Web-Based Emission Inventory Application:
Improving Data Integrity, Quality, and Processing Efficiency

Heather Hawkins

North Carolina Department of Environment and Natural Resources Division of Air Quality

225 Green Street Suite 714, Fayetteville, NC 28301

heather.hawkins@ncmail.net

The North Carolina Division of Air Quality (DAQ) has developed web-based interfaces for an emission inventory application. The purpose of this paper is to discuss quality assurance checks in these interface's that improve data integrity, quality and processing efficiency. To maintain order and understanding of each interface's QA checks, I will address them separately.

These web-based interfaces are the proverbial stone that kills two birds. They allow customers easier access to and faster reporting of their inventory data as well as allow the governing agency to spend less time manipulating the data and more time reviewing, trouble-shooting, and understanding the data. The principal benefits are improved data integrity, quality, and processing efficiency. Maximizing these benefits requires a system of well-designed quality assurance (QA) checks.

The DAQ created two web-based interfaces: the *facility interface*, where the facility enters and submits emission inventory data, and the *agency interface*, where the governing agency reviews, corrects, and approves the submitted data. Each of these interfaces has its own set of QA checks to ensure accuracy and completeness. From an agency standpoint, we want to ensure that submittals via a web-based interface are complete, so as to avoid delays in our review. From a facility standpoint, an incomplete submittal could constitute a violation, and the facility may incur civil penalties. Additionally, the governing agency is required to supply EPA with certain "mandatory" data elements described in the NIF 3.0 documentation. If the data are not supplied directly by the customer, then the agency may provide default information, compromising accuracy, integrity and efficiency.

Web-Based Emission Inventory Application

The North Carolina Division of Air Quality (DAQ) permits approximately 3100 facilities; approximately 400 Title V (Major) facilities that are inventoried annually, and approximately 2700 Synthetic Minor and Small facilities that are inventoried on a rolling 5 year cycle in conjunction with permit renewal. North Carolina DAQ developed a data management system to improve consistency and integration with permitting, compliance and inventory issues. The system, very basically, is a relational database supported by Oracle software. The design of this data system allows information to be stored in a centralized database and to be shared by all modules. The advantages are the elimination of redundant data-entry and the reduction of inconsistencies between permits and inventories, to name a couple.

The Emission Source Module (ESM) is the center-point module for several other modules, which are Emissions Data (ED), Air Emissions Reporting On-line (AERO), Permit Writer (PW), and Fees. The ESM is where emission sources and control devices for facilities are entered and stored. ED and AERO allow sources and devices stored in the ESM to be linked to actual emissions and emission release points for the purpose of entering and storing inventory data. Permit Writer allows the sources and devices stored in the ESM to be linked and permit stipulations to be created in order to generate the facility air permit. The Fees module uses actual emissions data that are stored in ED to calculate the annual permit fee for Title V facilities. Once source and device data are stored in the ESM they are viewable in ED, AERO, and PW. This allows users (agency and facility) to work with one consistent set of source and device data, thus reducing errors, double data-entry, and confusion. Below is a basic block flow diagram to illustrate the flow of data.

