UCD IMPROVE Technical Information #351F

Data Preparation and Reporting

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

January 7, 2025
Version 1.3

Prepared By: Signed by: Indu Thekkemeppilly Sivakumar 19F6B63B1B17443... Date: 1/10/2025

Reviewed By: DocuSigned by: Alex Murrain 2B7AC07297404D6... Date: 1/10/2025

Approved By: DocuSigned by: Marcus Langston 0A10CFCF79B0452... Date: 1/10/2025

UCDAVIS
AIR QUALITY RESEARCH CENTER

DOCUMENT HISTORY

Revision	Release Date	Initials	Section/s Modified	Brief Description of Modifications
1.0	03/14/2022	SRS	All	New TI created.
1.1	11/10/2022	DEY, ITS	9.2	Updated procedure for compiling report.
1.2	05/09/2023	ITS	9.4	Added a new section on preparing the site metadata report
1.3	01/07/2025	ITS, AMM	3, 9.3	Added new section on closing sites in UCD database.

TABLE OF CONTENTS

LIST OF FIGURES

1. PURPOSE AND APPLICABILITY

The purpose of this technical information (TI) is to provide guidance on metadata creation, data creation and modification, and compilation of reports. This document is intended to guide users on creating box information in the University of California, Davis (UCD), Interagency Monitoring of Protected Visual Environments (IMPROVE) database and changing box metadata. Additionally, users are instructed on how to add new site information to both the UCD IMPROVE database and the EPA's Air Quality System (AQS) database. The document also provides guidance on updating these databases when a site is closed. The procedure for compiling reports on the current field status of the IMPROVE data is also included in this document.

2. SUMMARY OF THE METHOD

Requests are received from the UCD sample handling laboratory to assist with data preparation including creating boxes in the UCD IMPROVE database and modifying existing box information. Boxes are created and modified in the UCD IMPROVE database using the UCD IMPROVE Data Management website.

On a quarterly basis, the current status of the IMPROVE samples is reported to IMPROVE related personnel. The report is compiled using the IMPROVE Status Views website and is in an Excel spreadsheet format.

New IMPROVE sampling sites may start within the network; the site metadata is to be added to the UCD IMPROVE database directly using SQL queries, to AQS directly in the AQS application, if the data from the site is to be delivered to AQS, and the National Park Service (NPS) is to be informed of the new site metadata.

3. DEFINITIONS

- **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.

- **Federal Land Manager Environmental Database (FED):** a database of environmental data managed by Cooperative Institute for Research in the Atmosphere (CIRA).
- **Interagency Monitoring of Protected Visual Environments (IMPROVE):** Federal $PM_{2.5}$ and PM_{10} 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.
- **NPS:** National Park Service.
- **PM:** Particulate Matter. $PM_{2.5}$ is particulate matter with diameters 2.5 micrometers (µm) and smaller. PM_{10} is particulate matter with diameters 10 µm or smaller.
- **SOP:** Standard Operating Procedure.
- **SQL:** database management system used by AQRC.
- **UCD:** University of CA—Davis.
- **Sample Handling laboratory:** The filter handling laboratory for IMPROVE at UCD AQMT.
- **Field group:** The group in charge of samplers and operation of IMPROVE network.
- **UCD IMPROVE Data Management Website**: A web-based application developed with .NET framework to interact with the IMPROVE database. Also referred to as the IMPROVE web app.

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

9.1 Data Preparation Tasks

9.1.1 Box Creation

Occasionally, the box sent to the site by the sample handling lab is lost either before it reaches the site or after sampling and before being received back at the sample handling lab at UCD. For boxes that are lost prior to sampling, a replacement box is created and sent to the site as soon as the sample handling lab is alerted to the lost box. If the Data & Reporting group are requested to assist in the creation of a new/replacement box, the following tool can be used: https://improve.aqrc.ucdavis.edu/Operations/BoxSchedules, which can be accessed by going to the IMPROVE web app, selecting the **Operations** tab and the sub-tab of **Schedule**. To create a box the following steps should be taken:

1. Go to the Box Schedules page and select the site via the drop-down menu next to **Sampler** for which the new box is needed. Click on the **Go** button to the right (Figure 1).
2. Scroll to the bottom of the page and click on the **Add New Box** (Figure 1) button on the left.
3. This will lead to a **Create Box** page. Enter the relevant date in the **InstallDate** option and check **this is a replacement box** option and click **Create**. By selecting the replacement box option, the lab station Id will be set to pre weigh.
4. Select **Add New Cartridge** and on the “Create Cartridge” page that is subsequently opened (Figure 2), various cartridge information can be added including Sampler Module ID (e.g., 1A, 2B, 3C, 4D), Install Date (this is the cartridge install date and can be found in the IMPROVE calendar), and Schedule Week (i.e., Week 1, Week 2, or Week 3). An example of install dates and schedule weeks is as follows: next upcoming box install date is on 04/20. The week 1 installment is 04/20, week 2 is 04/27, and week 3 is 05/04. The cartridge installment is always on a Tuesday regardless of it being a 2-3-2 or 3-2-2 box. Figure 2 shows an example of a Cartridge that is ready to be created. Once all relevant and required information is added, click **Create**.

Figure 1. Box creation page.

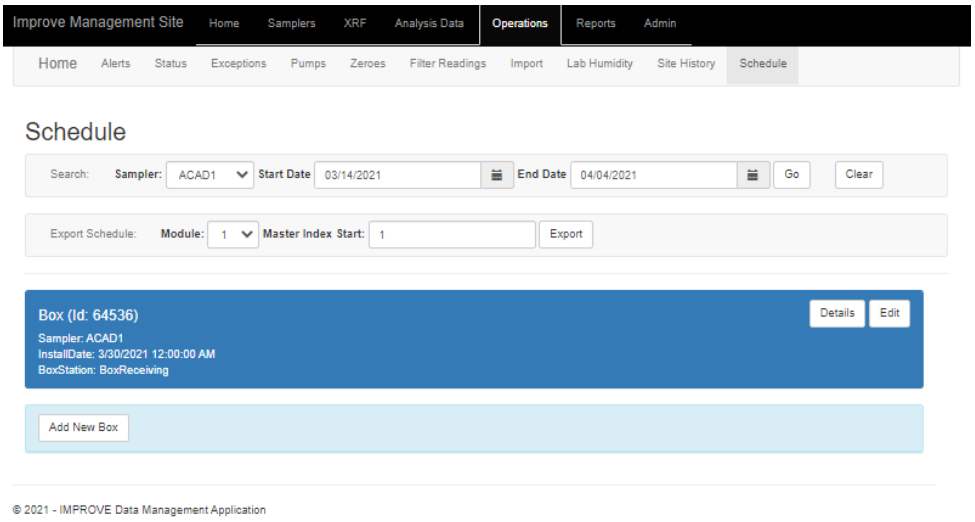
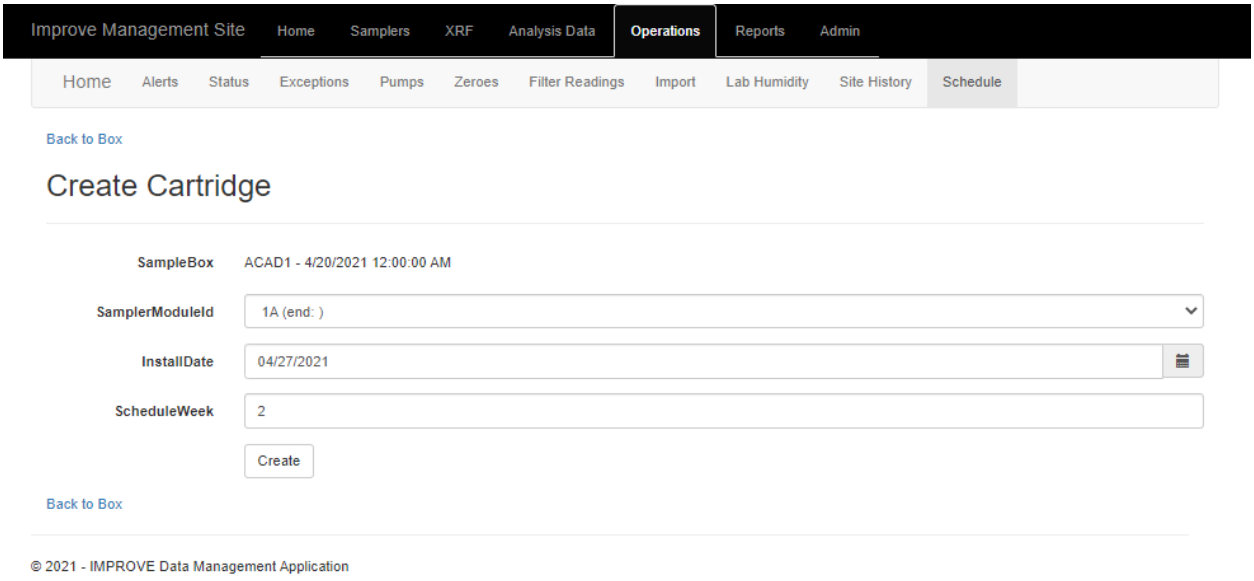


Figure 2. Cartridge creation page.

