

UCD IMPROVE Technical Information #351D

Data Delivery

*Interagency Monitoring of Protected Visual Environments
Air Quality Research Center
University of California, Davis*

*October 4, 2022
Version 1.0*

Prepared By:	<small>DocuSigned by:</small> <i>Indu Thekkemepilly Sivakumar</i> <small>19F6B63B1B17443...</small>	Date: <u>10/7/2022</u>
Reviewed By:	<small>DocuSigned by:</small> <i>Dominique Young</i> <small>BB55DBA34BAB407...</small>	Date: <u>10/7/2022</u>
Approved By:	<small>DocuSigned by:</small> <i>Marcus Langston</i> <small>0A10CFCE79B0452</small>	Date: <u>10/7/2022</u>

DOCUMENT HISTORY

Date Modified	Initials	Section/s Modified	Brief Description of Modifications
12/04/21	SRS	All	Document created.

TABLE OF CONTENTS

1. Purpose and Applicability.....	5
2. Summary of the Method	5
3. Definitions.....	5
4. Health and Safety Warnings	6
5. Cautions	6
6. Interferences.....	6
7. Personnel Qualifications	6
8. Equipment and Supplies	6
9. Procedural Steps.....	6
9.1 Submission to CIRA	6
9.2 Submission to AQS	8
9.2.1 File Generation and Checks	8
9.2.2 Submit Data to AQS	10
9.2.1 Address Data Errors in AQS.....	13
9.3 Submission to UCD CIA.....	17
9.4 Add Data Delivery Dates Delivery Dates to AQMT Status Page.....	18
9.5 Submission of Preliminary Data to CIRA.....	19
9.5.1 File Generation.....	20
10. Data and Records Management	21
10.1 Disaster Recovery Plan.....	21
10.1.1 Facility Recovery	22
10.1.2 Hardware Recovery Plan	22
10.1.3 Software and Data Recovery Plan	22
10.1.4 Data Security.....	22
11. Quality Assurance and Quality Control.....	23
11.1 Code Development	23
11.2 Bug Reporting.....	23
11.3 Data Validation.....	23
12. References.....	23

LIST OF FIGURES

Figure 1. Login screen for the EPA’s Exchange Network Services Center website.	10
Figure 2. Home screen of the Exchange Network Services Center website.....	11
Figure 3. Type AQS into the search bar.	11
Figure 4. Select the service named AQS Submit.....	12
Figure 5. AQS data submission form.....	13
Figure 6. EPA AQS Application login screen.	14
Figure 7. AQS screen after logging into the application	14
Figure 8. AQS pop-up screen after selecting Screening Group Access	15
Figure 9. AQS screen after selecting Batch from the options.....	15
Figure 10. AQS screen after selecting Correct -> Raw from the options	16
Figure 11. UCD CIA submission website home page.	17
Figure 12. UCD CIA data submission details page.	18
Figure 13. ‘IMPROVE’ Timeline tab on AQMT status page.....	19
Figure 14. File submission date update window.....	19

1. PURPOSE AND APPLICABILITY

The purpose of this technical information (TI) concerns the delivery of the resultant data from the Interagency Monitoring of Protected Rural Environments (IMPROVE). This document describes the procedure for preparing and delivering data to the Cooperative Institute for Research in the Atmosphere (CIRA)/Federal Land Manager Environmental Database (FED), the Air Quality System (AQS) database, and the CSN and IMPROVE Archive Database (CIA).

2. SUMMARY OF THE METHOD

The University of California, Davis (UCD) analyst prepares a delivery file of the validated IMPROVE data sets using custom tools in the *crocker* R package. Data are formatted into two FED formats and AQS format for delivery. The final data files are checked for correctness and then submitted to CIRA/FED via the National Park Service (NPS) and submitted to the Environmental Protection Agency's (EPA) AQS database. The Sample data are subsequently ingested into the CIA Database.

In addition, an unvalidated preliminary data set is to be processed and delivered to CIRA as single format data file on a routine basis.

3. DEFINITIONS

- **AQMT:** Air Quality Monitoring Team.
- **AQRC:** Air Quality Research Center.
- **AQS:** EPA's Air Quality System database.
- **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.
- **Chemical Speciation Network (CSN):** EPA's PM_{2.5} sampling network, with sites located principally in urban areas.
- **CIRA:** Cooperative Institute for Research in the Atmosphere.
- **crocker:** A custom software package in the R language that contains the data processing code used to produce, check, and post the final results.
- **CSV:** a comma-separated value file that is the common format for delivery files.
- **Energy Dispersive X-Ray Fluorescence (EDXRF):** An analytical technique used to determine the concentration of elements.
- **Federal Land Manager Environmental Database (FED):** a database of environmental data managed by Cooperative Institute for Research in the Atmosphere (CIRA)
- **Hybrid Integrating Plate/Sphere (HIPS):** An analytical technique for optical absorption.

- **Interagency Monitoring of Protected Visual Environments (IMPROVE):** Federal PM_{2.5} and PM₁₀ sampling network directed by the National Park Service, with sites located principally in remote rural areas.
- **IMPROVE database:** A SQL Server database that is the central warehouse of IMPROVE preliminary and final data at UCD.
- **Method Detection Limit (MDL):** A lower limit of detection specific to method of analysis and reported parameter.
- **NPS:** National Park Service.
- **PM:** Particulate Matter. PM_{2.5} is particulate matter with diameters 2.5 micrometers (µm) and smaller. PM₁₀ is particulate matter with diameters 10 µm or smaller.
- **SQL:** database management system used by AQRC.
- **UCD:** University of CA—Davis.

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. EQUIPMENT AND SUPPLIES

The hardware and software used for IMPROVE data ingest are described in the associated *UCD IMPROVE SOP #351: Data Processing & Validation*.

