

Investigation of Current SLT QA/QC Practices for Facility Inventories

1 Introduction

The goal of this research project is to understand the types of Quality Assurance (QA)/Quality Control (QC) procedures State, Local, Tribal Authorities (SLTs) implement when collecting facility inventory (FI) data to meet their air emissions program requirements, to ensure that the Combined Air Emissions Reporting System (CAERS) contains a complete set of QA/QC procedures. For compliance with the [Air Emissions Reporting Rule \(AERR\)](#) towards the creation of the [National Emissions Inventory \(NEI\)](#), SLTs must report emissions from their facilities (point sources) to the US EPA through the [Emissions Inventory System \(EIS\)](#). This system applies QA/QC checks to incoming data. While EIS QA checks and procedures are thorough and comprehensive to maintain NEI data integrity and validity, in addition, SLTs set their own specific QA/QC procedures to guarantee the data quality for their emissions inventories, beyond EIS QA procedures.

Those SLT-specific QA procedures are based on SLT rules and requirements for data fields included in NEI and for SLT-specific data fields not required for NEI by the federal program. Like EIS QA checks, SLT-specific QA checks will notify users of errors that will not allow them to save the data, and warnings that allow them to continue working but indicate there is a potential issue. The SLT emissions inventory reporting systems may also restrict user access to certain data fields to prevent reporting errors. By understanding the QA/QC procedures that SLTs apply beyond those required by the NEI, CAERS can be enhanced to include the same SLT-specific procedures or procedures that accomplish the same outcome in terms of data quality.

Because CAERS must have a standardized set of QA/QC procedures that meet EPA and SLT requirements, this report provides guidelines on types of SLT-specific QA checks that CAERS should adopt and suggests that at least the most prevalent checks be incorporated into CAERS for SLTs. The inclusion of the most complete set of QA checks in CAERS will ensure that these are performed as early in the reporting process as possible: at the point where the facility is reporting the data, instead of once the data is being sent to EIS months later.

When CAERS assists the facility in reducing or eliminating reporting errors early in the process, SLT and EPA staff can spend less time performing QA/QC procedures themselves, finding and correcting errors months after industry has sent in their reports, and sometimes after the data has already been submitted to EPA, and potentially sending reports back to industry for rework. Instead, both EPA and SLT staff can repurpose their time on performing more advanced QA and analysis of the data, and industry reporting time becomes more productive.

The CAER Product Design Team (PDT) has started conducting research on specific State, Local, and Tribal (SLT) authority requirements on facility inventory (FI) data. The Facility Research and Development (R&D) Team conducted a survey on FI creation and maintenance practices that SLTs follow for air emissions inventory reporting. Results and analysis of survey data is being presented in multiple parts.

In the first part, the team analyzed and reported on data sources and data flows that SLTs use for obtaining and updating FI information (facility site, emission units, site controls, site control paths, processes, release points, and release point apportionment) for their respective emission inventories. Recommendations made for CAERS were documented in the report titled “Investigation of current State, Local, and Tribal (SLT) data for facility inventories”. Appendices A through D contain detailed information about the survey and response data.

This report represents the second part of the Facility R&D team’s work. Here, the team has focused on QA/QC checks that SLTs need to perform on FI data for the purposes of ensuring CAERS has a complete set of SLT-specific QA checks, in addition to those required by NEI.

2 Background and Previous Work

Previous work has been conducted on QA/QC procedures by the PDT and for CAERS as follows:

- Starting January 2017, the CAER PDT conducted a study on QA/QC procedures for emissions inventory reporting. The conclusions of that study were documented in the final report on CAER QA/QC that can be found on the [CAERS PDT website](#). One aspect that study focused on was which QA/QC checks SLTs performed that were automated given the value of having automated QA checks in reporting systems. That project provided a suggested list of QA/QC checks for CAERS, as well as the possibility of sharing these checks with SLTs for use with their own systems via a shared service.
- The CAERS Minimum Viable Product (MVP), or first version of CAERS was released in 2020 and contained many of the suggested SLT QA checks that were added to the list from that study, as well as all EIS point source QA checks that were feasible to add, meaning, those that do not require a “call and response” between CAERS and EIS, a workflow that has not been built out yet but is planned for future releases. Georgia Department of Natural Resources (GADNR) staff assisted by providing feedback on necessary QA checks for their industry, as GA piloted the MVP with EPA. Since release of the CAERS MVP, through the Agile process, QA checks have been added to CAERS in response to feedback from industry and SLT users, as well as reporting errors that surfaced during reporting that can be prevented. In addition, every year when a new version of CAERS is released, it contains any new QA checks that EIS may have added where feasible to implement. See Appendix E for a detailed description of the EIS data submission process and a detailed analysis of EIS QA checks in Appendix F.
- CAERS currently applies FI QA/QC checks as follows (see Appendix G for details):
 - point source checks applied by EIS on incoming SLT reported data, so long as the check does not require a “call and response” between CAERS and EIS, a functionality that is still a work in progress,
 - additional QA checks requested by SLTs who participated in the PDT QA/QC R&D Team described above where feasible,
 - additional checks requested by current CAERS SLT users, including custom checks requested by some SLTs but not desired by others, and
 - QA checks that have been developed as use of CAERS has revealed their need.