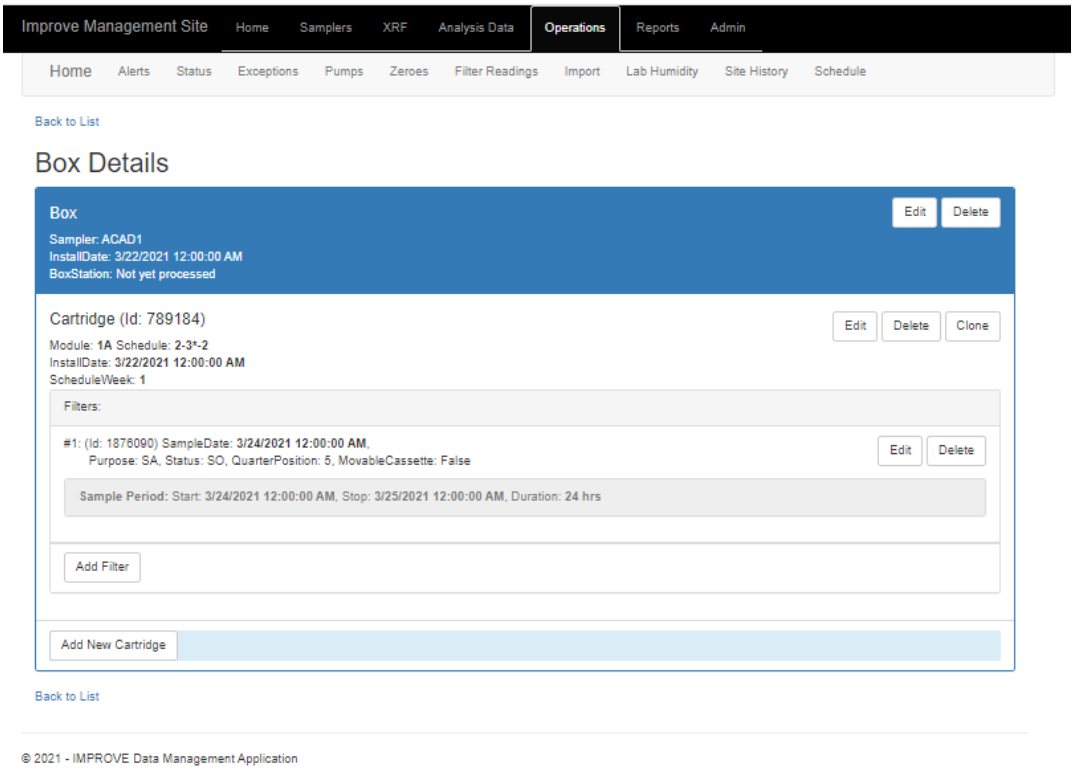


5. Once the cartridge is added, individual filters can then be added, one-by-one, using the Add Filter feature on the page loaded after creating the cartridge. After clicking on **Add Filter**, the "Create Filter" page is loaded. Every filter added requires the following information to be added: Cartridge Position (1, 2, or 3), Sample Date, Quarter Position (which can be found from the details of the lost box), Lot ID, and an indicator for whether it is a Moveable Cassette ("o-ring"). For the 2-3-2 boxes, the Moveable Cassette will always be the third position of the second week. For the 3-2-2 boxes the Moveable Cassette will always be the third position of the first week. All four modules will have the Moveable Cassette on the same date (position); as the

cartridges are cloned, this information will automatically be transferred. Once all information for a single filter is filled, click the **Create** button at the bottom of the page. A summary of the added filter is then displayed (Figure 3).

6. To add more filters, click **Add Filter** at the bottom of the summary information. Repeat the instruction from step 5 to add the filter information. The number of filters to be added depends on whether it is a three- or two-position week. The box schedule can be found under the Cartridge ID details (near "Module" details; Figure 3). For a three-position week, three filters will need to be added and for a two-position week, two filters will need to be added.
7. Once a cartridge is created and all relevant filters have been added, the cartridge can be cloned to create cartridges for other modules by clicking the **Clone** button to the right of the cartridge information on the Box Details page (Figure 3). The user should select the relevant Destination Module and Destination Lot from the drop-down menus and click **Create**.
8. Repeat steps 4-7 to create the cartridges for the second and third week.
9. Once the entire box has been created, the box details are to be sent to the sample handling lab to add filter pre-weights and filter barcodes for the PTFE filters. Make sure the current lab station ID (as described in section 9.1.2) is set to 2 so the sample handling lab can assign the pre-weights accordingly. If the box was created in place of a lost box, please proceed to section 9.1.2 for further actions that need to be taken.
10. If an item needs editing or deleting at any point of the box creation, the edit/delete options on the right-hand side can be used accordingly (Figure 3). To delete a box, all cartridges must first be deleted. To delete a cartridge, all filters within the cartridge must first be deleted.

Figure 3. Box creation page; after addition of filter.



9.1.2 Changing Current Lab Station ID and Assigning UF

When a replacement box is created in the case when boxes are lost prior to sampling, there are several additional steps to be performed to correctly assign data and other information to both the new box and lost box.

- For the new box, the sample handling lab assigns filter pre-weights. The current lab station of the box needs to be PreWeigh (Station ID = 2) to enable the sample handling lab to assign the weights. The box creation tool has the option to set the current lab station Id to PreWeigh. The field can be updated using the following SQL query if it wasn't selected during box creation., where *NewBoxID* is the ID of the newly created box:

  UPDATE [Improve_2.1].[filter].[SampleBoxes]

  SET CurrentLabStationId = 2

  WHERE Id = NewBoxID

  Check the update was successful by performing a SQL select query e.g.:

  *SELECT **

  FROM [Improve_2.1].[filter].[SampleBoxes]

WHERE Id = NewBoxID

- For the lost box, the current lab station needs to be updated to Finished (Station ID = 9). To do this, run the following SQL update query in the UCD database, where *LostBoxId* is the ID of the lost box:

  UPDATE [Improve_2.1].[filter].[SampleBoxes]

  SET CurrentLabStationId = 9

  WHERE Id = LostBoxId

  To check the update was successful, perform a SQL select query e.g.:

  *SELECT **

  FROM [Improve_2.1].[filter].[SampleBoxes]

  WHERE Id = LostBoxID

- For the filters in the lost box, the filter purposes are to be updated to UF (Unused/Lost Filter (Filter Purpose ID = 16) and can be updated using the following SQL update query, where *LostBoxId* is the ID of the lost box:

  UPDATE f

  SET f.FilterPurposeId = 16

  FROM [Improve_2.1].[filter].[Filters] f

  LEFT JOIN [Improve_2.1].[filter].[SampleCartridges] sc ON sc.Id = f.SampleCartridgeId

  WHERE sc.SampleBoxId = LostBoxID

After updating the filter purpose, review and confirm the filter purpose Id for the whole box is correct by running the following query.

*SELECT **

FROM [Improve_2.1].[filter].[Filters] f

LEFT JOIN [Improve_2.1].[filter].[SampleCartridges] sc ON sc.Id = f.SampleCartridgeId

WHERE sc.SampleBoxId =BoxID

9.2 Quarterly Field Status Report

A field status report is generated quarterly to report on the current status of all samples collected across the network for the previous quarter. Site status is evaluated relative to the regional haze rule criteria. The following information outlines the steps to generate the report and the checks to perform before delivery.

1. First, process flow data using the SQL execution code or the Crocker flow processing function as described in section 9 of UCD TI #351B and process flows

for the relevant date range to be covered in the quarterly field status report by changing the Start Date and End Date fields. Processing flows at this point ensures the most up-to-date flow data and subsequent statuses are reported.

2. Create the report spreadsheet:
    - For the first quarter of a new year, save a copy of the template report under another name, with the format of *IMPROVE Status Report YYYY Q#*. The template report is located at U:\IMPROVE\Status Reports\Status_Report_Template.xlsx. In the Site status report tab, update the year in the Terminal samples by quarter and percent complete by quarter columns to the current year. For example, for 2022, it should be A22, B22, C22, and D22.
    - For the second, third, or fourth quarter, find the last report and save it under a name indicating the relevant quarter number. Previous reports are located at U:\IMPROVE\Documentation\Quarterly Field Status Report
    - In the report there are four tabs:
        - Site Status Report
        - Status Flag Table
        - Flag Definitions: available from the database,

          *SELECT **

          FROM [Improve_2.1].[filter].[Statuses]
        - Sampler Locations: Determine if any sites are new, re-started, or have stopped during the relevant quarter by reviewing the date information in the *[Improve_2.1].[sampler].[Samplers]* and *[Improve_2.1].[module].[Modules]* tables in the production database. The *sampler.Samplers* table gives the site installation date, while the *module.*Modules table lists the first sampling date.
3. Populate the report:
    - From the IMPROVE Status page (https://shiny.aqrc.ucdavis.edu/ImproveStatus/), access the Network Status and Network Timeline tabs.
    - The Network Status tab provides a count of the different statuses used per site, the total number of terminal statuses, which quarter they occur in, the percent complete by quarter, and the number of consecutive invalid statuses. The status page has options to choose “filter status only”, “filter and flow statuses”, and “final statuses (filter, flow, and analysis statuses)”. For this report, the user is to only consider filter and flow statuses.
        - Change the Year and Ending Quarter fields to align with the reporting period.
        - Select the **filter and flow statuses option** from the Status Types drop-down menu.