Diagram 1. Flow of Data in System
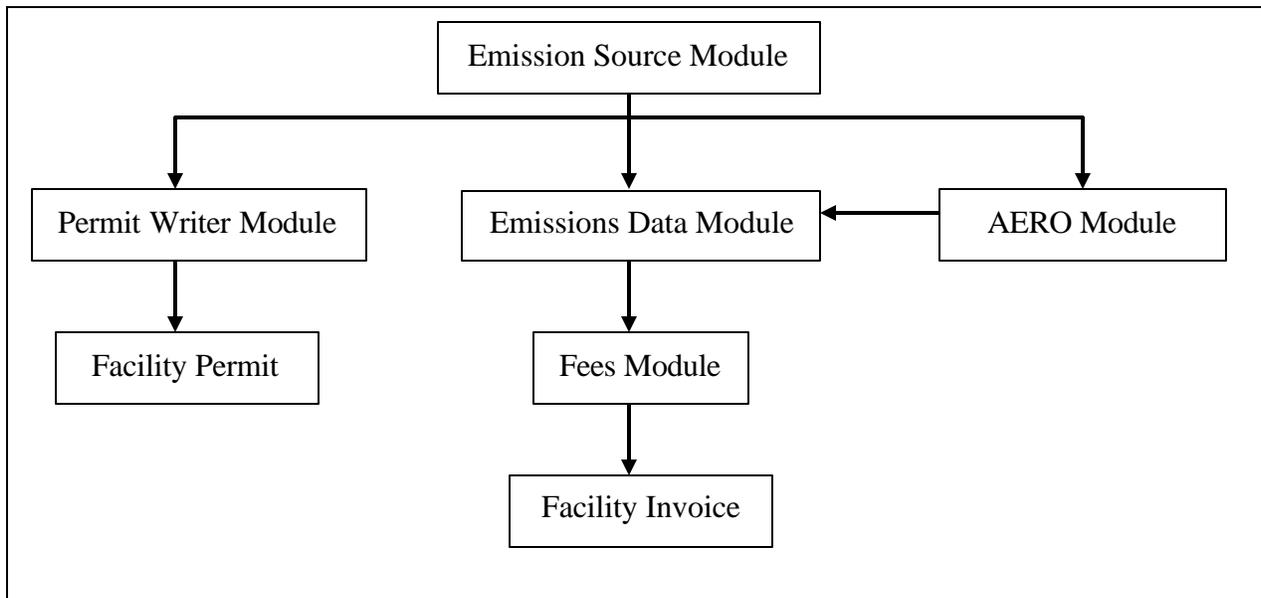
The Emissions Inventory Application

The key to a successful web-based inventory application is completely and accurately defining the data requirements prior to development of the application. Defining data requirements means determining what data will be *obligatory* from facilities, what data will be *requested* from facilities, how to ensure receipt of all the data, and how to control the quality of those data. Choosing not to address these issues up front will limit the system and could seriously detract from the goals of improving data integrity, quality and processing efficiency.

Defining Data Elements

A great starting place for defining required data elements is EPA's NIF 3.0 specifications. EPA has labeled all of their required data fields as "mandatory." Include all the mandatory data fields in the required data element list. Additionally, determine which fields your state or local government program will require that are not already listed and include those in the required data element list. The list will be complete when it includes all the fields that you think are "must haves." Once the required data element list is complete move on to the "nice to have" data element list. EPA has labeled these fields as "necessary." Include in the "nice to have" list all data fields that you think EPA or your state or local government program may need in the near future. Increasing data integrity is a gradual process and is directly related to how completely the system's data elements have been defined. Data integrity can be defined as the soundness and completeness of data. Only when facilities submit complete and known (actual) data to the state and local government agencies does progress toward improved data integrity begin. When government agencies accept incomplete or inaccurate data a trend of decreased data integrity begins.

Quality Assurance Methods

A web-based inventory application that is structured to allow optimal data integrity, quality and processing improvements must have detailed quality assurance checks. The goal of a quality assurance check is to ensure that all required data are submitted as well as to control the accuracy of the data. There are two categories of quality assurance checks: immediate and delayed. An immediate quality assurance (IQA) check verifies that a required data element has been completed and will stop the user from moving forward in the system until the requirement has been met. An IQA check can be performed on a screen-by-screen basis (when the facility user clicks on the Save button) and also on a field-by-field basis (when the user moves to another data field). A delayed quality assurance (DQA) check verifies that a required data element or a required data condition has been completed, but is viewed at the <u>end</u> of the data entry process and will only stop the submittal of the inventory, until the requirement has been met. DQA checks can be displayed as running total reports, updated each time the Save button is pressed. This allows the user to check the DQA reports throughout the data entry process.

The main reason a well-designed application will include both types of QA checks is to create a more user-friendly process. An application could feasibly incorporate only one of the two QA checks, and still improve data integrity, quality and processing efficiency. However, unless you create something that facilities will want to use and will understand, then getting them

to use it correctly will be a lost cause. For example, a facility user will not be very happy if he/she goes through the data entry process and receives no IQA check warnings, then, upon reaching the end of the process, finds a long list of DQA checks that have to be satisfied before submittal. On the other hand, if the facility user is stopped from proceeding, via an IQA check, until required data are completed on a screen-by-screen basis, then when the user reaches the end of the process there will be few or no DQA checks that require completion before the prized inventory can be submitted. So, it is a fine line between: 1) what do you want to allow the facility user to leave that screen without doing and 2) what will they have to do at the end of the process to submit the inventory. An application that is well balanced will contain more IQA checks to help the facility user correctly enter data on a screen-by-screen basis. The DQA checks are reserved mainly for required data condition summaries, and also for a few required data elements.