9. PROCEDURAL STEPS

After Level 2 data validation is complete, the data files are submitted to CIRA, AQS, and UCD CIA databases. The date of delivery is added to the AQMT Status IMPROVE Timeline page at <https://shiny.aqrc.ucdavis.edu/aqmtstatus/>.

9.1 Submission to CIRA

Export files for CIRA (FED) are created using the *improve_export_fed* and *improve_export_wide* functions in the *crocker* package, in which the year, month, and server for both functions are entered. The functions create “skinny” and “wide” versions of the dataset, and both are submitted. To generate the “skinny” format export file, run the following command in the R environment:

```
[FED] <- crocker::improve_export_fed(year = [YYYY], month = [MM], server =  
'production', write_file = TRUE)
```

To generate the “wide” format export file, run the following command in the R environment:

```
[FED_wide] <- crocker::improve_export_wide(year = [YYYY], month = [MM],  
server = 'production', write_file = TRUE)
```

By default, the *write_file* argument is TRUE, so if a normal delivery file is created, no extra step is needed.

The files are saved under *U:\IMPROVE\FED Export*, named ‘IMPROVE_Data_YYYY_MM_server’ and ‘IMPROVE_WideData_YYYY_MM_server’ (e.g., ‘IMPROVE_Data_2017_02_production’), respectively. These files are compressed into a zip folder and are emailed to the CIRA correspondent(s) as an attachment.

To export a subset of the data, in the file export code, set *write_file* = FALSE. Instead of writing the output file directly, the functions return the data frame to the user’s R environment. Using the *dplyr* package functions, the user should filter the results to only specific parameters and sites before writing the output. After creating the data frame with the selected subset of data, use the *write.csv* function to export the custom file as follows:

```
write.csv(df, file = [filename], na = '-999', row.names = [FALSE], quote =  
[FALSE])
```

The file name argument should specify the path to which you want to save the data as well as other pertinent information regarding the contents of the file. The user can specify the file name as a separate variable in their R environment before running the *write.csv* function. An example of the filename variable is as follows:

```
filename <- paste0('U:/IMPROVE/FED Export/', 'IMPROVE_Data_', year, '_', month,  
'_', server, '.csv')
```

The following checks are performed on the skinny format files:

- Open the CSV file and make sure the following columns exist: Id, FilterId, Sampler, ObjectiveCode, SampleDate, Status, Parameter, Value, Uncertainty, MDL, Unit, POC, ModuleTypeCode.
- In the CSV file, filter the Status column to 'UN' and review the parameters listed in the Parameter column: only PM₁₀ or PM_{2.5} should be listed. If non-mass parameters are listed, inform the software group to fix the issue.
- For the BYIS1 site, a value for the fAbs parameter is not reported for the routine module. To check this, filter the data in the Sampler column to 'BYIS1' and 'RT' in the ObjectiveCode column, and select 'fAbs' in the Parameter field. Review the data in the Value field and confirm values are reported as -999.
- For filters with pending analysis results, use the filter option in the CSV to select the particular filter Id from the FilterId column and confirm the values for all parameters for that filter are reported as -999 and only the 'NS' status is listed in the Status column.
- In the Status column, click on the filter button and inspect the available statuses. Make sure that the listed statuses are currently active.

The following checks are performed on the wide format files:

- Make sure the 'UN' status is only applied to the mass. This can be done by opening the CSV file and using the find and search option for 'UN'. Review the results and confirm 'UN' is only listed under the columns named PM₁₀_flag or PM_{2.5}_flag.

In the CSV file, spot check the number of rows per site equals the number of sample dates in that month. For example, for the month of February, nine rows are expected. Other months are expected to have 10 or 11 rows.

9.2 Submission to AQS

9.2.1 File Generation and Checks

AQS data export files are created using the *improve_export_aqs* function in the *crocker* package. To generate the AQS delivery file, run the following command in the R environment:

```
[aqs] <- crocker::improve_export_aqs([YYYY], [MM], server = "production", filename = [NULL], action = ["keep"], site = ["XXXXX"], param = [NULL], del_type = ["I"])
```

This command will generate a formatted text file suitable for AQS delivery containing all data for the year ([YYYY]) and month ([MM]) and save to the location specified in *filename*, with the default file name and path in the format of '[aqs_path]/AQSResultsOutput_[today]_[year]_[month]_[server].txt'. A typical command for routine monthly data delivery can be run as follows:

```
aqqs <- crocker::improve_export_aqs(2020, 04, 'production')
```

The function has the capability to generate an AQS delivery file down to the parameter level (*param*) and/or particular site(s) (*site*). If *action* is specified as “keep”, only the specified records are retained in the delivery file whereas “drop” will remove the specified records. In addition to generating delivery files to add data to the AQS database, the function can also be used to create delivery files to update or delete data within AQS by specifying the *del_type*. To display the helper documentation for the function, the user can run *?improve_export_aqs* in the R Studio console.

Once the file has been generated, various checks should be performed by running the following commands in R Studio:

- Check for existence of duplicate records:

```
duplicates <- find_duplicates(aqs, c('StateCode', 'CountyCode', 'SiteID', 'POC', 'SampleStartDate', 'AQSPParameterCode', 'AQSMMethodCode'))
```
- Check the number of unique parameters:

```
if(count(unique(aqs[c("AQSPParameterCode")])) == 45) {print('ok')}
```
- Check for existence of records with null values and no null code:

```
noVal.null <- aqs %>% dplyr::filter(Value == "" & NullCode == "")
```
- Check for existence of records with non-null values and a null code:

```
null.noData <- aqs %>% dplyr::filter(NullCode != "" & Value != "")
```
- Check for existence of records with a qualifier validity flag and a null code:

```
null.noData <- aqs %>% dplyr::filter(NullCode != "" & Qualifier1 != "")
```

In addition to checking the data file for issues that would result in a failed AQS delivery, the validator also reviews the data further to obtain information on the data set as a whole. The validator can compare this with similar information from previous months of data to determine if the current month of data is reasonable or if there is an unexpectedly large increase in the number of invalid records, for example.