- Download the full table by clicking on the **Download** button. A .csv file is downloaded.
- Compare the columns in the Site Status Report tab to the content of the downloaded spreadsheet; add new columns to the Site Status Report tab as needed.
- Compare the sampler details in the Site Status Report tab to the content of the downloaded spreadsheet; add sampler details to or remove sampler details from the Site Status Report tab as needed.
- Confirm that included flags are allowed (for example, the RF flag is no longer used). Investigate cases where unallowed flags are applied; work with the Sample Handling Laboratory to resolve.
- Add the flag, definition, and result to the Flag Definitions tab of the report spreadsheet if not already listed.
- Copy/paste content from the downloaded spreadsheet to the Site Status Report tab.
- Color the relevant fields:
    - Percent Complete by Quarter:

      < 75%, yellow
      < 50%, red

    - Consecutive Terminal Samples:

      > 7, yellow
      > 10, red

    - Annual Completeness:

      < 75%, red
      Note that the Annual Completeness column should only be colored for the fourth quarter (Q4) report. Report uncolored values for the first, second, and third quarter reports.

- Check formatting for consistency, including font style and type, coloring, and shading.
- Consecutive Terminal Samples may require merged cells to report a single number per row. Select all the cells to be merged; navigate to the MS Office Home tab and select Merge Across (drop down menu by Merge and Center); click okay for each row for the cells highlighted.
- Update the date and quarter details at the top of the Site Status Report tab.
- The Network Timeline documents the most severe filter status for each site and date. For this report, check the box to “Include flow statuses”.; Change the Year to be relevant.
- Do a search for NF statuses. If any NF statuses are found, process the flows again using the SQL execution code or the Crocker flow processing function. If the NF statuses are for the most recent

quarterly period, run the code in SQL or R, changing the Start Date and End Date fields; accordingly, if successful, a date/time of completion will show in the window:

If the NF statuses are for a small set of filters/sites/dates, confirm why this is the case and edit the SQL execution code or input the appropriate arguments in the Crocker flow processing code to run on the specified filter, date range, and/or site.

- Do a search for no statuses. For sites with no statuses, determine if it is a new site or a site that's closed or if there is a reason, such as paused shipments, or the site is temporarily offline.
  If the site is new or closed, there may be blank records prior to the start date/ end date; if so, leave as-is but make sure the site is not falsely reported as failing the Regional Haze Rule criteria. Also check and ensure that the annual completeness column in the Site status report tab is reported correctly.
  If shipments are paused, work with the Software & Analysis Group and Sample Handling Laboratory; records may need to be added and/or the OL status may need to be manually inserted.
- If no NF or blank statuses are found, the data can be downloaded to be included in the report. There is a checkbox option defaulted to include the collocated module data on the network timeline page. As the Network Status page does not include statuses of filters from collocated modules, the data downloaded from the Network Timeline page should also only use filter statuses from routine modules. To do this, unselect the **Include collocated modules** option and download the data by clicking on the **Download** button. The default name of the downloaded .csv filter is "IMPROVE_network_timeline.csv". As the PHOE5 site is a collocated site with all four modules, it is considered an independent site in the Network Status table. To include the data from this site in the quarterly report, re-select the **Include collocated modules** option and download the file by clicking on the **Download** button. The file name is "IMPROVE_network_timeline_cl.csv". Copy the line containing the PHOE5 information from this file and paste/insert it into the IMPROVE_network_timeline.csv file in the row under PHOE1.
- Compare the site list from the IMPROVE_network _timeline.csv with the sites listed in the Sampler Locations and Status Flag Table tabs in the Site Status Report tab. Update site and date information as needed after confirming with the Data & Reporting Group Supervisor; some sites are for special studies and are not included in this report.

- Copy/paste all content from the IMPROVE_network_timeline.csv to the Status Flag Table tab.
- Color the relevant fields:
  - QD, TO, CG, and LF flags, yellow
  - Null/terminal flags, red. SD status is terminal for regional haze rule.

4. Perform checks prior to delivery:
   - Verify that the color coding is correctly assigned.
   - Status Flag Table tab: Look for blocks of red (invalid) and SO flags. Investigate using JIRA and/or follow up with the Sample Handling Laboratory.
   - Status Flag Table tab: Spot check to ensure that the number of terminal flags corresponds to those reported in the Site Status Report tab.
   - Status Flag Table tab: Confirm that the sites listed are also shown in the Site Status Report tab and the Sampler Locations tab.
   - Flag Definitions tab: Confirm that the formatting and color coding is correct.
   - Sampler Locations tab: Confirm that new sites have been added.
5. Send it to the Data & Reporting Group Manager for review.
6. Once reviewed and approved, the Data & Reporting Group Supervisor will deliver to various personnel including IMPROVE site operators, NPS, and EPA staff via email. A summary of site losses is to be included in the body of the email.

9.3 Updating UCD Database and AQS Site Information

Whenever a new site starts within the network and the data will be delivered routinely to AQS, both the UCD database is updated with AQS related information and the AQS database is to be updated to add the new site and its associated monitors (where monitors are the AQS term for parameters). The site details should also be sent to NPS prior to data delivery when a new site starts sampling. In instances where an IMPROVE site ceases operations, the end dates within various UCD IMPROVE database tables are updated to mitigate potential data processing issues. Furthermore, corresponding updates to end dates are also applied to AQS sites and monitors.

9.3.1 Updating UCD Database

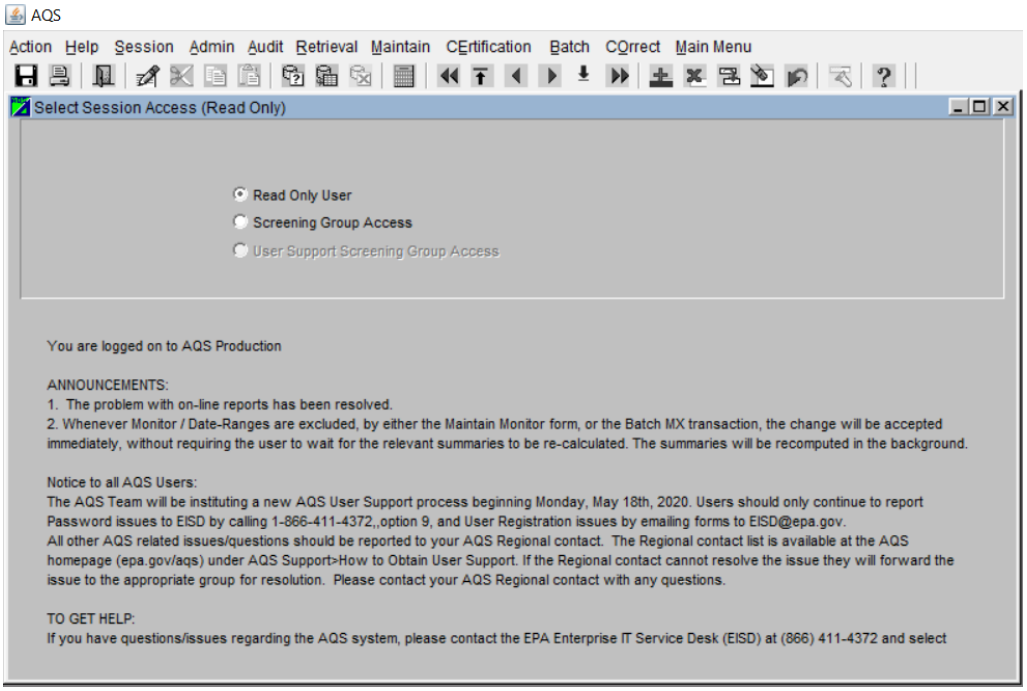
9.3.1.1 AQS Site ID

An AqsSiteId needs to be assigned to the new site in the sampler.Samplers table in the UCD database (Improve_2.1), which consists of State Code, County Code, and Site Id. Further information can be found in sections 3.2.3-3.2.5 of the AQS Data Coding Manual (https://www.epa.gov/sites/production/files/2015-09/documents/aqs_data_coding_manual_0.pdf). If the state and county where the site is located are known, then the associated codes can be found by searching FIPS online

(Federal Information Processing Standards/Series, e.g. https://geonames.usgs.gov/apex/f?p=138:1:::NO:1:P1_SHOW_FIPS55,P1_SHOW_ADV,P1_SHOW_ANTAR:Y,,) or the State and Counties codes on the AQS webpage (https://aqs.epa.gov/aqsweb/documents/codetables/states_and_counties.html). The user should query the AQS database to determine if a site already exists at the same location. To do this, the user should follow these steps:

