDataValidFlags include the null codes and validity codes, respectively. Null codes and validity codes are joined to corresponding filter data by the unique combination of SampleRequestID and ChannelID. LabBlanksDataTransfer contains a single record for each Lab Blank filter record sent with each shipment batch, with no sample operational data included. The files received from Wood are downloaded and stored on the UCD file server (U:\CSN\FromWood\To Be Imported).

Filter records (FilterDataTransfer_[xxx].csv and LabBlanksDataTransfer_[xxx].csv files from Wood) are ingested to the UCD CSN database through the UCD CSN Data Management website, one file at a time. Figure 1 shows a screenshot of the upload page. For each file, the analyst first loads in “test” mode (by having the “TestOnly” box checked as shown in Figure 1), which will perform import validation, but will not save any changes to the database. Filter records are subjected to the automated validity checks as shown in Table 1. The analyst will review the results of the validation and determine

which, if any, records fail to upload due to validation errors. Once the analyst has reviewed the output messages in the “test” mode, the upload should be performed again with the “TestOnly” box unchecked to ingest the data into the database. After upload, the analyst will store the source files on the UCD file server (U:\CSN\FromWood\Imported). If there are too many output messages and/or the analyst wants to review a reduced number of messages relating to specific columns, the analyst can run the ingest process in test mode and check various boxes in the “IgnoredColumns” section.

Figure 1. Filter data upload page from the UCD CSN Data Management website.

Table 1. Automated validity checks performed by the UCD CSN Data Management website during the filter data upload process.

| Check | Action |
|---|-----------------|
| Number of columns in header matches number of columns in row | Warning message |
| Any columns not found (or renamed) | Import aborted |
| Filter record matched more than one site or did not match any sites | Warning message |
| More than one batch found in the import | Warning message |
| Number/date columns fail to parse into number/date | Warning message |
| Existing records in the database; if multiple matches, generates message | Warning message |
| If matched existing record, checks for changed fields | Warning message |
| IntendedUseDate after SamplerStartDate | Warning message |
| SamplerStartDate more than a day after IntendedUseDate | Warning message |
| SamplerEndDate more than 25 hours after SamplerStartDate | Warning message |
| Calculated SamplerEventId does not match one in record | Warning message |
| SamplerEventId plus Channel position do not uniquely identify the record | Warning message |
| More than one LotNumber for PTFE filters in the import | Warning message |
| Flow and ambient parameters are out of specification | Warning message |
| Calculated sample volume is not within 10% of the reported sample volume | Warning message |
| FilterType column only consists of expected types (e.g. PTFE, nylon, quartz). | Warning message |

Null codes (FilterDataNullFlags_[xxx].csv file from Wood) and validity codes (FilterDataValidFlags_[xxx].csv file from Wood) are uploaded through the data management website as shown in Figure 2. Filter records must be loaded prior to the null and validity codes. The following steps are followed for both files (FilterDataValidFlags_[xxx].csv and FilterDataValidFlags_[xxx].csv). Files should first be loaded in “test” mode, which will perform import validation, but will not save any changes to the database. Null codes and validity codes are subjected to the automated validity checks as shown in Table 2. The analyst will review the results from the automated checks and investigate and resolve records that failed to upload because of validation errors. Similar to the previous step, the ingest process should be performed again with the “TestOnly” box unchecked. After ingest, the analyst will store the source files on the file server (U:\CSN\FromWood\Imported).

Figure 2. Null code and validity code upload page.

Table 2. Automated validity checks performed during the null code and validity flag upload process.

| Check | Action |
|--|-----------------|
| Number of columns in header matches number of columns in row | Warning message |
| Any columns not found (or renamed) | Import aborted |
| Code record matched more than one filter or did not match any filters | Warning message |
| SetNumber or IntendedUseDate do not match the matched filter record | Warning message |
| Number/date columns fail to parse into number/date | Warning message |
| Code does not match existing AQS Code | Warning message |
| Codes apply to more than one batch | Warning message |
| More than one null code applied to filter (also create FilterComment); ranks codes and marks extra codes as duplicates | Warning message |
| The same code is applied to a filter more than once | Warning message |
| NullCode import tries to use any non-terminal codes; QualifierCode import tries to use any terminal codes | Warning message |
| Filter is not marked invalid but a null code is assigned to the filter | Warning message |

8.2 Ion Analysis Results

Ion analysis results are sent from RTI to UCD via email in .csv format.