Facility Interface

A basic description of the North Carolina DAQ's facility interface for the inventory application will help clarify future references. AERO is composed of eight main data screens: Emission Source, Emission Group, Control Device, Control System, Emission Release Point, Operating Scenario, Operating Scenario Summary, and Closeout. The facility user may create, modify or end-date emission sources, emission groups, control devices, control systems, and emission release points. The hub of the interface is the Operating Scenario screen. Operating Scenario is a term coined by DAQ, the fields on this screen correlate with EPA's emission, emission process, and emission period records. This is where all emissions-related data are entered, and where control systems and emission release points are associated for sources/groups. The Operating Scenario Summary screen is a view of all sources/groups and their associated operating scenarios. The Closeout screen contains the DQA checks and guidance for a successful inventory submittal.

IQA Checks - Improving Processing Efficiency

Ensuring that the inventory application simply captures all of the required data elements previously listed is easy enough. IQA checks can alert the facility user when required data are omitted and stop forward movement in the application, on a screen-by-screen basis, until required data are completed. Some of the data fields that are required in AERO include: emission source ID, SCC, operating schedule, annual throughput, control device ID, and emission release point ID. The IQA checks that are performed on each data screen will stop the facility user on that screen if one of the required fields is incomplete. The user must complete the field, identified by a pop-up message, before being allowed to proceed to the next data screen. For example, in AERO if the facility user fails to enter the emission source ID for a newly created emission source before saving the data record, then the IQA check will stop progress and warn that the emission source ID must be completed before saving the new emission source. The user must complete the field before being allowed to proceed to the next data screen. The complete submittal of all required and "nice to have" data elements means that

reviewers will spend less time checking completeness and more time reviewing and understanding data.

DQA Checks - Improving Processing Efficiency and Data Quality

As stated before, a well-balanced application will include only a few DQA checks for required data elements. Waiting until the end of the data entry process to inform the facility user that he has missing data in 100 different operating scenarios (OS) will probably ensure he will never attempt the web-base interface again. However, at the same time there may be instances where data elements that have been defined as "required" may not be readily known by the facility user at the time of data entry. Allowing the data record to be saved without all the required fields is okay as long as those required fields are included in a DQA check. If the user fails to go back and complete the missing data fields, then the DQA check will display the missing data elements on the Closeout screen and prevent inventory submittal until the required fields are complete.

In AERO, the check for "Operating Scenarios without Emission Release Points" is an example of a DQA check for a required data element. EPA says it is mandatory for every process record to have an emission release point record associated. On the OS page the user is required to enter emissions-related data, including annual throughput and units, operating schedule, seasonal throughputs, actual emissions, and affiliated control devices (CD) and emission release points (ERP). However, we have found that users often will not know all the ERP parameters when they start entering the emissions. AERO is designed to require facility users to enter all emissions related data before the OS record can be saved, while at the same time allowing the user to forego entering the ERP data at that time, and only require it before final submittal of the inventory. In the end this promotes improved processing efficiency for the facility user because the data entry process doesn't stall while waiting for ERP related information.

Another example from AERO is the DQA check for "Emission Sources (ES) With Incomplete Data Status." This check is a running total of all ES's that 1) do not have a completed OS, and 2) are not marked 'Not Operated' or 'Not Required to Report'. This is a wonderful example of a check for a required data condition. The facility user may inadvertently skip one or more emission sources (especially for those facilities that have hundreds of sources) but this check will catch those and display them for the user to correct, and will not allow submittal of data until all ES's are appropriately marked. Without this check the facility could submit an inventory with unaccounted emissions for several sources. The governing agency can process the inventory faster, spending more time reviewing the inventory for content than for completeness, thus improving efficiency and quality.