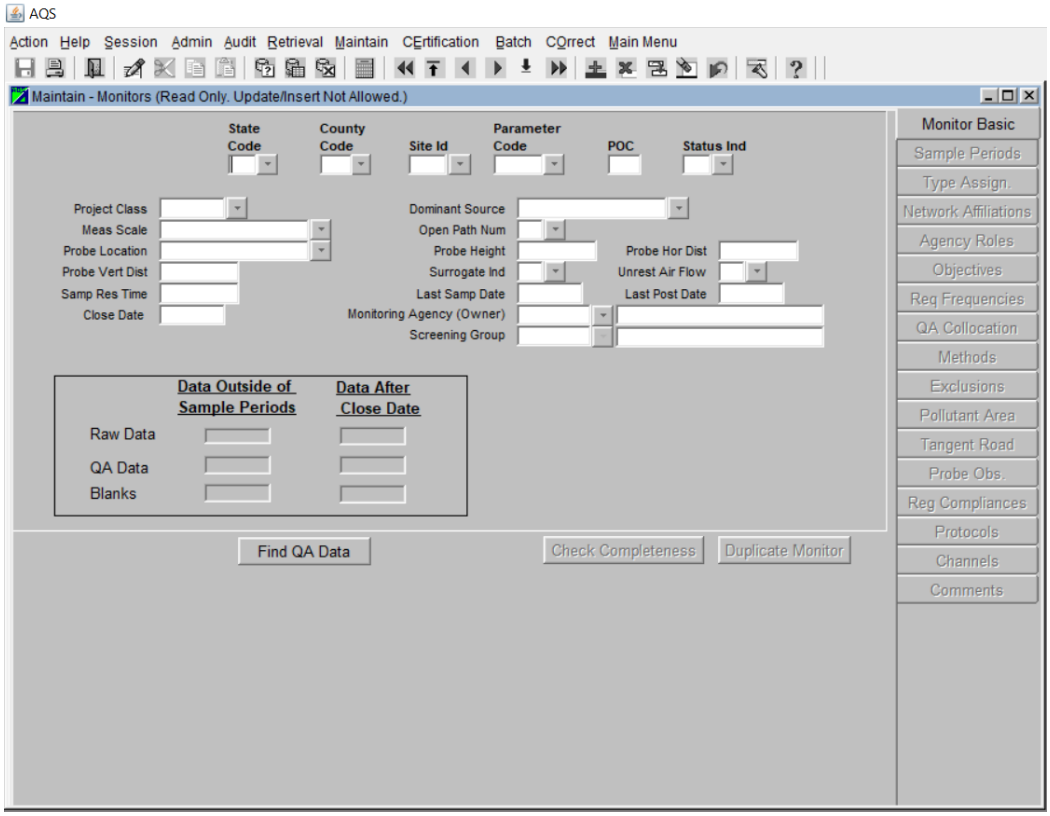
- Log into the AQS application
- Select the Read Only User option (Figure 4)
- Go to Maintain
- Select Monitor (Figure 5)
- Type in the State Code and County Code
- Click on the 'Execute Query' button.

Figure 4. AQS screen after logging into the application.



Please select the Read Only User option.

Figure 5. AQS screen after selecting Monitor from the Maintain option.



The site does not already exist in AQS if no results are returned. If any results are returned, the user should review the details to confirm if the parameter code and/or the POC is one that is reported as part of the IMPROVE network. By convention, if the site does not already exist in AQS, the Site Id assigned is '9000'. The user should repeat the steps above to include the Site Id of '9000' and executing the query again to confirm the site and/or monitors does not already exist.

Another way to check what site IDs are already in use and thus what are 'available' is to review the 'Site Listing' spreadsheet via https://aqs.epa.gov/aqsweb/airdata/download_files.html#Meta. Note that this is updated only every 6 months so it may not be the most up-to-date. By convention CDPHE uses the lower site numbers and NPS has used the 9000 series. Reach out to NPS, CIRA and air-resources to confirm what site Id's are already in use.

If 9000 is already in use, get the next available count, make sure it's not already in use, and confirm with various agencies (e.g. CDPHE) that we are not in conflict with anyone.

Once the State Code, County Code, and Site Id are known, the UCD database can be updated as follows:

- Query the database using the following SQL query to find the relevant site record, where 'XXXX#' represents the four-character site name plus the number, typically 1.

  *SELECT **

  FROM [Improve_2.1].[sampler].[Samplers]

  WHERE Name = 'XXXX#'

- Update this site record with the newly generated AQS site ID, a nine-digit ID comprising the State Code, County Code, and Site ID, with no separation.

  UPDATE [Improve_2.1].[sampler].[Samplers]

  SET AqsSiteId = '#########'

  WHERE Name = 'XXXX#'

9.3.1.2 AQS Parameters and POCs

In addition to updating the Samplers table in the database with the AQS site ID, POCs (Parameter Occurrence Code) need to be added to the analysis.AqsPOCs table. In AQS, POCs are assigned per parameter. If there was no existing site in AQS, POC = 1 for all parameters, except for the coarse mass parameter (PM_{10}-$PM_{2.5}$), which is assigned POC = 5 by convention. If there are existing collocated sites in AQS, the next smallest different number is to be used, e.g. POC = 2.

To add the parameters and POCs to the database, specifically the *analysis.AqsPOCs* table, a SQL insert query can be written using the starting format below where each set of values is for a different parameter:

INSERT INTO [Improve_2.1].[analysis].[AqsPOCs] (SamplerName, ObjectiveCode, Parameter, POC)

Values (SamplerNameX, ObjectiveCodeX, ParameterX, POCX), (SamplerNameX, ObjectiveCodeX, ParameterY, POCX), …

Alternatively, to add to the database in bulk, an R script can be written and used, ensuring that the outputs from each step of the script are reviewed along the way.

9.3.2 Adding Site and Monitors to AQS

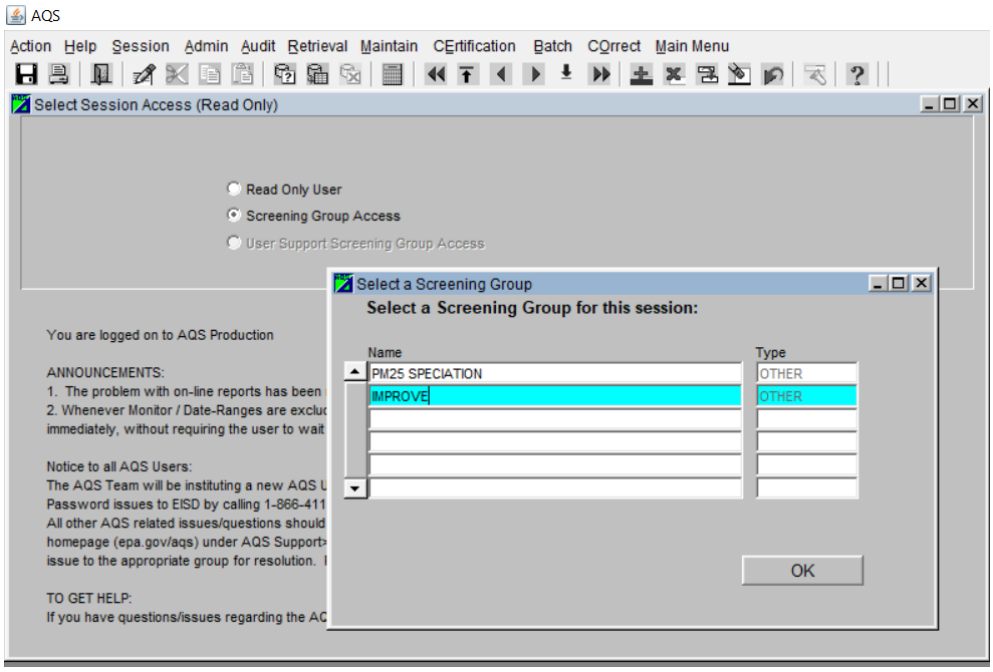
9.3.2.1 Adding a New Site to AQS

To add a new site to AQS the user should follow the steps below:

- Log into the AQS application.
- Select the **IMPROVE Screening Group Access** option (Figure 6).
- Go to **Maintain**.

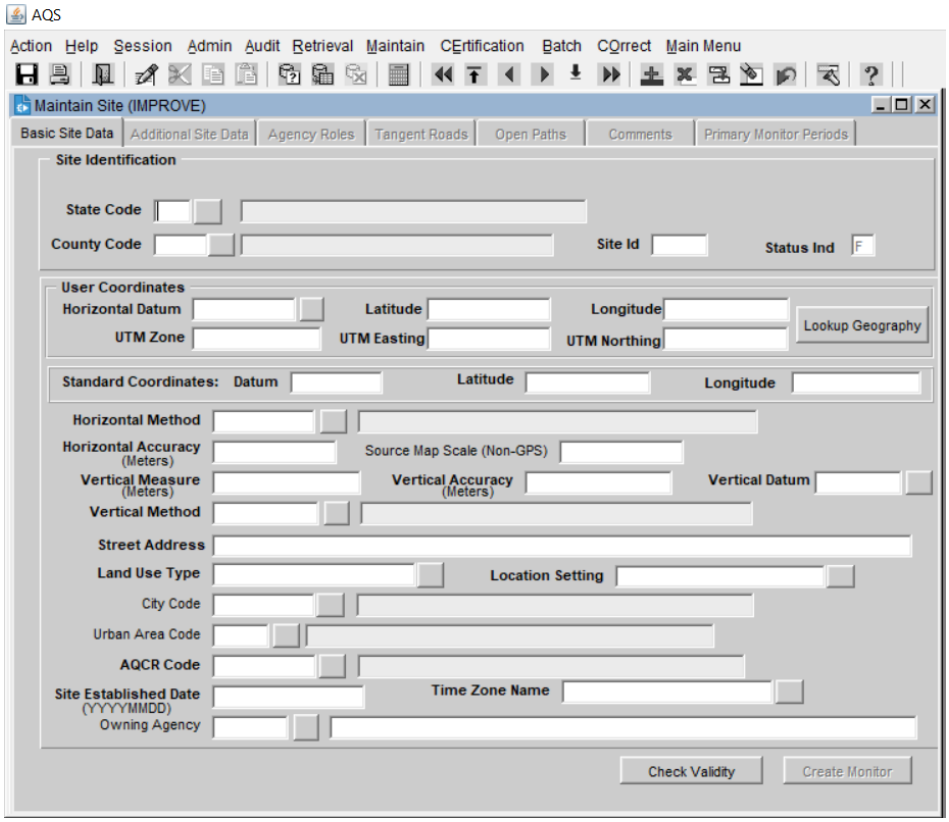
- Select **Site**.
- Click **Cancel Query** (Figure 7).
  - This allows the user to click on the **Check Validity** button at the bottom of the window once various details have been entered.