3 Method

In April 2022, the facility inventory team conducted a questionnaire survey among SLTs (see details in the document and appendices of the “Investigation of current State, Local, and Tribal (SLT) data for facility inventories” report). The survey contained only two questions related to QA/QC procedures. SLTs were asked:

- if there are any SLT-specific QA/QC procedures or restrictions (in addition to EIS QA/QC checks) for facility inventory components,
- if SLTs have encountered issues and problems with their data in preparing facility inventories.

If the respondents answered yes to one of these questions, they were asked to explain further.

The team collected a suite of SLT-specific QA procedures through the survey and communications with SLT EI staff. Using the information from the responses, the team’s tasks included:

1. analyzing the QA/QC procedures and practices that SLTs use for facility inventory data,
2. comparing the current types of quality checks that SLTs apply with those already in CAERS, so that CAERS may adopt any additional SLT checks that could improve the accuracy and quality of the data further.

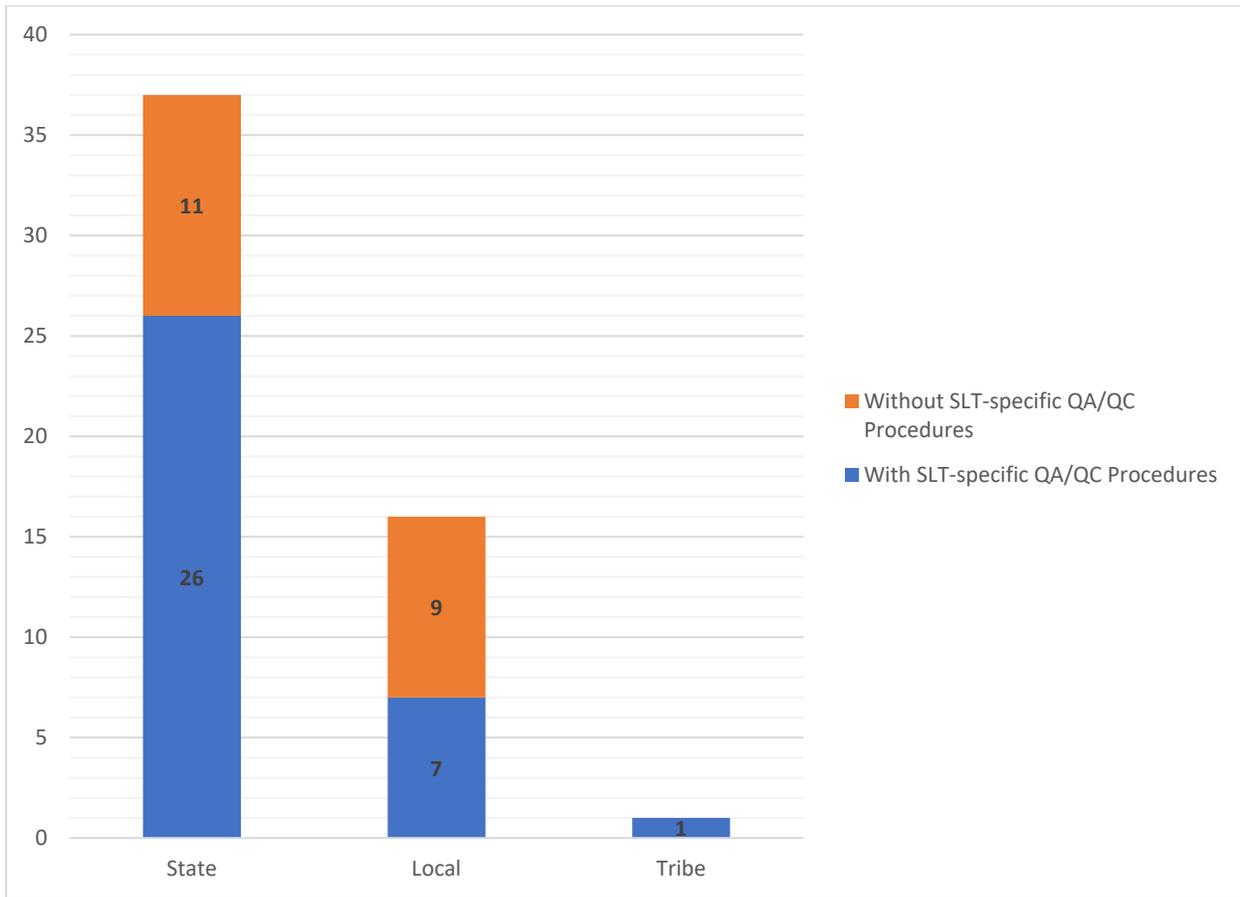
4 Analysis of Current SLT QA/QC Procedures