- Count the number of records with a null value and null code:

```
noData <- aqs %>% dplyr::filter(NullCode != "" & Value == "")
```
- Count the number of invalid and valid records:

```
no.Nulls <- aqs %>% dplyr::group_by(NullCode)%>%  
dplyr::summarise(n_nulls = n())%>% dplyr::arrange(n_nulls)
```

For each dataset, the data validator keeps a record of the null codes reported and the number of records with each null code to put each month into context with previous months of data. The log file is located at U:\IMPROVE\Data_Validation\Status change list\AQS_null_code_count_per_month.xlsx.

- After running the command above, the validator can confirm the total number of invalid and valid records in the file matches the total number of records in the dataset by running the following command: `print(sum(no.Nulls$no_nulls))`
- Further, the number of invalid records (those with a null code) can be counted and compared with the results from the check for existence of records with non-null values and a null code:

```
no.Nulls.total <- no.Nulls %>% dplyr::filter(NullCode != "")
print(sum(no.Nulls.total$no_nulls))
```

9.2.2 Submit Data to AQS

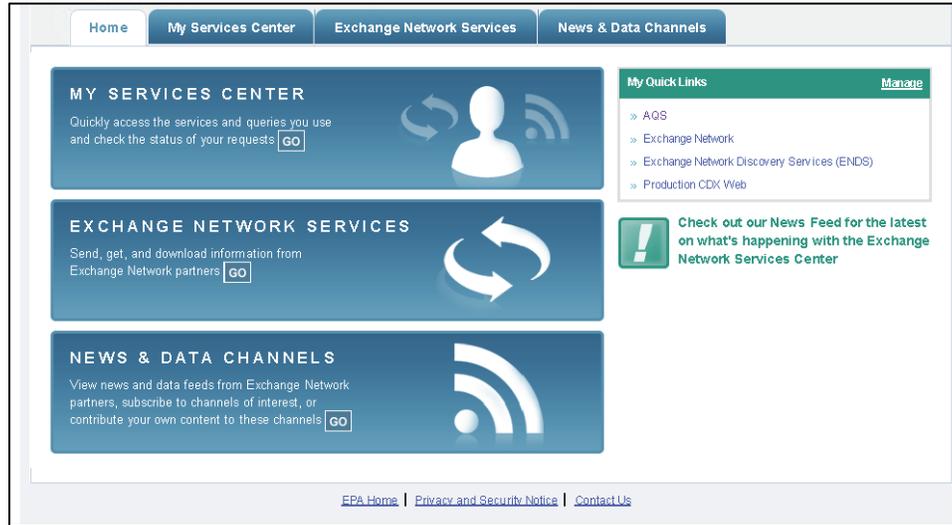
Once the checks have been completed, the data can be delivered to AQS. To submit batch data files to AQS, open a web browser and navigate to the EPA Exchange Network Services website, <https://enservices.epa.gov/login.aspx> (Figure 1). Use credentials to login.

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

The screenshot shows the login interface for the EPA Exchange Network Services Center. At the top left, there is an 'Environmental Information' link and the 'exchange Network' logo. The main header reads 'SERVICES CENTER'. On the right, there are links for 'Help' and 'Contact Us'. The central content area is titled 'SERVICES CENTER' and includes a brief description of the tool, a note about account requirements, and a 'Request an Account' link. Below this are three icons: a circular refresh icon, a person icon with a signal wave, and a standard RSS feed icon. A 'Warning Notice' section at the bottom contains a disclaimer about the system's use. On the right side, a 'Login' form is displayed with the following fields: 'Username:' (text input), 'Password:' (text input), and 'Domain:' (dropdown menu with 'default' selected and a 'Not sure?' link). A 'Login' button and a 'Forgot Username or Password' link are also present in the form area.

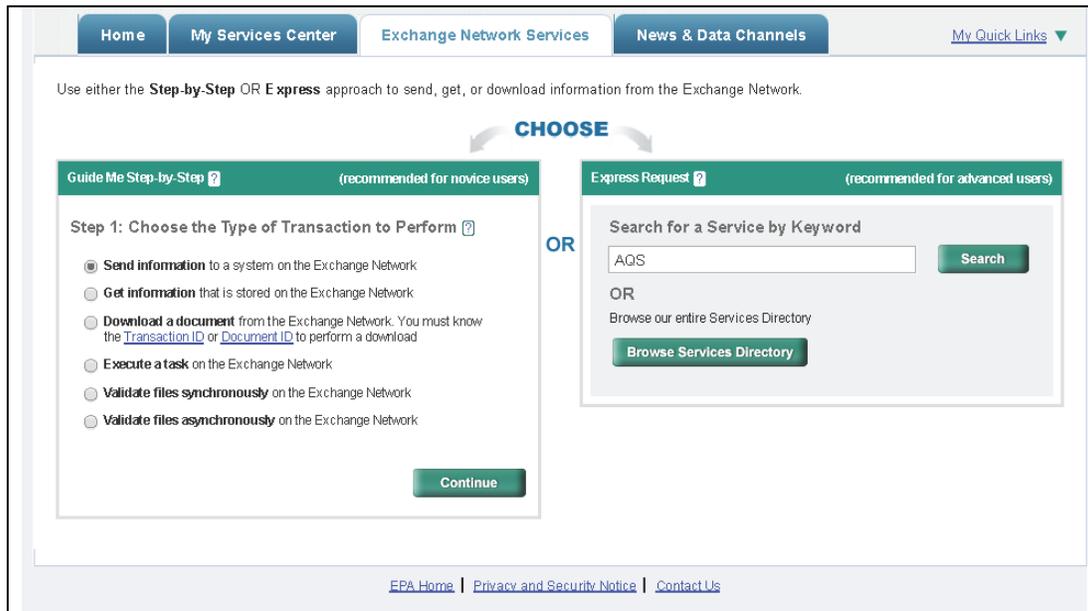
Following login, the home screen is accessed (Figure 2). For efficiency, add the AQS service to the home screen **My Quick Links** bar; however, it is also possible to search for the AQS submission form. To search, use the **Go** button of the **Exchange Network Services** bar.