Figure 6. AQS screen after logging into the application.



Please select the Screening Group Access option and then IMPROVE. Next, select **Site** from the Maintain option and click **Cancel Query** (Figure 7).

Figure 7. AQS screen Basic Site Data to input.

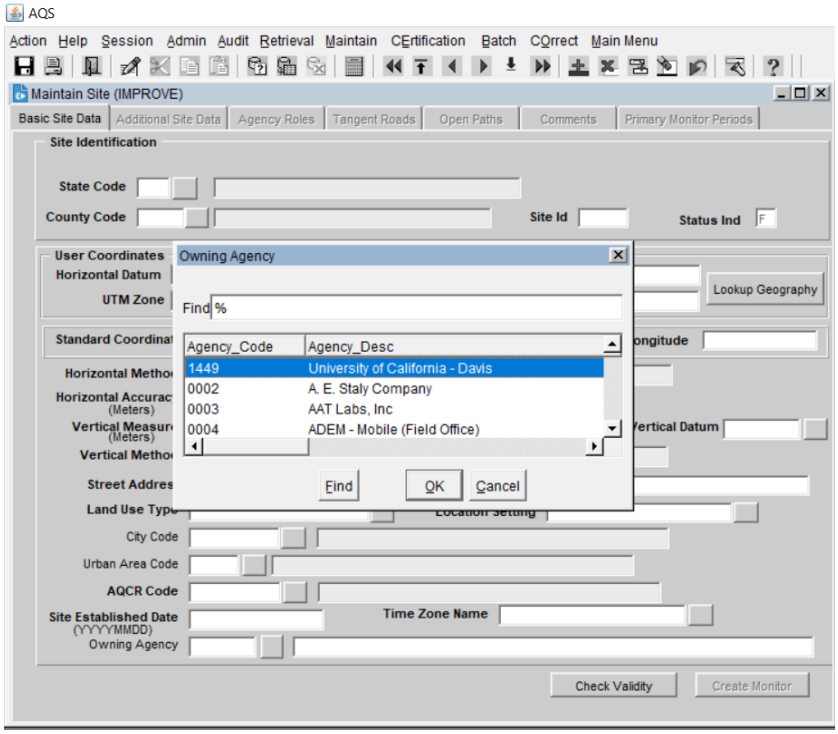


The fields to be completed are detailed below:

- Site Identification
  - The State Code, County Code and Site Id should all be known from the previous section. Enter the codes; the associated names will fill in automatically.
- User Coordinates
  - Horizontal Datum: WGS84 (by convention).
  - Latitude and Longitude: find in the IMPROVE Data page, under Sites, select the relevant site. Note that W longitude should be entered as negative. i.e., 105 W is entered as -105.
  - Horizonal Method: 103(GPS) (by convention).
  - Horizontal Accuracy: 5 meters (by convention).
  - Source Map Scale (Non-GPS): 10000 (by convention)
  - Vertical Measure: Site elevation; find in the IMPROVE Data page, under Sites, select the relevant site.
  - Vertical Accuracy: 5 (by convention).

- Vertical Datum: NAVD88 (by convention).
- Vertical Method: 001 (by convention).
- Street Address: If not already known, the site operator may have to be contacted to obtain this information.
- Land Use Type: If not already known, the site operator may have to be contacted to obtain this information. Options for this field are: Residential, Commercial, Industrial, Agricultural, Forest, Desert, Mobile, Blighted Areas, Military Reservation.
- Location Setting: If not already known, the site operator may have to be contacted to obtain this information. Options for this field are: Urban and center city, Suburban, Rural.
- AQCR Code: Use the drop-down menu, select the code listed; there should only be one.
- Site Established Date: Find in the *sampler*.*Samplers* table in the UCD database.
- Time Zone Name: Use the drop-down menu, select the option listed.
- Owning Agency: If not already known, the site operator may have to be contacted to obtain this information. To search for the agency and obtain the relevant agency code, use the drop-down menu and in the ‘Find’ box, type in the details of the agency (Figure 8) and select the appropriate affiliation from the returned results.

Figure 8. AQS screen when searching for Owning Agency details.



Once all the fields have been entered, the user should click **Check Validity** at the bottom of the screen. A Row_Errors window may appear with Error Descriptions left blank. If not blank, the user should return to the site's window and correct the errors. Upon completion, the details are saved by clicking on the save button at the top left of the AQS window (under Action). Follow any additional prompts, e.g. click **Lookup Geography**, and save again.

For confirmation the site is saved, the user should navigate to a fresh Maintain Site window, enter the State Code, County Code, and Site ID and execute query. The full site details should be displayed and the Status Ind (status indicator) should have changed from F to P, meaning the site is now in production.

9.3.2.2 Adding Monitors to AQS

In addition to opening the parent site, the monitors (parameters) need to be opened. There are two methods for adding monitors: Batch processing (preferred because of high efficiency) and manually adding monitors, one at a time.

For batch processing, the user should follow these steps:

- Navigate to the template text file, monitor_template.txt, at U:\IMPROVE\AQS\AQS_Documentation.
- Save this as a new file, with a file name that indicates which site it is for.

- Update the State Code, County Code, Site ID, and POC, if necessary by performing a Find and Replace in Notepad (or other application).
- If this is a new site, then the only date in the file should be updated to be the site start date. If the site is being closed, then an end date needs to be added as well (this field is currently blank in the template).
- Navigate to the Exchange Network website and submit the file in the same way that raw data is delivered as described in section 9.3.2.

To add monitors manually, one at a time, the user should follow these steps:

- After adding the site to AQS as detailed in section 9.3.2.1, the user should ensure they are in the IMPROVE Screening Group Access session and navigate to the Maintain Monitors window (Figure 5; section 9.3.1.1).

The following details the information that must be added to each specified tab to open a parameter (monitor):

- Monitor Basic: enter State Code, County Code, Site Id, Parameter Code, and POC for the parameter that is being opened.
- Sample Period: enter the date used for Site Established Date when creating the site in the Begin Date field.
- Type Assign: enter **EPA** as the Monitor Type and Begin Date is the same date as the Site Established Date.
- Network Affiliations: enter **IMPROVE** for the Monitor Network Code and the Begin Date.
- Agency Roles: enter a row each for Agency roles of ANALYZING, COLLECTING, REPORTING and PQAO, list the Agency Code as “0745” (National Park Service), and the Begin Date. If the Site Established Date is before 2007-01-01, the Begin Date should be entered as “20070101”.
- Objectives: select **GENERAL/BACKGROUND** and enter “0000” (not in an urban area) for the UA Represented field.

The user should save the entry and confirm that the Status Ind in the Monitor Basic tab is “P”. To add more monitors for the same site, the user should click on the **Duplicate Monitor** option at the bottom of the Monitor Basic tab and enter the appropriate parameter code details. The user should follow these steps until all relevant monitors are opened.

9.3.3 Updating NPS

The NPS needs to have the site details for any new site that starts sampling in advance of data delivery. The NPS requires the following information be sent:

- Site Name
- State

- County
- AQS Code
- Latitude
- Longitude
- Elevation
- Start Date
- End Date (assumed to be blank at present)
- Sponsor/Agency
- Location Description
- Rural or Urban (or other demographic code)
- Land Use code, if any
- Photos of the site.