8.2.1 Ions

The ions data are delivered in two .csv files named after the batch of data that is being delivered with the replicate (RTI termed ‘duplicate’) analysis results file name ending with “_DUP”:

1. AXXXXXXXX.csv
2. AXXXXXXXX_DUP.csv

The Xs represent the batch number; for example, A0000053 for batch 53. All .csv analysis records are ingested to the database through the UCD CSN Data Management website, one file at a time. Figure 3 shows a screenshot of the IonsData upload page. For each file, the analyst first loads in “test” mode (by having the “TestOnly” box checked as shown in Figure 3), which will perform import validation, but will not save any changes to the database. Records are subjected to the automated validity checks as shown in Table 3. The analyst will review the results from the automated checks and investigate and resolve records that failed to upload because of validation errors. The ingest process should be performed again with the “TestOnly” box unchecked. After upload, the analyst will store the source files on the file server (U:\CSN\FromRTI\Imported).

Figure 3. Ion analysis results upload page.

CSN Management Site Home Analysis Data **Import** Admin Log off

Upload Filters Flags Mass Carbon **Ions** DART Contractor Flags FTIR

Upload RTI Ions Data

Old DRI import is here: [link](#)

Data file No file chosen

TestOnly Runs through the import process but doesn't save the changes to the database.

Filename:

Messages (0)

Table 3. Automated validity checks performed during the ions analysis results upload.

| Check | Action |
|--|-----------------|
| Number of columns in header matches number of columns in row | Warning message |
| Any columns not found (or renamed) | Import aborted |
| Number/date columns fail to parse into number/date | Warning message |
| Flag does not match existing AQS Code or extra codes (i.e. RS/RP) | Warning message |
| Existing records in the database; if multiple matches, generates message | Warning message |
| If matched existing record, checks for changed fields | Warning message |
| Record identifier did not match any filters | Warning message |

8.3 Elemental and Carbon Analysis Results

Elemental and carbon analyses are performed at the AQRC.

8.3.1 Elemental Analysis

Results files created by the PANalytical EDXRF software are automatically ingested on a schedule by a software service. The results files are transmitted to a directory on the PC connected to the PANalytical EDXRF analyzer (C:\PANalytical\Transmission). A Windows Service (internally named XRF Data Transfer) is installed on each individual PC connected to a PANalytical EDXRF analyzer and monitors the transmission directory checking it every hour for any files created. The results files are standard text files with the extension *.qan*. The file names are the EDXRF analysis dates and times in the format YYYYMMDDHHMMSS.qan. The results files and contents are parsed by the service and ingested into tables in the UCD CSN database.

8.3.2 Carbon Analysis

The software application used to run the carbon analyzer automatically stores data acquired during an analysis in a comma-delimited text format for later computation, display, and printing. The text file containing raw carbon data is automatically ingested into the UCD CSN database by a software service. Upon ingestion, the areal densities of OC (transmittance and/or reflectance), EC (transmittance and/or reflectance), and TC, as well as OC1, OC2, OC3, OC4, EC1, EC2, EC3 and OP (Pyrol C) (in $\mu\text{g C}/\text{cm}^2$) are automatically calculated and stored.

8.4 Mass Data

Filter masses for specific sites (typically specially designated sites) are determined at Wood and the results are sent to UCD via email as MassTransfer_[xxx].csv files, where [xxx] represents a number corresponding to the delivery batch. Mass analysis data is ingested to the UCD CSN database through the UCD CSN Data Management website. Figure 4 shows a screenshot of the upload page. The data analyst will first load in "test" mode (by having the "TestOnly" box checked as shown in Figure 4). The analyst will review the results from the automated checks and investigate and resolve records that failed to upload because of validation errors. After upload, the analyst will store the source files on the file server (U:/CSN/FromWood/Imported/Mass).