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



Next, the option for a step-by-step guide and a search bar presented (Figure 3); type **AQS** into the search bar.

Figure 3. Type 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, choose the service that has **AQS Submit** specified in the **Service Name** field (usually the third option listed).

Figure 4. Select the service named AQS Submit.

The screenshot shows the 'Services Directory' page with a search filter set to 'AQS'. The search results table lists several services, with 'AQS Submit' selected. The table columns are: Service Transaction, Dataflow, Service Name, Service Description, Node, and Service Provider.

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	NUDEP
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

Fill out the submission form, specifying email address, AQS user ID, screening group (IMPROVE), the file type (FLAT), the final processing step (POST), and whether to stop on errors (YES). See Figure 5 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. Monitor progress of the data submission through the web portal. If the submission to AQS is fully successful, no further actions are necessary; the data have been delivered.

Figure 5. AQS data submission form.

9.2.1 Address Data Errors in AQS

If there are data records that are considered by AQS as statistical evaluation issues or critical review issues, the submission process will stop, despite there being no data errors. The data will only have been staged in AQS and needs to be posted. The user will receive two emails at the email address entered during the submission (Figure 5): one from AQSDEV and one from cdx@epa.gov, which include details of the data submission. The AQSDEV email also includes other details such as links to the load report and statistical and critical review report. Once the user has reviewed the emails to confirm the data are ready to be posted as no row failed to load, the user should log into AQS directly to post the relevant data.

To do this the user should follow these steps:

- Launch the AQS Application at <https://www.epa.gov/aqs> and enter their AQS credentials (Figure 6);
- Select the Screening Group Access option (Figure 7);
- Select IMPROVE (Figure 8);
- Go to Batch;
- Highlight the relevant row for the submitted file on the Process by File tab;
- Click on the 'Post File' button.

Figure 6. EPA AQS Application login screen.

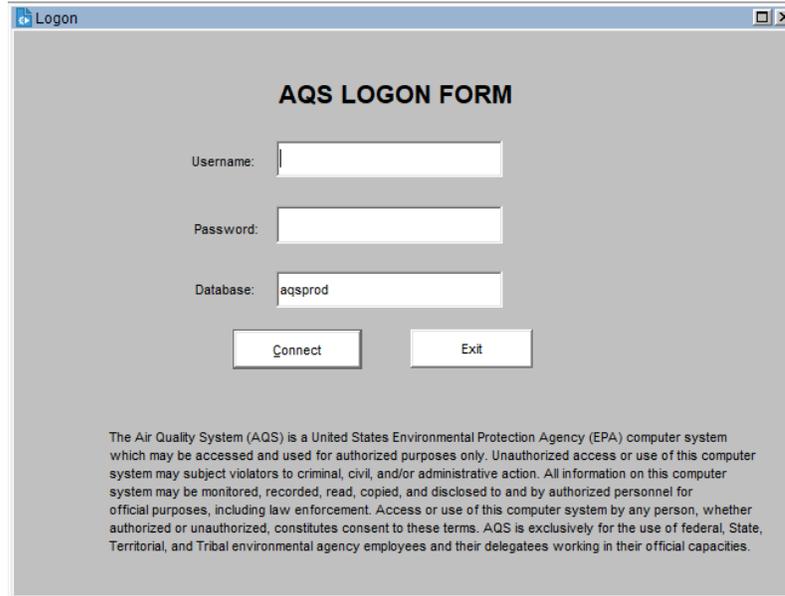


Figure 7. AQS screen after logging into the application; select the Screening Group Access option.

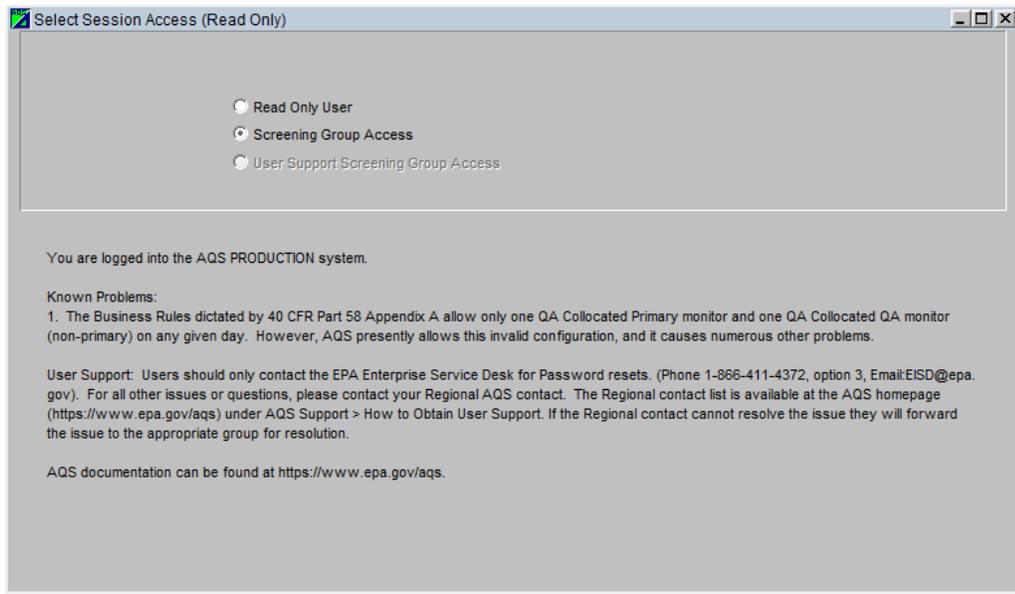


Figure 8. AQS pop-up screen after selecting Screening Group Access; select IMPROVE screening group.