9.3.4 Updating UCD-CIA Database

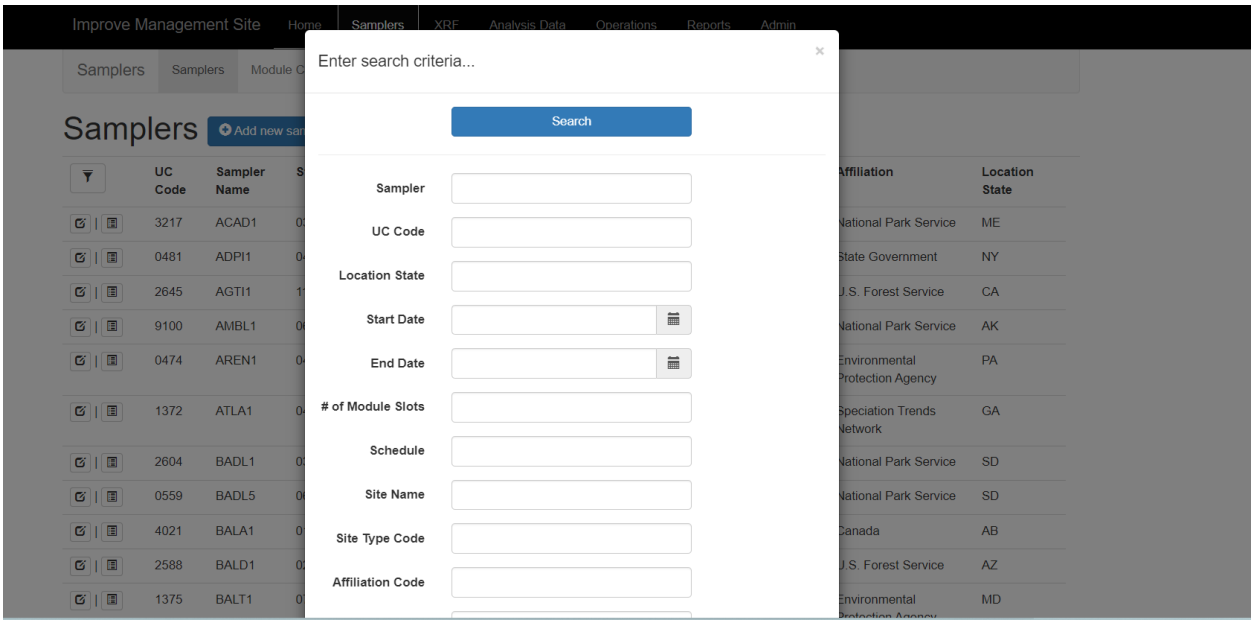
Whenever a new sampler is added, the sites table in the UCD-CIA database needs to be updated. The request should be directed to the software group. The software group uses the EPA's master AQS site list to add information to the database. If a new sampler is not added to the UCD-CIA database prior to data submission, the data will stage but will not migrate. Once the relevant sampler information is added to the database, the staged data will successfully be posted to the UCD-CIA database when the SQL query to migrate data is next run (typically every night).

9.3.5 Updating Site Closure in UCD Database

When an IMPROVE site becomes non-operational, the ending dates are updated in various UCD IMPROVE database tables to prevent data processing issues. Specifically, the *sampler.samplers, module.modules, module.ModuleAnalysisPaths, sampler.SamplerLocations, and sampler.TemperatureProbeVersions* tables have an end date field that needs updating. This can be accomplished through direct update statements in the database. The recommended approach is using the IMPROVE web app.

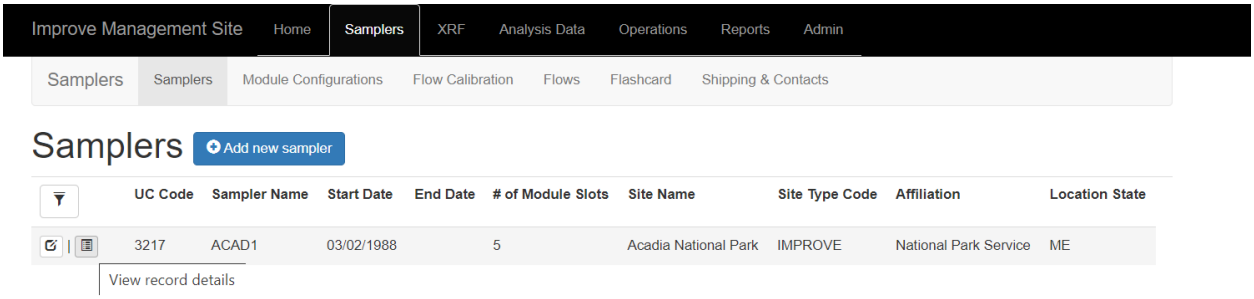
To update the end dates, please navigate to the samplers tab within the IMPROVE web application (https://improve.aqrc.ucdavis.edu/Samplers). Utilize the filter option shown in Figure 9 to select the sampler that is scheduled to become non-operational.

Figure 9. Sampler list page with filter option selected.



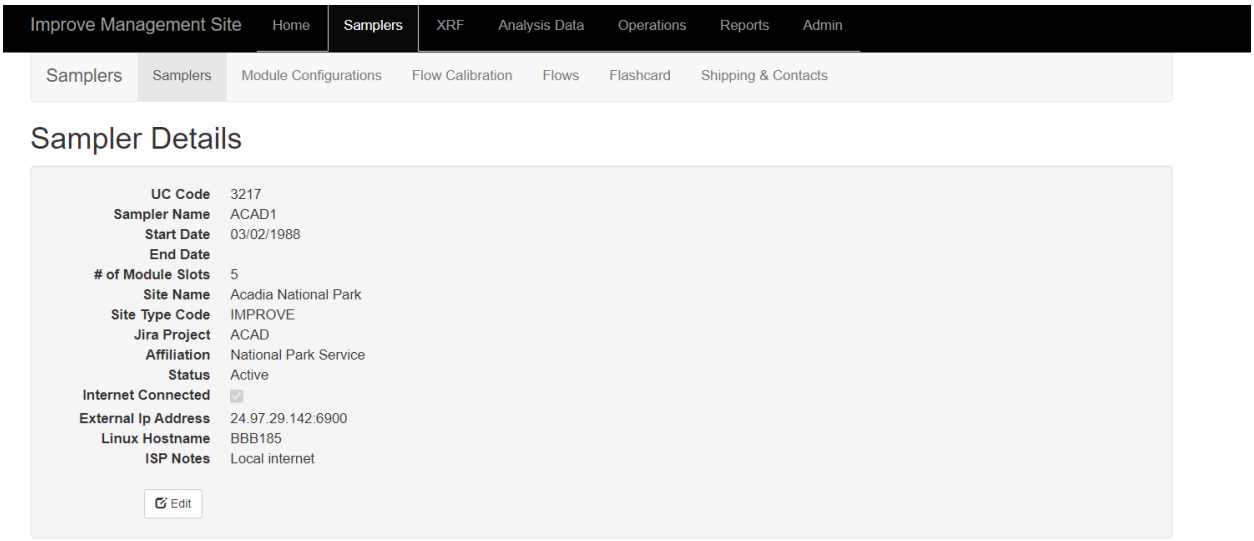
To update sampler information, input the sampler's name in the specified field and proceed by clicking the **Search** button. This action will generate a page containing the sampler's details. Subsequently, click on the **View record details** button as seen in Figure 10 to be directed to a page presenting all the requisite information.

Figure 10. Selected Sampler Record.



The page contains fields to update the end date in all five required tables. To update the end date for the sampler, access the “Sampler details” section and click the **Edit** button (see Figure 11). Within the Edit sampler page, locate the “End Date” field. The end date can be manually entered in MM-DD-YYYY format or by utilizing the calendar option. After adding the end date, confirm the changes by clicking the **Okay** button to finalize the process.

Figure 11. Sampler Details.



To update the end dates of the modules and module analysis paths, each module (four per site or five in cases of sites with collocated modules) must be individually modified. To update the module end dates, select the **View** button seen in Figure 12, which will redirect to a page containing the module details and module analysis path details. Within the respective sections, select the **Edit** button shown in Figure 13 and proceed to update the end date either by manual entry in MM-DD-YYYY format or through the calendar option. Upon completing the modifications in the Module details, confirm the changes by selecting **Okay**. For the Module Analysis Paths, click **Save** to confirm the changes.

Figure 12. Sampler Modules.