4.1 SLT QA/QC procedures beyond EIS from Survey Responses

Fifty-four jurisdictions responded to the survey (Appendices A-D). Thirty-four SLTs among the 54 SLTs that responded to the questionnaire indicated they conduct SLT-specific QA/QC procedures or apply data restrictions in addition to QA checks applied by EIS. Figure 1 shows the number and type of jurisdictions that have additional QA/QC procedures to those applied by EIS: 26 of 37 (70%) of State respondents, 7 of 16 (44%) local respondents, and the single tribal authority respondent all apply QA checks in addition to EIS.

Since the questionnaire was not explicitly designed to ask detailed questions about SLT QA/QC practices, the results provided only general information from SLTs who provided additional explanation about their procedures. The following information was gathered from their comments in the survey, as well as follow up conversations, about the types of procedures SLTs utilize to QA their data. CAERS should be able to assist SLTs in applying customized QA procedures so that when an SLT uses CAERS, it does not lose these capabilities:

Figure 1. Number and Type of Jurisdictions with QA/QC Procedures beyond EIS Checks



1. Restrict edits for certain data elements to prevent erroneous reporting from facilities.

- a. Do not allow facilities to change their facility site IDs and emission unit IDs to maintain the integrity between the EI data and permitting data and the consistence of IDs across years.**

For example, Idaho does not allow facilities to enter any unit agency (IDDEQ) ID. The system notifies EI staff every time a unit has been added. At that point, staff assign an IDDEQ ID to the new emissions unit. Once the ID is assigned, the state never changes it. However, ID has a field for the facility to add their own IDs, and these are allowed to change.

While many SLTs that have QA/QC procedures beyond EIS procedures have this restriction, there are different restrictions for other data elements related to identification of data components.

Currently, CAERS does not assign IDs automatically. However, facility reporters are not allowed to change agency IDs. This is because if they were to do so, EIS would receive these units and consider them new, because they wouldn't have an agency ID that EIS recognizes. The result would be duplicate units in that facility's inventory. However, CAERS does alert the SLT that a new unit has been created, by issuing the facility a warning that requests that the user check with their SLT as to if/what naming convention the SLT requires. The warning also alerts the SLT that a new unit has been created.

- b. Do not allow facilities to change their process IDs, control IDs, and release point IDs.**

- For example, Massachusetts and Illinois uses this restriction.
- In Michigan, facilities can update the names for these data elements but not the IDs.
- Wyoming’s system generates noneditable IDs for all elements of the facility tree (facility, emission units, processes, controls, etc.), while a separate data field also exists for the operator to enter their internal, company IDs for each element.
- Montana does not allow facility to update emission unit and process IDs.

As with units, CAERS currently can issue QA warnings if a new component is created. It does not allow reporters to delete a previously reported component. To remove it, it must be marked shut down.

c. Do not allow facilities to change all or part of their facility-level information.

- For example, OK, Forsyth County in North Carolina, Missouri, and Rhode Island do not allow facilities to change their facility name/address.
- South Carolina does do not allow facilities to make facility-level (i.e., name, address, EIS category, location, contacts, etc.,) changes in the reporting.

Currently, CAERS does not allow reporters to alter the facility level information. The SLT is enabled to edit most of those fields except those that EIS won’t allow changes to (for example, the facility coordinates). Also, CAERS can be customized further, for example, GADNR does not allow industry to edit the facility NAICS.

d. Do not allow facilities to change any process-level information, excluding those in the EIS Point submittal, within EI systems.

In Maine, a facility needs to call EI staff for changes because these details are in the facility license issue and changes trigger a licensing update.

e. Do not allow facilities to change process SCCs.

Delaware, NC, and Texas are examples of this.

For these two items, CAERS would be customized further to restrict specific process-level information, and/or SLTs can allow the facilities to edit these but get warnings so they may ensure that industry has made any edits correctly.