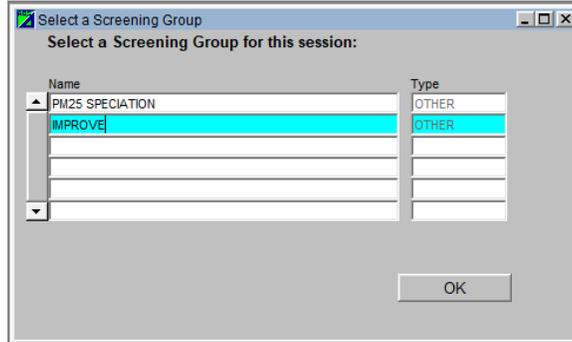


Figure 9. AQS screen after selecting **Batch** from the options. Click on any field of the row for the file which is to be posted. Files to be posted will show a Process Status of 'LOAD-COMPLETED' and the Skip'd Monitors and Records Posted fields will be empty.

HISTORY AND STATUS

Submission Date	File Name	User Name	Records In File	Date (last)	Process Status	LOAD			POST		
						Recs Loaded	Recs Not Loaded	Stat/CR Finding Count	Records to Post	Skip'd Monitors	Records Posted
20220224 16:37	AqsResultsOutput_20220224		65160	20220225 00:25	LOAD-COMPLETED	65160	0	510	65160		
20220223 20:10	AqsResultsOutput_20220223		79640	20220224 01:10	POST-COMPLETED	79640	0	669	0	0	79640
20220215 14:29	AqsResultsOutput_20220128		72400	20220215 15:12	POST-COMPLETED	72400	0	503	0	0	72400
20220215 13:29	AqsResultsOutput_20220128		72400	20220215 14:11	POST-COMPLETED	72400	0	499	0	0	72400
20220215 12:24	AqsResultsOutput_20220128		72400	20220215 13:13	POST-COMPLETED	72400	0	398	0	0	72400
20220214 19:11	AqsResultsOutput_20220208		72400	20220214 19:59	POST-COMPLETED	72400	0	495	0	0	72400
20220214 17:48	AqsResultsOutput_20220208		79640	20220214 18:40	POST-COMPLETED	79640	0	370	0	0	79640
20220214 16:24	AqsResultsOutput_20220208		71995	20220214 17:19	POST-COMPLETED	71995	0	456	0	0	71995
20220214 14:25	AqsResultsOutput_20220208		71950	20220214 15:53	POST-COMPLETED	71950	0	412	0	0	71950
20220214 11:42	AqsResultsOutput_20220214		71950	20220214 13:41	POST-COMPLETED	71950	0	281	0	0	71950
20220207 16:34	AqsResultsOutput_20220207		71950	20220207 20:27	POST-COMPLETED	71950	0	55	0	0	71950

PROCESS CONTROL

Process selected file through:

Results and Reports:

PROCESS FLOW

```

graph TD
    FILE((FILE)) --> ENSC[ENSC]
    ENSC --> Stage[Stage]
    Stage --> Transfer{Transfer Error?}
    Transfer -- N --> Load[Load]
    Transfer -- Y --> Diagnose[Diagnose and fix problem]
    Diagnose --> FILE
    Load --> Errors{Errors?}
    Errors -- Y --> Correct[Correct]
    Correct --> Load
    Errors -- N --> RawData{Raw Data?}
    RawData -- Y --> Post[Post]
    RawData -- N --> Done((DONE))
    Post --> Done
    Done -.-> LoadSummary[Load Summary and Errors]
    Done -.-> RawDataInventory[Raw Data Inventory]
    Done -.-> StatCR[Statistical and Critical Review (Stat CR)]
  
```

If there are data records with errors, the submission process will stop. Again, the user will receive two emails. In the case of data errors, the AQSDEV email will also include a link to a text file containing the records that failed to load. The user should review the data in the text file along with the load summary and detail report to determine the cause of the error and determine how to resolve the issue. Regarding the staged data in AQS, the records that did not load can either be corrected directly in AQS or corrected externally and a new data file be submitted through the process outlined in section 9.2.2. The standard procedure at UCD is to correct the data outside of AQS and resubmit the file for tracking and documentation purposes.

Information on how to correct data in AQS can be found in the AQS User Guide (https://www.epa.gov/sites/default/files/2018-07/documents/aqs_user_guide_2018_2.pdf, from page 47) which can be found on the EPA's AQS website at <https://www.epa.gov/aqs>. Select 'AQS Users Guide' under Documentation and select the AQS User Guide PDF towards the bottom of the webpage.

In brief, to correct data externally and submit another file, the user first needs to delete the data from AQS staging by following these steps:

- Navigate to the Correct option after logging into the IMPROVE screening group;
- Select 'Raw';
- Enter details of known records with issues
 - For example several records for a particular Parameter Code may have failed to load. In this case, enter the relevant Parameter Code in the appropriate field.
- Click on the 'Execute Query' button.

Figure 10. AQS screen after selecting **Correct** -> **Raw** from the options, entering information on failed records (e.g., Parameter Code), and clicking on the 'Execute Query' button.