IQA Checks - Improving Data Quality

In order for the IQA check to be efficient in controlling the <u>quality</u> of the data entered, you must have a firm understanding of each data element and its parameters. Go back to the required and "nice to have" data element lists and create a new column. In this column record the parameters, if applicable, for each of the data fields. For example, for the field "Days (of operation) per Week" on the Operating Scenario screen, the parameters might be that the number entered for this field has to be between 1 and 7. However, the field, "Operating Scenario Description" on the Operating Scenario screen probably will not have parameters, because the facility can enter any number of different descriptions. IQA checks can verify that values entered in a specific data field fall within the set parameters. These checks alert the facility user when values entered fall outside the parameters and stop user progress, on a field-by-field basis, until a data value within the defined parameters is entered. In AERO, if the facility user enters the value 25 in the data field "Hours (of operation) per Day", the IQA check stops the user from continuing with a warning that the acceptable data values for this field are between 1 and 24. Once the user enters an acceptable data value the data entry process can continue. The goal is to provide the facility user with as much data guidance, via the IQA checks, as possible during data entry to ensure that he enters accurate data. This type of IQA check will improve the quality of submitted data and increase government processing efficiency.

IQA and DQA Checks - Improving Data Integrity

The cumulative effect of the IQA and DQA checks in the facility interface is increased data integrity. Without the QA checks the facility user could submit an inventory without all the required data. Data integrity is compromised when data submittals are incomplete. In addition, when incomplete inventories are received, governing agencies may insert default data in incomplete fields instead of spending the time to request the information from the facility. Along the same lines, when EPA receives state emissions data with gaps they may also fill them in with default data. At both of these points the integrity of the data has been further compromised because the soundness of the data is questioned. Furthermore, without the QA checks the facility user could enter inaccurate data. The QA checks do not prevent all mistakes and inaccuracies, but they do catch some of the more obvious and common ones. The more data fields that can be checked automatically against defined parameters, the more accurate (sound) the data submitted.

Table 1. Example IQA Checks - Facility Interface (AERO)

Screen Name in AERO	Description of QA Check	Parameters (if applicable)
Emission Source	Source ID for new sources	< 21 characters; No duplicates
Emission Source	Start Date for new sources	< 01/01/200X (X=current year)
Emission Source	Source Type for new sources	
Emission Source	Emission source description for new sources	
Emission Group	All sources have same source type for new groups	
Emission Group	Emission group description for new groups	
Control Device	Device ID for new devices	< 21 characters; No duplicates
Control Device	Device Category for new devices	
Control Device	Device Type for new devices	

Control System	Device ID's in new systems	No duplicate devices in system
Emission Release Point	ERP ID for new ERP's	< 21 characters; No duplicates
Emission Release Point	ERP description for new ERP's	
Emission Release Point	Height for new ERP's	
Emission Release Point	Diameter (if circular) for new ERP's	
Emission Release Point	Length (if rectangular) for new ERP's	
Emission Release Point	Width (if rectangular) for new ERP's	
Emission Release Point	Temperature for new ERP's	Between 50-100 degrees F (system allows higher/lower, with warning)
Emission Release Point	Velocity for new ERP's	Between 16.7-66.7fps (system allows higher/lower, with warning)
Emission Release Point	Volumetric Flow Rate for new ERP's	
Emission Release Point	Start date	< 01/01/200X (X=current year)
Emission Release Point	Lat/Long	Must be within appropriate county Lat/Long's
Operating Scenario	OS Description for new OS's	
Operating Scenario	Annual Throughput and Units for all OS's	> Zero
Operating Scenario	Seasonal Percentages for all OS's	Sum must = 100%
Operating Scenario	Operating Schedule for all OS's	Hours/Day – between 1-24 Days/Wk – between 1-7 Wks/Yr – between 1-52
Operating Scenario	Operation Start and End Dates	Between 01/01/200X and 12/31/200X (X = current inventory year)
Operating Scenario	Operation Start and End Times	Between 0000 and 2359

Table 2. Example DQA Checks - Facility Interface (AERO)