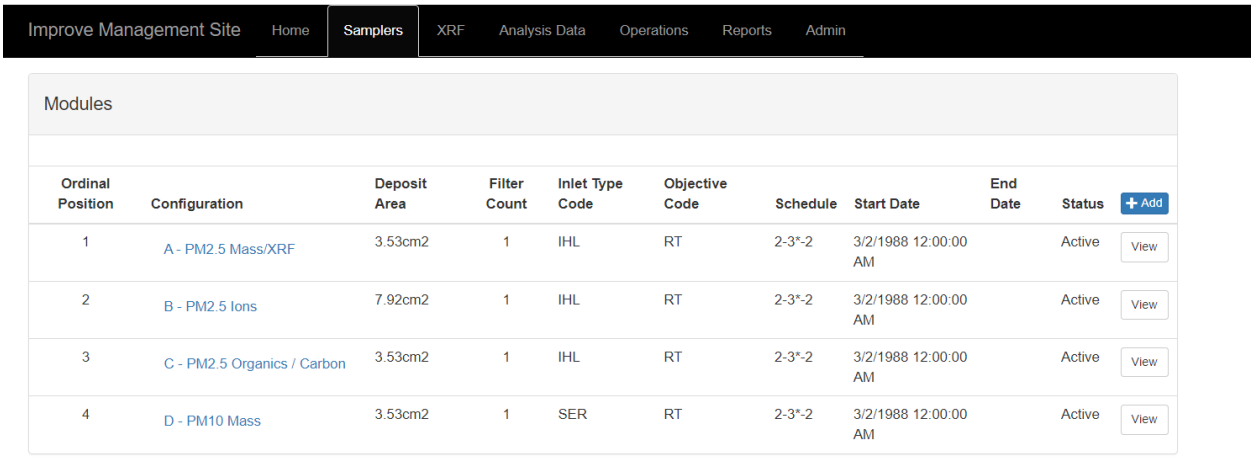
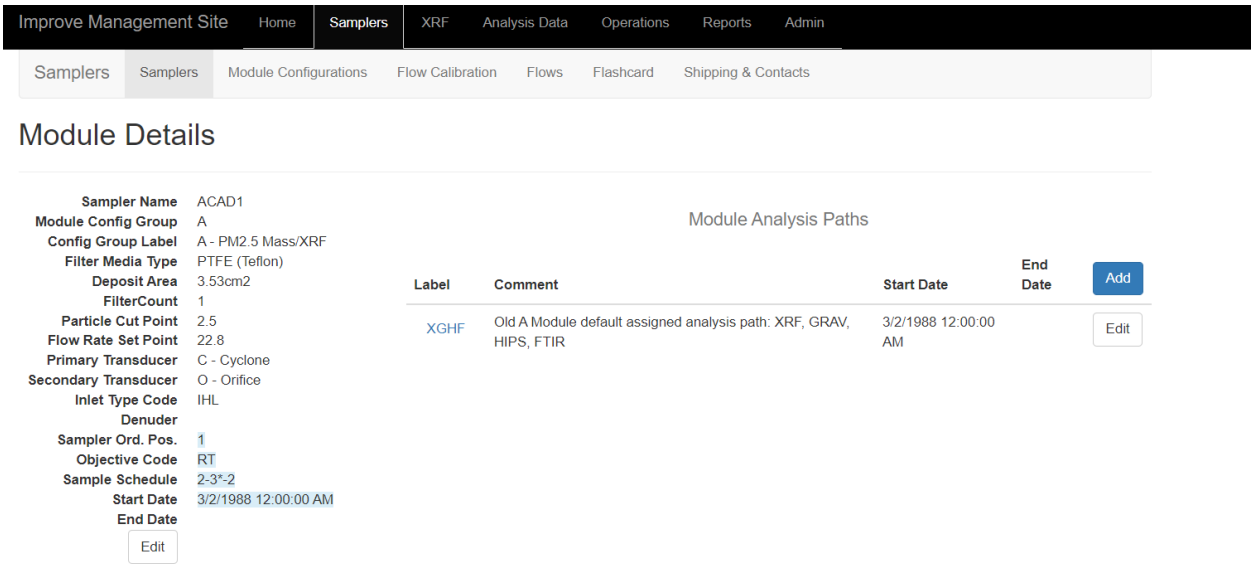
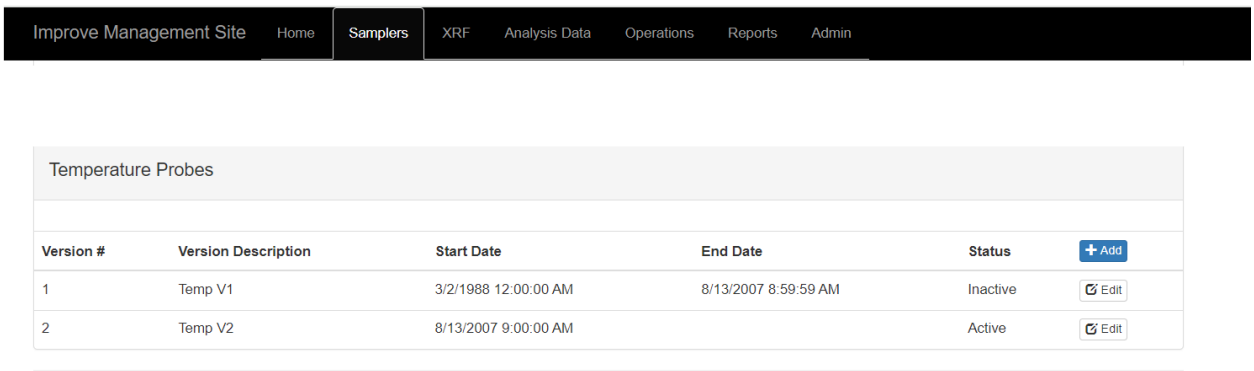


Figure 13. Module Details.



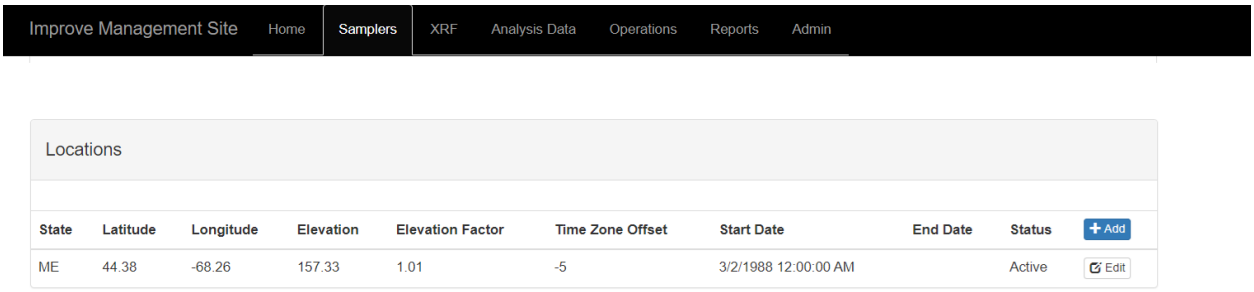
The temperature probe section contains a list of all active and inactive issues. To update the end dates for the temperature probes, navigate to the Temperature probes section shown in Figure 14 and click the **Edit** button for the **Active** temperature probe. Update the end date by manually entering it in MM-DD-YYYY format or by using the calendar option. After making the necessary changes, confirm the modifications by selecting **Okay**.

Figure 14. Temperature Probes.



To update the end dates for the sampler locations, go to the Locations section shown in Figure 15 and click the **Edit** button. Update the end date by entering it manually in MM-DD-YYYY format or by using the calendar option. After making the necessary changes, confirm the modifications by selecting **Okay**.

Figure 15. Sampler Locations.



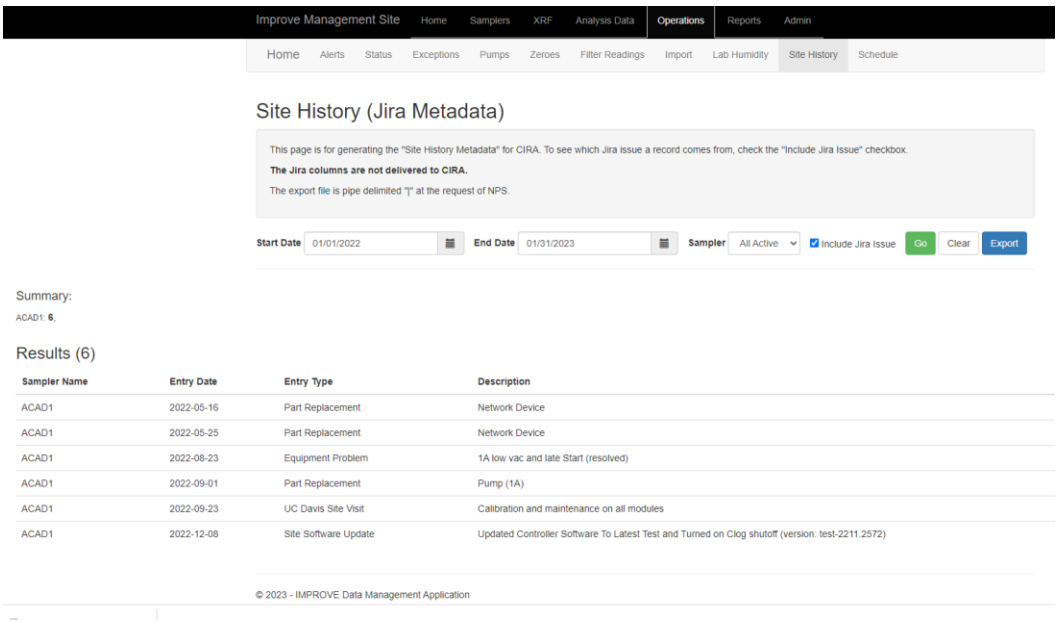
9.4 Annual Site Metadata Report

9.4.1 Generating the report

UCD produces an annual report called the Site Metadata Report (also known as the Site History Report) that documents site-related issues. UCD collects the information from our internal issue tracking system, Jira, using the Site History (Jira Metadata) tool on the IMPROVE web app. The report is sent to the National Park Service (NPS) after the annual redelivery of data for the year that is completed. To prepare the report,

- Access the Site History tool using the following link. (https://improve.aqrc.ucdavis.edu/Jira/Metadata)
- Alternatively, the link can be accessed from the Site History subtab in the Operations tab in the IMPROVE Web app also known as **UCD IMPROVE Data Management Website** https://improve.aqrc.ucdavis.edu/.

Figure 16. Site History Screen Before Generating the Report.



- To create the report, choose January 1st as the "Start Date" and December 31st as the "End Date" for the reporting year as shown in Figure 16. Select "All Active" from the sampler dropdown menu. You can generate the report with or without Jira issues by selecting the "Include Jira Issue" check box. The report should be sent to NPS without JIRA issues, so leave the check box unchecked. For reviewing the report JIRA issues, make sure to check the "Include Jira Issue" box.
- Users have two options for accessing the report: either review it on the webpage or download a text file in the format of .txt. To generate the report on the webpage, click the "Go" button. Once generated, the screen will display a view like Figure 17. To export the report to a text file, click the "Export" button. The user doesn't have to create a report on the website to get a text file, but keeping the webpage open for reference while checking the exported file is recommended.
- In the U:\IMPROVE\Documentation\Annual Site Metadata Report\ folder on the U drive, create a folder for the relevant year and save the text file in that location. The file name will be in the format: “SiteHistoryYYYY-MM-DDtoYYYY-MM-DD” where the first date is the date of the first Jira ticket for the year included in the report, and the second date is the date of the last Jira ticket for the year included in the report. If the dates differ from 01-01 and 12-31 of the relevant year (YYYY), check with the software and field groups to ensure the date range is correct. Consider adding ‘_wJira’ to the file name when saving if the “Include Jira Issues” check box option was selected when the report was generated.