State Code	County Code	Site Id	Parameter Code	Action	Session Date	Duration	Reported Method	Unit	Code	Coil Date	Coil Time	Rpt Samp Value	Null Data Code	C
78	020	0001	86101	S	U	20220201	7	105	805	20200104	00.00	13.07851		3
78	020	0001	86101	S	U	20220201	7	105	805	20200107	00.00	5.67203		3
78	020	0001	86101	S	U	20220201	7	105	805	20200110	00.00	17.85896		3
78	020	0001	86101	S	U	20220201	7	105	805	20200113	00.00	16.06202		3
78	020	0001	86101	S	U	20220201	7	105	805	20200131	00.00	3.60980		3
78	020	0001	86101	S	U	20220201	7	105	805	20200119	00.00	12.42042		3
78	020	0001	86101	S	U	20220201	7	105	805	20200122	00.00	8.53120		3
78	020	0001	86101	S	U	20220201	7	105	805	20200125	00.00	6.10694		3
78	020	0001	86101	S	U	20220201	7	105	805	20200128	00.00	2.57481		3
78	020	0001	86101	S	U	20220201	7	105	805	20200116	00.00	10.42865		3

Column Name	Error Code	Error Message
RELATIONAL	20718	Production record doesn't exist to update or delete

A FIPS code that identifies one of the 50 states or other countries. - LOV
Record: 1782/1782 | List of Values

The data returned are from the AQS staging table and are those that have not yet been loaded into AQS (e.g., Figure 10 shows records that failed to load for the Coarse Mass parameter (code 86101)). The user should click on the down arrow and review the

information at the very bottom of the window to confirm the number of pulled records matches what was expected. The user should verify these data are to be corrected and click on 'Delete All Selected'. This button will delete all records that were returned from the query; the records do not all have to be selected.

If there are multiple types of issues, the user should repeat the steps outlined above until all data that failed to load are deleted.

To correct the data, the user may need to fix the data directly using tools and techniques described in UCD 351#C Data Validation, reprocess the data as described in UCD 351#B Data Processing, generate another AQS file as described in section 9.2, and submit to AQS.

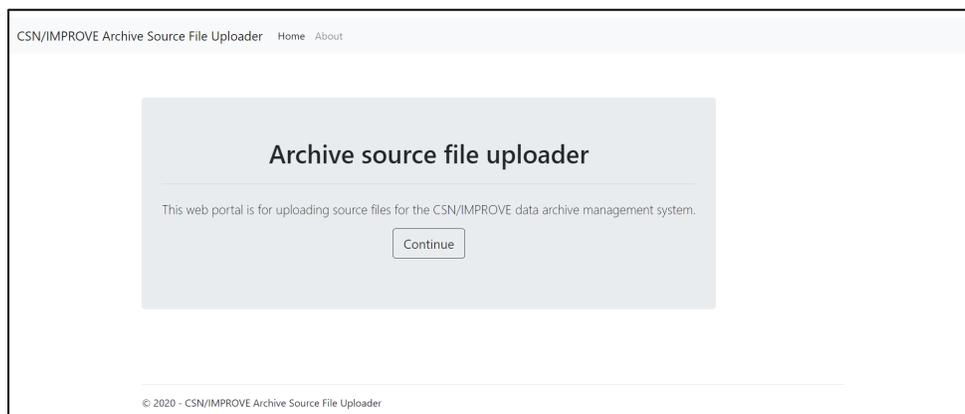
Alternatively, if the data can be easily corrected in bulk, such as if the correction is to change the Action Indicator to "I" from "U", the user can utilize the records in the failed records link in the AQSDEV email. The user should click on the link to the text file, which opens a web browser tab. The user should copy the data – except the last line of the page that notes the number of records that were posted to the file – and paste in a text file such as Notepad. The user can correct the data using 'find and replace' and save as a new AQS file using the same naming convention as is used for regular files, with additional information to note the contents of this particular file. The file can then be submitted to AQS as a typical data file.

9.3 Submission to UCD CIA

The CSN/IMPROVE Archive (CIA) is a database of the complete record of CSN and IMPROVE data coupled with a web-based visualization and analysis tool.

1. Open a web browser and navigate to the UCD CIA submission website, <https://cia-uploadportal.azurewebsites.net/> (Figure 11).

Figure 11. UCD CIA submission website home page.



2. Click the **Continue** button in the center of the page.

3. Specify the network of choice that you will be delivering the data for, which in this case is **IMPROVE**. See Figure 12 for an example.

Figure 12. UCD CIA data submission details 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

SUBMIT | Cancel

© 2020 - CSN/IMPROVE Archive Source File Uploader

4. Click **Browse** and select the file generated/submitted successfully to AQS.
5. Once the file is selected, click **Submit**; the next page will indicate if the submission was successful.