2. Set allowances or conventions for editable data elements, particularly, for entering a new component.

- For example, in NC, facilities can enter a new emission unit or control device, but the EI system adds a U- to the IDs of both to flag them as new and not in their air permit.
- Minnesota’s on-line emission reporting system allows facilities to add new emission units, processes, controls, and release points, and auto creates IDs in the sequences of the existing once.
- In ID, process IDs are automated and are numbered 1, 2, 3, etc.
- The restrictions or conventions in number 1 and 2 are controlled and enforced by the SLT emissions reporting systems. For example, Kansas uses a customized version of Windsor's State and Local Emissions Information System (SLEIS) that restricts what information

facilities can change. Other SLTs using SLEIS (such as DE, SC, and Hawaii) or their own EI reporting systems (such as MN and WY) have similar approaches to restrict editing data elements although the restricted data elements may vary from SLT to SLT.

- If facilities want to make changes for the restricted data elements, they must get approval from the EI staff or through permitting amendments/revisions.

More detailed or customized restrictions or QA checks of this nature would have to be built out in the SLT CAERS module.

3. Require additional data elements to make sure information for critical data elements is properly reported.

For example, Kansas allows facilities to change operation status (such as shutdown of a process), but changes must be quantified with dates to be reported properly.

CAERS will be able to allow SLT-specific data fields and/or codes in future as part of the SLT's module.

4. EI staff have full control of the EI data submitted to NEI.

- For example, in Southwest Clean Air Agency, Washington, all data submitted by facilities are on forms provided by the agency, and all data is independently verified by the agency staff. Facilities are not allowed to alter the forms or the provided information (e.g., stack info, EU info, etc.).
- In Jefferson County, Alabama, point sources (major sources) are required to submit/update all emission units/stack info that generate emissions to our Permitting Section during permit applications/renewals to operate within the county. The Permitting section verifies and documents all information reported. The EI group obtained this info by checking the Permitting database/spreadsheet for EI purposes.

For data tied into an SLT's permitting data base, see the facility inventory workflows that were discussed in the previous report by this team. See "Investigation of current State, Local, and Tribal (SLT) data for facility inventories" report. For the forms that the SLT might require, CAERS can intake the data that the SLT does allow the facility to edit, and can prevent edits of other data, per the required customization. CAERS may also generate reports in an SLT-required format if needed.

5. Analyze and run QA checks on data before the EI cycle begins and after EI reporting is complete.

- For example, OK does an analysis to determine which facilities need to report to the state for the upcoming cycle, prior to collecting inventories from their permitted facilities. Many QA checks are run during this analysis. It also runs in-house created queries on SLEIS data to identify other errors after the reporting season is complete, such as for operating status.
- Wisconsin runs SQL queries to QA the data and has several QA checks built into a QA report.

CAERS uploads a previous year report into the new report as a starting point for the reporter to begin. At that point, any new QA checks that may be available in CAERS can be run so the reporter knows of any errors they may have. In addition, CAERS does not allow reporters to submit their reports unless they have passed all critical errors. All warnings can be observed by the SLT reviewer as well, so that if there

is an issue, the reviewer can send the report back to the facility before sending the data to EIS. SLT-specific QA checks are included in the global checks both before and after reporting.

Also, in CAERS an opt in/opt out questionnaire was created as a customization for the state of GA. This means facilities can determine if they will be reporting for a specific inventory year, and GADNR requires them to attach an analysis to demonstrate that, if they have opted out, that this is appropriate for the facility. Such a process could be further customized for new SLTs desiring a similar approach.

6. Conduct special QA/QC activities for geographic data.

In Illinois, records group staff conduct QA/QC for uniqueness at the address level. In addition, GIS staff are used to locate facility coordinates and check for potential duplicate entities.

While CAERS does not have a “call and answer” capability with EIS at this time, to verify duplicates, a future functionality would allow this type of QA check to be applied before the report is certified and submitted. At this time, however, latitudes and longitudes of facilities are locked so that an SLT must request an unlock to change these. And the SLT can set the facility latitudes/longitudes for a new facility when it is entered in CAERS.

7. Build conditional QA checks in the EI system.

- In Kentucky, all processes must have a numeric value reported for annual throughputs (cannot be blank).

CAERS could issue a QA check for an SLT that does not want blank values.

4.1.1 QA/QC from PDT Call Discussions

Connecticut has many and complex QA/QC that they didn't specify in the survey, and other SLTs might also have additional QA checks not captured by the survey. Therefore, several conversations were held to discuss SLT QA procedures further with PDT members. These PDT call discussions provided additional QA measures as well as more details on the SLT QA procedures identified in the survey. This section presents the information collected from discussion with PDT members. Besides built-in EIS QA checks in the SLT emission reporting systems, there are also other techniques used in QA procedures. The analysis here focuses more on additional QA measures and techniques.