Figure 4. Mass analysis results upload page.

Table 4. Automated validity checks performed during the mass analysis results upload.

| Check | Action |
|---|-----------------|
| Number of columns in header matches number of columns in row | Warning message |
| Any columns not found (or renamed) | Import aborted |
| Mass record matched more than one filter or did not match any filters | Warning message |
| SetNumber or IntendedUseDate do not match the matched filter record | Warning message |
| Number/date columns fail to parse into number/date | Warning message |
| Checks for out of spec Weight Types | Warning message |
| Records from more than one batch | Warning message |
| More than one valid (Analysis QC code = 1) mass record per filter | Warning message |

8.5 Ingesting Revised Files

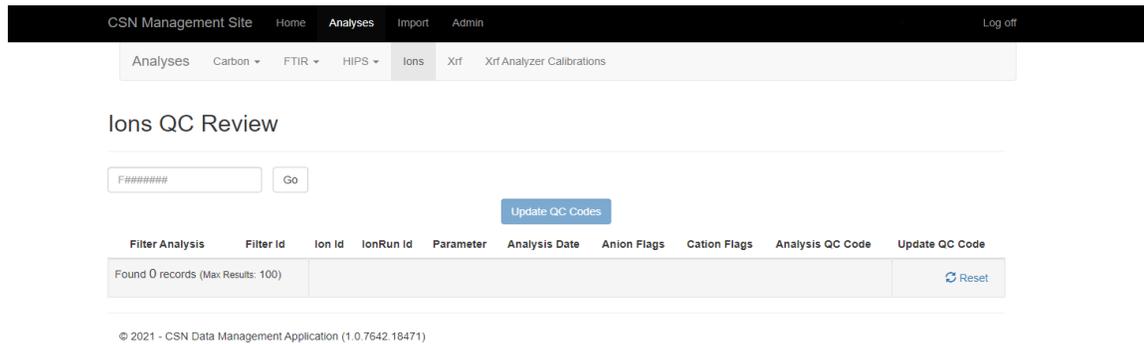
In the event that data corrections are made by Wood or RTI, new files will be provided for re-ingestion. The new files will be uploaded using the same processes described above. The ingest processing will identify any changed records in the case of the filter data files.

For revised filter data files from Wood, the analyst will first run the ingest process in “test” mode (by having the “TestOnly” box checked); to confirm what changes will be made, the analyst needs to also have the “OverwriteExisting” box checked. The analyst will scrutinize the changed records to ensure that they are correct before re-running the process in “overwrite” mode only (by having the “TestOnly” box unchecked and the “OverwriteExisting” box checked). Only changed records will be overwritten.

For revised Ions data files from RTI, the analyst will first run the ingest process in “test” mode and scrutinize the added records to ensure that they are correct before re-running the process (by having the “TestOnly” box unchecked). No records will be overwritten, only added to the database. Each record in the revised file should have an ‘RS’ flag

present and only the species requiring updating will have different values. The analyst will then use the Ions QC Review tool in the UCD CSN Data Management website to edit the Analysis QC Codes. Figure 5 shows a screenshot of the Ions Analysis QC Review page. Filtering by Filter Id, the Analysis QC Codes for each Ion will be updated so the confirmed valid reanalysis results are reported (by setting the analysis QC Code = 1) and the original results be invalidated (by setting the analysis QC Code = 0). For results which do not require changes, the reanalysis results with the 'RS' flag present and have the same values as originally provided results will be reported as Reanalysis results (by setting the analysis QC Code = 2).

Figure 5. Mass analysis results upload page.



9. EQUIPMENT AND SUPPLIES

The hardware and software used for CSN data ingest 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 the internal UCD JIRA website and can submit, track, and comment on issue reports.

11. REFERENCES

Not applicable.