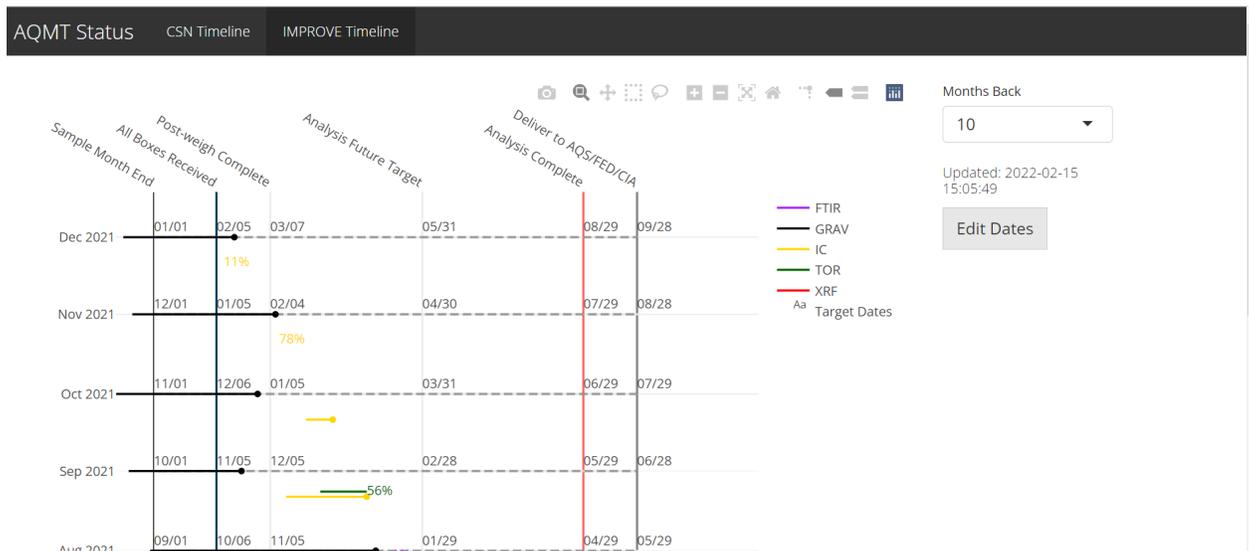
Note that the CIA upload portal will only accept files with a single action indicator. Most initial delivery files have only one action indicator (I - ingest). But the redelivery files can have more than one action indicator. If the upload fails, the file needs to be separated based on the action indicator and uploaded separately.

9.4 Add Data Delivery Dates Delivery Dates to AQMT Status Page

After delivering data to CIRA, the user should perform the following steps:

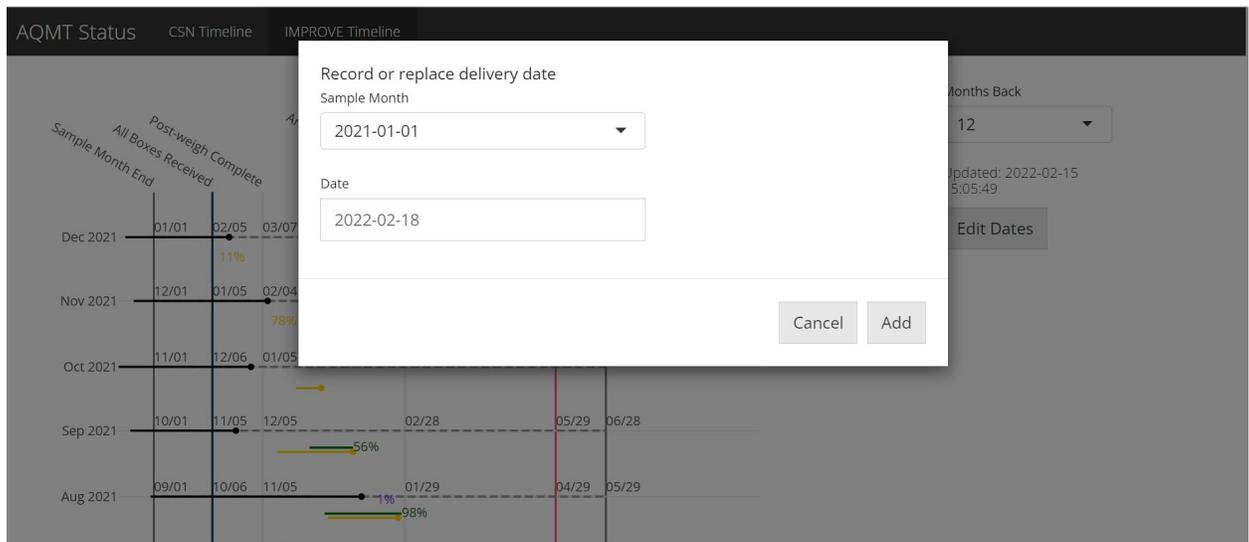
1. Add the IMPROVE delivery dates to the IMPROVE Timeline on the AQMT Status page (<https://shiny.aqrc.ucdavis.edu/aqmtstatus/>).
2. After accessing the web page, go to the 'IMPROVE Timeline' tab (Figure 13) and click on the Edit Dates button.

Figure 13. 'IMPROVE' Timeline tab on AQMT status page.



3. Select the sample month and add the delivery date and click on the Add button (Figure 14).
4. After adding the file delivery date, confirm that a tick mark is added to the timeline for the month.

Figure 14. File submission date update window.



9.5 Submission of Preliminary Data to CIRA

An unvalidated preliminary data set will be delivered to CIRA every month. Preliminary data differ from validated data in several ways:

- Validation has not been performed or is incomplete.
- Blank corrections are estimated from past data.

- MDL and uncertainty values are not calculated or reported.
- Statuses may include those used for internal validation purposes only and are not delivered in final data.

9.5.1 File Generation

The `improve_export_preliminary` function in the `crocker` package is used to create the preliminary data export file. To generate the preliminary delivery file, run the following command in the R environment:

```
[preliminary_data] <- crocker::improve_export_preliminary(start_year=[YYYY],  
start_month=[MM], end_year=[YYYY], end_month=[MM], server = 'production',  
write_file = FALSE)
```

The argument `start_year` is the numeric Year of the first sample month to process and `start_month` is the numeric month of the sample month to process. To determine the starting Year and Month to process, visit the IMPROVE timeline tab on the AQMT status page as mentioned in section 9.4 (<https://shiny.aqrc.ucdavis.edu/aqmtstatus/>) and look for the last month of data that was delivered to CIRA, AQS, and CIA. The start year and month of preliminary data to be processed and delivered will then be the month after that. The argument `end_year` is the numeric Year of the last sample month to process. The `end_month` is the numeric Month of the last sample month to process. To determine these end dates, the IMPROVE timeline tab can be reviewed again. The most recent month with completed 'GRAV' analysis can be used as the end date for processing and delivering the preliminary data, however gravimetric analysis does not have to be completed to run preliminary processing.