1. Calculate release point parameters by SLTs.

In CERS V2.0, the units of measures for release point operation parameters must be specific Imperial units. For example, release point stack height and diameter must be in feet, exit gas temperatures must be in Fahrenheit, exit gas velocity must be in feet per minute or feet per second, and exit gas flow rate must be in actual cubic feet per minute or actual cubic feet per second. Facilities might have trouble with unit conversions. ID calculates the values for facilities. ID also performs the calculation of diameters for non-circle stacks.

In future, additional conversions could be added so that more conversions are possible in CAERS. Also, SLTs may indicate that they would like to verify release point parameters and then have them locked so reporters may not edit them.

2. Special efforts on geographic coordinates.

Geographic coordinates use an intersection of two lines of latitude and longitude to determine the geographical point of a facility site or a release point, for example, such as a latitude of 46.992611 and a longitude of -93.604936. Accurate identification of geographic coordinates is critical for using emission data in risk assessment and air quality planning. SLTs take different approaches to QA/QC these data. Examples are shown below:

- TX locks out facilities off editing geographic coordinates. If facilities want to change coordinates, they must map the coordinates and make the coordinates more accurate. Therefore, facilities spend a lot of time correcting geographic coordinates. If facilities use the bulk upload, they will get an undone message for the coordinates, and EI staff will have a chance to look at the coordinates.
- NC sets reference points for geographic coordinates for the entrance point (front door) of a facility. Front door is the main office building that could be a substantial distance from the street address, for example, a facility with a long entrance drive. NC does not use street addresses to determine geographic coordinates. In NC, 3 facilities have the same street address but separate office buildings.

As described above for other data, customizations would allow more data fields (coordinates for different location points of the facility) and these could be non-editable by the reporter as needed by the SLT.

3. Check the consistency between facility inventory in SLT EI system with data in the permitting system.

For example, if a permit lists controls, then ID EI staff check to ensure they are listed along with emission units, and everything has been submitted to the emission inventory.

In CAERS SLTs may enter data for a new facility, and thus, enter the data as it is shown in the permit. As described above, facility reporters could be prevented from editing certain data fields, or QA checks could be issued, per the SLT's preference, so the SLT may verify that any edits by the facility reporter are correct and align with the permit.

4. SLT EI systems auto fill missing information.

- The MN's emission inventory rule requires facilities that have certain types of state registration permits report facilities total emissions. Permitting data in the master database contain only facility-level information, no sub-facility-level information, such as emission units, processes, and release points. The state EI system collects SCC-level emissions through online reporting from facilities and auto generates the sub-facility-level information based on SCCs.
- NC and ID systems automatically assign a new process with an ID, where the ID number is assigned sequentially to each additional unit/process.

While adding a capability that automatically assigns a new ID for a facility component, following an SLT naming convention, could be part of an SLTs module in future. Another option is to have QA checks that

verify that the naming convention has been followed, or warnings that allow the SLT to see if a new ID provided by industry has been assigned correctly.

5. Restrict facility's ability to delete certain facility inventory records.

ID does not let facilities delete controls, release points, emission units, or processes, unless facilities mark them as permanently shut down or a newly added emission unit.

CAERS has the same process as ID and does not allow facilities to simply delete previously reported sub-facility components. These must be either temporarily or permanently shut down.

6. Use dropdown lists to enforce valid and acceptable reporting information.

Many SLTs have their own codes for facilities to report dates in the EI systems. MN provides dropdown code lists for SCCs, pollutants, site control types in the EI reporting system.

CAERS has the same process where a code cannot be entered directly but must be selected from a drop down both in the user interface and the bulk upload template. If the SLT were to enter an incorrect code in the bulk upload template, CAERS will not allow that data to be uploaded. CAERS will also not allow an outdated code (such as a retired SCC, for example) to be used and will force the user to choose a valid code.

7. SLT EI systems perform QA check with start and end dates to ensure the validity of components.

Oklahoma is using SLEIS to program the QA checks with start and end dates. When a component comes into play, e.g., a new emission unit, a start date must be entered for it. If retiring a component, an end date must be entered, and people cannot use these units past their end dates.

This functionality is not available in CAERS yet, but SLTs could add their own start and end dates as SLT-required data fields associated with each sub-facility component if they wish to do so in their module.

8. Get notifications when facilities alter the existing facility inventory.

ID EI staff receive notifications when there are new units entered, so EI staff can be sure to review what the facility did before the facility submits.

CAERS already has this functionality and is applying it for reports from ID and ME.

9. Send an announcement for starting the current emission inventory reporting with notes for common issues observed in the last emission inventory.