Display Screen in AERO	Description of QA Check
Closeout	Operating Scenarios Without Emission Release Points
Closeout	Operating Scenarios With Incomplete Data Status
Closeout	Emission Sources With Incomplete Data Status
Closeout	Operating Scenarios With < 100% Emissions Accounted for Through ERP
Closeout	Control Devices Without Category and Type
Closeout	Emission Source Without Source Type (first layer of SCC)

Agency Interface

The NC DAQ structured the facility interface and the agency interface (Emissions Data module, "ED") very similarly. Therefore, the basic description of the agency interface (ED) is the same as previously described for AERO. ED is the interface used to review inventories submitted via AERO as well as for DAQ staff to enter inventories received in paper format. In either case, the same QA checks apply in ED as in AERO. Without some final QA checks data integrity, quality, and processing efficiency can still be compromised.

Receiving complete, accurate and timely inventories from facilities is only half the goal of the inventory application. The other half is sending EPA complete, accurate and timely inventory data. Just as the DQA checks in the facility interface center around increasing facility

processing efficiency and ensuring completeness and accuracy for facility submittals, those same DQA checks in the agency interface will help improve agency-processing efficiency and ensure completeness and accuracy in the submittals to EPA. In addition to the QA checks already discussed in AERO, there are several QA checks specific to ED.

There is one required data element that DAQ determined to be too complex for many smaller facilities to accurately complete, and therefore decided to defer that requirement to ED. EPA requires a Source Classification Code (SCC) for every process record. There is an additional IQA check in ED to verify that every new OS record created has an SCC. The agency user cannot leave the screen until the SCC field is complete. However, this IQA check does not catch the OS records without SCC's that were created and submitted via AERO. Therefore, ED also has a DQA check for "OS's Without SCC's." This check must return a null value before approval of the inventory is possible, and an inventory must be approved before it can be sent to EPA. Should the inventory be approved without this information and the data submitted to EPA, it would be rejected as not meeting the standards. Incorporating these IQA and DQA checks ensures the completion of a required data field prior to approving the inventory. IQA and DQA checks will prevent erroneous submittals and the subsequent waste of resources for corrections and re-submittals to EPA.

ED also includes a few DQA checks to verify the completion of data fields critical to internal tracking and processing. For example, to ensure timely review and approval of inventories, the DAQ created the data field "Data Entry/Review Date Completed". The DAQ user must complete this field prior to inventory approval. The field is included in several reports and will identify inventories that have yet to be entered or reviewed. This DQA check prevents inventory approval until this data field is complete. Similarly, since new sources, devices, operating scenarios and emission release points may be submitted via AERO, DAQ felt it crucial to implement a way to ensure the review of these elements. The DAQ user must address each new element, listed above, prior to inventory approval. It may be that a new element was entered incorrectly or may even be a duplicate, in those cases the DAQ user has rights to delete, versus approve, that element. The DQA check prevents approval until all new elements are accepted or appropriately addressed.

Table 3. Example IQA Checks – Agency Interface (ED) (In addition to all IQA checks in AERO)

Screen Name in ED	Description of QA Check	Parameters (if applicable)
Operating Scenario	New OS's must have SCC	

Table 4. Example DQA Checks - Agency Interface (ED) (In addition to all DQA Checks in AERO)

Screen Name in ED	Description of QA Check
Facility Approval	Operating Scenarios Without SCC's
Facility Approval	Sources, Groups, Control Devices, Control Systems, and Emission Release Points Not Marked Accepted
Facility Approval	Approval Type selected
Facility Approval	Data Entry/Review Date Completed

Summary

Switching from the traditional paper inventory format to a web-based format provides for immediate gain in data processing efficiency. However, incorporating extensive QA checks throughout the web-based interface is the key to a three-fold gain: improved data quality, integrity, and processing efficiency. Increased data quality and integrity is essential if EPA is to use the data to protect and improve air quality in the United States. Obtaining complete, accurate, and valid data is necessary for risk assessments, development of State Implementation Plans (SIPs), identification of environmental problem areas, regional and national air modeling, pollution reduction verification, and trending for policy and regulation development, to name a few. The safeguards that quality assurance checks provide bring the level of data quality and integrity to an all-time high. However, the future is wide-open for more sophisticated applications, providing avenues for more intricate and encompassing quality assurance checks.