By default, the `write_file` argument is FALSE. If the write file argument is not specified or set to FALSE, but the function is assigned to a variable then the command generates a data frame containing all data for the year and month range. The user can review the data in their R environment. If the `write_file` argument is set to TRUE, a delivery file in CSV format will be saved to `U:/IMPROVE/FED Export/Preliminary/`, named 'IMPROVE_Preliminary_Unvalidated_YYYY-MM-DD' (e.g., `IMPROVE_Preliminary_Unvalidated_2022-05-05`), where YYYY-MM-DD is the date the file was created.

At the time of processing, many analyses will be incomplete. When running the code, warning messages for missing analyses along with other usual validation messages will be printed to the R console. These warnings will not interrupt the processing. The user is not to act on these messages. However, if the processing stops because of an error, the user is to inform the Software Group about the issue.

Once the data frame is generated, review the 'SampleDate' column to confirm all expected months are included. If data for any months are missing, especially the more recent months being queried, it could be because processed flows do not exist for those

samples. The status of processed flows can be checked with the Data & Reporting Group Manager or the Sample Handling Laboratory to determine if the flow can be processed for the missing months and subsequently proceed with re-processing and delivering the data. If the flow cannot be processed for more recent months, only report data through to the last month of data that was processed for flow. If the flow data can be processed, use the SQL query or the `improve_process_flow` function in R as described in Section 9 of IMPROVE TI 351B_Data Processing to derive the daily average flow rate and elapsed time (ET).

Once the data have been finalized, the data frame can be saved into a CSV file by running the following command in the R environment, where *YYYY-MM-DD* in the file name is the date the file was created:

```
write.csv(preliminary_data, 'U:/IMPROVE/FED  
Export/Preliminary/IMPROVE_Preliminary_Invalidated_YYYY-MM-DD.csv')
```

A typical command for standard monthly data delivery can be run as follows:

```
preliminary_data <- crocker::improve_export_preliminary(2021, 05, 2021, 12, server  
= "production", write_file = TRUE)
```

Once a file has been generated, the user is to compress the CSV file into a zip folder and email to the CIRA correspondent(s) as an attachment.

10. DATA AND RECORDS MANAGEMENT

The IMPROVE data are stored in Microsoft SQL Server Databases at UC Davis. The production database is run on a dedicated Windows Server with a RAID array for storage and with offsite backups. Our development and test database environments are virtual machines. To test back up recovery, our development and testing environments are regularly restored from the production backups.

Data management is handled through custom software that interfaces with the UCD IMPROVE database. The primary applications for data ingest and management were developed on the .NET platform. Data processing and calculations were developed as R software packages. In addition, to support data validation and operational monitoring, several interactive visualizations have been developed using the R Shiny platform.

10.1 Disaster Recovery Plan

The scope of recovery activities will depend on the nature of the disaster. Please refer to *UCD IMPROVE SOP #351: Data Processing and Validation* for specific information.

10.1.1 Facility Recovery

Private security services patrol the laboratory building on a regular basis (including nights, weekends, and holidays). In addition, campus facilities and maintenance staff are on call at all times.

Databases, file servers, and web server virtual and dedicated machines operate primarily out of the Metro IT data center in Hoagland Hall on the UCD campus. Metro IT has a highly-available, disaster recoverable virtualization environment. Weekly backups of the virtual hard drives are taken offsite and stored in the Campus Data Center. In the event of a disaster in Hoagland, critical machines will be mounted at the Campus Data Center. The Drew Avenue laboratory is directly connected to the main campus internet. In the event that connection is disrupted (such as through a construction accident), connections will be switched to a local backup server until service can be restored.

10.1.2 Hardware Recovery Plan

The campus network of IT Administrator staff allows for rapid response to server failure and recovery issues.

10.1.3 Software and Data Recovery Plan

10.1.3.1 UCD Laboratories

Raw and processed analysis data produced with the UCD laboratories are saved and available for use at any time on the computers associated with each instrument, including the PANalytical Epsilon 5 EDXRF, MTL Automated Weighing System (gravimetric mass), Hybrid Integrating Plate and Sphere (HIPS).

Operational flow rate information from samplers in the field is automatically transferred nightly to a file processing server. As a backup, the flow data are stored on SD cards and delivered to the sample handling lab along with the exposed filters.

Data from all analyses, along with the flows, are scheduled to automatically transfer to a central Microsoft SQL Server database located at a data center on the UCD campus. Differential backups are performed daily, and full backups are performed weekly.

10.1.4 Data Security

UCD access policies: Access to databases and computers associated with this project is limited to authorized project personnel by use of access control lists for files, programs, and database access. Access to laboratory and office space is controlled by keycards.

Password policies: Unique passwords are issued to each employee by the UCD campus system administrator. Password integrity is monitored by the UCD campus system administrator.

Termination policies: System access is revoked for terminated personnel. The IT Administrator disables domain accounts and passwords upon termination of employment.

Virus protection: Microsoft Endpoint Protection is used for virus scanning and protection. All staff are required to complete annual cyber security awareness training.

11. QUALITY ASSURANCE AND QUALITY CONTROL

11.1 Code Development

Software for data management, processing, and validation is developed in-house by professional software engineers. Source code is managed through a code repository. Development of code changes and new applications is conducted on a development environment that parallels the production environment. Prior to deployment in production, all code changes undergo testing within a separate test environment. The testing, which is conducted by developers, managers, and users, is targeted both at the identification of software bugs and the confirmation of valid data equivalent to the production system.

11.2 Bug Reporting

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

11.3 Data Validation

Data integrity is enforced within the UCD IMPROVE database via unique primary keys and non-nullable records. Data completeness and data quality are thoroughly checked through the data validation process, as described elsewhere in this SOP.

12. REFERENCES

Not applicable.