For example, facilities may use the 'End Date' for a process as the end date for the emission inventory. It should only be entered if no emissions are being reported for the process and the process will no longer be used. MN sent this note with the announcement of collecting the 2021 emission inventory.

CAERS has the capability to allow SLTs to send email notifications to industry reporters through CAERS itself. As reporting progresses throughout the year, QA checks can be added to prevent prevalent and previously un-anticipated errors from propagating further. Finally, annual trainings address aspects of CAERS that may have been confusing or where errors in reporting were observed to prevent them from happening in future.

10. Contact facilities when problems are observed and cannot be explained by the existing in-house information.

For example, in MN, information for site controls is from the state master database. However, some information could be incorrect, such as the end operation date. To confirm the correct information, EI staff need to contact facilities with the operation status to make sure the control efficiencies from the site control could be included in emission calculation.

SLTs using CAERS are still able to contact individual reporters to clarify questions.

4.2 Discussion and Recommendations

In the previous section, CAERS capabilities that were available at the time of this study were mentioned. In this section, we summarize functionalities that would still be desirable to have in CAERS in its future functionality.

4.2.1 QA Checks from SLT Survey

1. Do not allow facilities to change process SCCs

While it is currently possible to edit an SCC in a process in CAERS, this feature is not desired by many and thus, the CAERS team should explore preventing this ability for SLTs that want to avoid SCC changes, and/or issuing warnings when an industry reporter has modified an SCC to ensure the SLT is made aware of and agrees with that change.

2. Allow reporters to change their own facility component IDs without altering the agency ID's that the SLT and EIS recognize.

In future, and given feedback provided by current SLT reporters as well as previous and current PDT SLT members, the CAER team would like to explore the ID option to allow for the facility to relabel their components without this action affecting the Agency and associated EIS IDs. This would give the facility reporter flexibility while allowing the SLT to maintain certain control over the desired Agency IDs.

3. Set allowances or conventions for editable data elements, particularly, for entering a new component.

Currently, CAERS does not automatically create an ID for a specific SLT. However, given that the SLT may see warnings if a new component has been created, the SLT may require the facility to enter the data using their naming convention, and return the report if the convention has been violated.

In following EIS convention, CAERS will also not allow a sub-facility component ID to be re-used if one already exists. For example, a unit ID may not be used more than once. So for example, a new Unit that is accidentally labeled with an ID already in use by another Unit, will generate a critical error.

In future, the CAERS team would like to explore customization of ID's for SLTs. While it may prove difficult to incorporate specific functionality to create specific types of ID's automatically for each SLT, it may be possible to issue SLT-specific QA checks that, for example: alert the user that a specific

number has already been used by comparing the Id's for specific components with those marked PS in previous year reports, or that an expected prefix or suffix has not been added, may be the most quick and easy way to incorporate such customizations.

4. Require additional data elements to make sure information for critical data elements is properly reported.

In future, CAERS will allow SLTs to collect additional data they would like to enter for facilities, or that they require from facilities for their specific programs.

5. EI staff have a full control of the EI data and facilities cannot access -the EI system.

As mentioned above, while CAERS already restricts changes in facility level data, the previous document from this team describes more sophisticated workflows of facility inventory data that would have to be built to allow EI staff more control if desired.

6. Conduct special QA/QC activities for geographic data.

Future functionality where a "call and answer" workflow is possible between CAERS and EIS is desired so items such as potential duplicates, previously used ID's, for example, can be checked before the reporter submits their report, so that the report does not contain errors that could potentially trigger errors from EIS.

SLT-specific geographic information and QA checks would have to be added as functionality in CAERS as well.

7. Build conditional QA checks in the EI system.

As described above, customized QA checks are possible in CAERS. Conditional QA checks can be applied in CAERS, as can be stricter QA checks than for the federal program, so long as they are not in contravention of federal program requirements, and so long as the SLT has the legal authority to impose any additional restrictions or require additional information from industry.

8. Calculate release point parameters by SLTs.

CAERS allows for conversions in the system but does not include all UOMs. However, soon it will be possible for SLTs to retrieve their data automatically at which point necessary conversions can be performed. In future, CAERS could also issue certain reports to SLTs in specific UOMs for SLTs who prefer metric system UOMs.

SLTs may customize CAERS and add data fields or codes to CAERS for their own uses. For example, in future, they may be able to collect additional data on release points as needed.

4.2.2 QA/QC from PDT Call Discussions

1. Allow SLTs to calculate release point parameters.

An SLT may indicate they would like to verify and then lock the release point parameters. This customization could be built in CAERS.