Figure 17. Site History screen after generating the report.

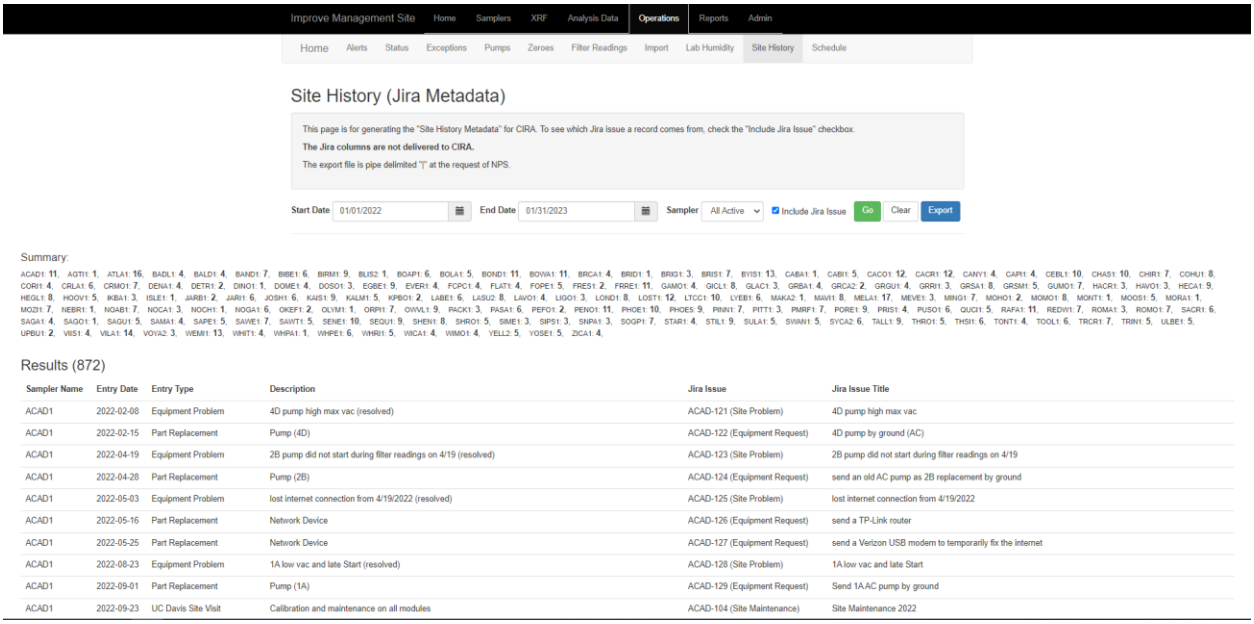


Figure 17 above was generated by selecting Start and End dates, all active sites, Jira issues and clicking the Go button.

The text file for the site history report contains four columns: Sampler, EntryDate, EntryType, and Description. If the file includes a Jira issue, it will have two more columns: JiraIssue and JiraIssueTitle. The files are delimited with a pipe character.

9.4.2 Checking Site History report files

Various data analysis tools are available in the R environment to check the report. In order to review the text file in the R environment, the analyst will need to load the tidyverse package using the command given below.

library(tidyverse)

To read the file into the R environment, use the command below:

[
[Hist_repo_wjira] <- utils::read.table(file path, header = TRUE, sep = "|")

The argument *'file path'* has two parts; the path to the site history text file generated and the file name of the text file as mentioned in the steps for generating the site metadata report using the Site History tool. The file path is usually in the format "U:/IMPROVE/Documentation/Annual Site Metadata Report/Year/SiteHistoryYYYY-MM-DDtoYYYY-MM-DD_wJira.txt"
An example of this command that was used to generate the site history report for 2022 is as follows:

Hist_repo_wjira <- utils::read.table("U:/IMPROVE/Documentation/Annual Site Metadata Report/2022/SiteHistory2022-01-03to2022-12-29_wJira.txt", header = TRUE, sep = "|")

After loading the file in the R environment, it's important to check if the number of records matches the number in the web app mentioned in section 9.4.1-Generating the report. In case of any discrepancy, report the issue to the software group and attach the downloaded file.
In addition, the following checks can be performed:

- Confirm that the dataset's earliest and latest dates align with the reporting year and filename. This can be achieved by executing the following commands in the R environment:
  min(Hist_repo_wjira$EntryDate)
  max(Hist_repo_wjira$EntryDate)

If the dates at the beginning and end of the file do not correspond with the file name, please notify the software team.

- To make sure that all items in the file have Entry Dates, run the code given in the R console. If there are any Jira issues without an EntryDate, contact the field group and request them to update the relevant fields.

  hist_reportj_nodate <- Hist_repo_wjira %>% dplyr::filter(EntryDate == "")

To make it easier to check the file for accuracy, you can group entries by their EntryType column. Common entry types include Equipment Problem, Part Replacement, UC Davis Site Visit, Site Software Update, Filters Problem, Sampler Flow Adjustment, Sampler Flow Check, Other Problem, Networking problem, External Audit, Power Problem, and Structural Damage Problem.

Use the following R code to identify the EntryType and the corresponding number of issues in the report. If an EntryType is missing, contact the field group to verify its exclusion. Additionally, report the number of issues discovered during checks to the field group and confirm their accuracy.

```
hist_report_types <- Hist_repo_wjira %>% dplyr::select(EntryType) %>%  distinct()

hist_report_types_no <- Hist_repo_wjira %>% group_by(EntryType) %>%
summarise(num = n())
```

To review each EntryType in more detail, use the *filter* function in the dplyr package. An example of the R code is given below.

```
hist_report_EP <- Hist_repo_wjira %>% dplyr::filter(EntryType == "Equipment
Problem")
```

Once each entry type is separated into lists, the user should perform various checks. Some checks are performed by stepping through the report. For example, when reporting Site issues and Equipment Problems, the description column usually specifies the affected module. However, if all four modules are affected, the module may not be mentioned. To avoid confusion, make sure to include the module in cases where only some modules are affected. Also, for the Equipment problem, ensure it's marked as resolved in the Description. Another check that can be done is to check if any two consecutive Jira issues have a similar title. If so, please get in touch with the field group and determine if it's appropriate to have both items included in the report.

Certain checks can be performed efficiently using data analysis tools available in the R environment.

For example, the Entry Date should match the report's year and the Jira Issue Title. The user can utilize the code below to identify Jira issues that have mismatched Entry Date years and Jira Issue Titles.

```
library(stringr)

hist_report_vis <- Hist_repo_wjira %>% dplyr::filter(EntryType == "UC Davis Site
Visit")

hist_report_UCD_year <- hist_report_vis %>% mutate(year = substr(EntryDate, start =
1, stop = 4)) %>%  mutate(match =str_detect(JiraIssueTitle, year))

hist_report_UCD_year_mismatch <- hist_report_UCD_year %>% dplyr::filter(match
== "FALSE")
```

To ensure the proper assignment of Jira tickets to the sampler, check that the correct Jira issue key is used for sites with the same project code. As an example, Phoenix has two active sites:

PHOE1 and PHOE5. The Jira issue key for PHOE1 is PHOE, and PHOE5 is PHOEV. You can verify this by reviewing the results of the following command and confirm the sampler name matches with the Jira Issue.

View(subset(Hist_repo_wjira, grepl("PHOE", Sampler)))

Another check is to determine whether all active samplers in the year of interest have a Jira issue. For sites that do not have a Jira issue, contact the field group and confirm it is correct. The following R code is used to do the checks.

```
Active_samplers <- crocker::db_query("select s.Name as
Sampler,SiteName,s.AffiliationCode from [Improve_2.1].sampler.Samplers s
left join [Improve_2.1].sampler.SamplerLocations sl on sl.SamplerId = s.Id
left join [Improve_2.1].sampler.Locations l on l.Id = sl.LocationId
where EndDate is null or enddate > 'yyyy-mm-dd", database, server)
```

The *enddate* is the last day (12/31) of the prior year of reporting year i.e., for 2022, the *enddate* is 2021-12-31. The *database* is '[Improve_2.1]', and the *server* is 'production.'

The sites with affiliation code SPECIAL, except for BYIS1 and RENO, are not reported to NPS. To exclude the sites that are not reported from the site list, the following R code can be used:

```
Active_samplers_reduced <- subset (Active_samplers,!(!( Active samplers $Sampler
%in% c('BYIS1','RENO1','RENO2','RENO3')) & Active_samplers
$AffiliationCode=="SPECIAL"))

noJIraIssue <- anti_join(Active_samplers_reduced, Hist_repo_wjira, by="Sampler")
```

After resolving all the issues, follow the steps mentioned in section 1.1.1 to generate a new text file. Verify the corrections by conducting the checks again. To generate the final report in text format, use the Site History tool and uncheck the "Include Jira Issues" box. Send the report to the NPS by email.

10. QUALITY ASSURANCE AND QUALITY CONTROL

10.1 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.

10.2 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.

10.3 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.

11. REFERENCES

UCD IMPROVE SOP #351: Data Processing and Validation