2. Check the consistency between facility inventory in SLT EI system with data in permitting system.

In future, CAERS will allow SLTs to include the permit numbers and types for their facilities so that they may easily reference the permits, as needed. And for more sophisticated workflows between SLT FI and CAERS, see this team's previous report as described in the Background and Previous Work section of this document.

3. SLT EI systems auto fill missing information.

An SLT wishing additional auto-calculated data from CAERS may not necessarily need to generate it in CAERS but may generate the needed information when CAERS pulls data back into its own system. For example, if certain facilities are not going to report emissions inventories, the SLT may be able to generate the sub-facility data in their own system by SCC. Conversely, the SLT may be able to autogenerate the relevant information at the SCC level for facilities that only report totals in an SLT system, and have that data sent to CAERS from where it can be sent to EIS with its other information.

4. SLT EI systems perform QA checks with start and end dates to make sure the validity of components.

In future CAERS would need to track start and end dates for QA checks. More discussion is needed as to which QA checks should be versioned and which checks should be applied retroactively to previous year reports even if the QA check is new.

5 Future Work on Facility Inventory

Beyond the survey questions covered in this part of the research, the team identified future work that is needed to allow SLTs to have their facility inventories in CAERS. Such research should include:

1. Handling emissions for a parent facility with multiple child sites.

- For example, a nonmetallic facility in MN has only one facility ID, but multiple operation sites. The MN emission inventory rule only requires those facilities to report total emission and pay emission fees based on the facility ID, not on the individual operation sites. MN cannot handle the situation in the current EI system, therefore, takes hard copies from those facilities and manually enters total emissions for the facility IDs to the EI system.
 - Alaska has the similar situation and uses the same approach as MN for nonmetallic facilities.
 - Portable facilities also present challenges, such as asphalt plants/rock crushers that can move all over the state, cannot be assigned a specific borough/census area.
2. Handling one facility where different parts of the facility are owned by different companies.

In WY and CT, one location has two identically named facilities. For example, in oil and gas facilities in WY, the well is owned by one company but the natural gas dehydration unit(s) at the well site are owned and operated by another company. The dehydration facility owner gives it the same name as the well site facility name.

Appendix A. Original Letter and Questionnaire

See Appendix A. Original Letter and Questionnaire.docx

Appendix B. EIS facility staging table requirements

See Appendix B. Facility Staging Requirements.xlsx

Appendix C. Responses to the Questionnaire

See Appendix C. All Original Responses.pdf

Appendix D. Data analysis

See Appendix D. SLT Facility Inventory Research.xlsx

Appendix E. EIS Data Submission

EIS Data Submission

Data submitted by SLTs for NEI reports via EIS must undergo QA checks. The EIS Quality Assurance (QA) Environment is used as a preliminary quality assurance step prior to making an official submission to the Production Environment. Users are encouraged to use the QA environment as frequently as necessary to help ensure that the production submission is of the highest quality.

The QA Environment will apply checks to submitted data that ensure file integrity for submission purposes and will apply checks that may reference data stored in the EIS Production Environment. Most importantly, this is the QA stage that will give users advance notice that certain data will be rejected if they are submitted to the Production Environment. EIS issues:

- critical errors that must be corrected for the report to be accepted,
- warnings, which alert the SLT that, while technically correct, the submission may still contain issues that the SLT may want to review.

Any errors in the data will be noted in feedback reports that will provide users with a listing of errors that need to be corrected (for example, missing data, inconsistencies in data, invalid files), and indicate how the submitted data would be integrated into the EIS Facility Inventory. After correcting all errors in data, users can make the official submission to the EIS Production Environment.

In the EIS Production Environment, as part of the submission, the same checks as those used in the QA Environment are run during the batch submission process. The results of these checks are logged in EIS. Users again receive a feedback report that indicates critical errors and potential issues upon submitting their XML file to the Production environment. Users must correct the problems with data content, or the XML document structure listed in the report and resubmit the file to ensure all data submitted is checked.

In the Production Environment, when users make additions, deletions, or edits on a limited data set, QA checks will be run only on the data associated with or related to the data which have been changed or added. Users will immediately see the impact that minor additions may have.

In addition, EIS may prevent certain data from coming in entirely. For example, the geographic coordinate information for facility sites is protected and have been verified using Google Earth for accuracy, so that information cannot be overwritten or edited with an EIS submission. For data of that kind, the SLT must reach out to EPA and EPA may then unprotect the data to allow for the submission to overwrite it.

When users make single record additions or edits to the EIS Facility Inventory data on the EIS screen, EIS will run checks only associated with the single record data that were changed or added by the online transaction.

Besides schema validation checks and file validation checks, the EIS Gateway will also prevent the following cardinality errors:

- Duplication of XML Elements within a Complex Type
- Duplication of Complex Data Types

- Duplication of Major Data Blocks

QA checks in EIS

As of 11/02/2022 EIS performs 896 QA checks for data content and format during SLT data submissions (see Appendix A). These checks are automatically performed when data are uploaded to the EIS by SLT submittals as well as by EPA loads. There are two levels of checks: critical and warning, for a variety of checking types. Table E 1 shows the number of EIS QA checks for each check type under each check level.

Table E 1. Statistics of EIS QA Checks

Check Type	Critical	Warning	Grand Total
Calculation	5	2	7
Cardinality	18	6	24
Code	92	6	98
Comparison	9	2	11
Conditional	164	22	186
Duplication	3	92	95
Format	150	45	195
Present	140	8	148
Range	124	8	132
Grand Total	705	191	896

The Facility Inventory in EIS is the permanent, continually maintained inventory of large stationary sources and voluntarily reported smaller sources, which serves as the basis for all point emissions reported to the EIS. It contains information about facility sites and their physical location, emissions units, emissions processes, release points, controls, control paths, and regulations.

While many QA checks are run to verify the facility inventory data quality, about 416 out of 896 QA checks are for point source emissions and sources other than point sources (nonpoint sources, mobile sources, and events). An analysis based on best judgement shows that there are 480 QA checks applied to facility inventory data. These are listed in Table E 2 and broken down by the simplified data components used in this study for the investigation of current State, Local, and Tribal (SLT) data for facility inventories.

The 34 QA checks on the first row are for the document header table in the facility inventory data submission or for all data components, such as, for example, check 620 (FIPS County ID Code must match value in code list), check 621 (FIPS State ID Code must match value in code list), and check 874 (Program System Code must match value in list of registered codes).

Table E 2. Number of EIS QA Checks for Each Sub-facility Component

Component	Number of QA Checks	Included QA Checks Applicable to Other Component	Other Applicable Component
All and Document Header Staging Table	34		
Facility Site	116		
Emission Unit	47		
Site Control	59	1	Control Path
Control Path	73		
Process	43	8	Emission Unit
Release Point	85		
Release Point Apportionment	23		
Not for Facility Inventory	416		

While most QA checks are unique to one data component, certain QA checks are common for a couple of data components. For example, 8 QA checks for the regulation staging table are applicable for both the emission unit data components and the process data components because regulations could be at either emission-unit level or process level. Check 838 (PM percent control measure reduction efficiency dependency), requires that PM2.5 percent control measure reduction efficiency cannot be larger than PM10 percent control measure reduction efficiency in either the site control data component or the site control path data component. In those cases, the common QA checks are only included in one data component but with additional information in the third and fourth column of Table 2 to avoid double counting the number of QA checks.

After Data are Loaded to EIS

EPA staff and contractors take additional steps to review the data used in the NEI. The QA/QC work performed after data loaded to NEI is basically for point source emissions that are emissions at facilities with specific latitude/longitude locations. The NEI is a composite of SLT-submitted data and EPA-generated data to use when SLT data are not available, mainly for hazardous air pollutants (HAPs). The reason is HAP emissions are voluntarily reported in many SLTs. Therefore, some states and pollutants are not reported to EIS. On the other hand, NEI data are largely compiled from data submitted by SLTs.

Prior to release, EPA staff generate maps for review and run many comparisons of the data to other data, such as compare inventories across versions and years, calculate and compare VOC versus summed VOC-HAP totals by county and SCC and identify outliers. EPA provides feedback to SLTs during the compilation of the data on critical issues (such as potential outliers, missing data), and requesting assistance in reviewing and editing as needed. EPA also builds on-line Data Completeness Report that shows number of point sources required to report and the number of facilities reported, a percentage completion, a percentage completion metric for the

expected HAPs, and an indicator for which facilities have “outliers” (either high or low values or missing altogether).

In the development of EPA's environmental justice mapping and screening tool, EPA asks SLTs to perform the point source review that serves as an opportunity for additional review of hazardous air pollutant (HAP) emissions in conjunction with draft modeled risk results from initial modeling of a draft inventory and identify corrections that would better estimate risks at communities near these facilities when they are modeled for AirToxScreen.

Although the above QA/QC activities focus on emissions, issues related to facility inventories could be identified, such as missing facilities in reporting, mistakes in release point parameters and apportionments, site controls, and site control paths. SLTs may correct issues and problems identified by all those EPA's QA/QC activities for their respective SLT facility inventories.

6 Appendix F. EIS QA Checks and Analysis

See Appendix F. EIS QA Checks and Analysis.xlsx.

7 Appendix G. CAERS QA Checks

See Appendix G. CAERS QA Checks.xlsx