WRAP EDMS Database Administrator (DBA) Activities Documentation
August 22, 2006

1. *What steps are followed within Pechan and outside with Tom, when someone requests registration?*

EDMS is intended for the public. All users must register before they can access the data. They have to complete a brief registration form on-line and they have to provide a valid e-mail address. At the time of registration, users are asked to indicate if they would like to receive an EDMS-related newsletter via email. If a user wants to have rights to upload data or be a Data Analyst, then the EDMS DBA requests permission from Tom Moore. Once permission is granted, a Data Owner is provided specific rights to upload inventories into EDMS and a Data Analyst is provided with rights to query and export data. All these steps require the DBA to manually update the appropriate tables and records using scripts or an interactive database tool.

- Registration occurs within 24 hours on normal business days
- New registrants are notified via email once their account is activated
- Standard registrations give users “General User” status
- The registrants and their access level are tracked under the User Management section of EDMS. In this section of EDMS, a user’s status can be changed, passwords can be reset, and if user wants, their account can be deleted.

2. *What is the process of steps undertaken when data are uploaded by data owners/providers? Are these steps different for state/local agencies, tribes, WRAP contractors and if so how?*

The WRAP Steering Committee makes a preliminary decision on the quality assurance and gap-filling measures to be applied to each inventory before it is initiated in the EDMS. The DBA then manually sets up the procedures to be applied as the data are uploaded by the data providers. The data provider uploads a specific time period NIF-based inventory in a ZIP format and selects the inventory to which the data is to be applied in EDMS. Once the file is uploaded, a record is created of the upload in the data management section of EDMS. There are a series of steps that are performed before the inventory data gets into the production database:

- When a submittal from a data provider is processed, the data is added first to the Transfer Database. The DBA manually checks to ensure that the data has been applied to the proper inventory, which is based on year and the version. For example a data provider might submit calendar year 2002 emissions data to “Inventory 8: Actual Version 1.2.” The DBA checks that the data submitted is for the year 2002 and is applied correctly to this version of the 2002 inventory. If the submittal should fail at any point, the entire submittal run is rolled back and removed. If there is a discrepancy, an error file is generated for the database administrator to review and manually remove the data.

- In the second step, in transaction stage, certain potential irregularities in data format are addressed, such as trimming spaces and upper casing of certain fields. Quality Control routines diagnose duplication errors, referential integrity errors, valid range errors, null checks for required fields, and valid value errors. This report is returned to the data provider for review. The DBA controls the amount of quality assurance applied to each submittal by manually setting up the scripts to be run against the data.
- In the third step-- the Staging Database-- quality assurance routines, such as gap filling for missing sites, counties, SCCs or pollutants are performed, consistent with the version and typing restrictions enforced by the EDMS. The inventory is analyzed by the DBA to compare it to the anticipated submittals. Gap-filling methodologies and procedures are finalized with the WRAP Steering Committee. The database administrator and data analyst then manually develop and applied scripts for the QA/gap-filling to be applied.
- After QA routines and diagnoses of the staging database are complete, the data are manually moved to the production database by the DBA. No changes occur in the production database. The production version houses all locked emission inventories.
- State/local/tribal agencies and WRAP contractors can upload data only if they have been granted data owner access. The procedures for all data providers are the same regardless of jurisdiction. Any geographic/sector specific overlap must be resolved between the data providers that submit the overlapping data. The DBA can manually run data queries to determine data overlap while data are in the Transaction or Staging database.

The QA/QC routines are also explained in detail in the EDMS Quality Assurance Project Plan.

3. *Describe the completeness and reasonableness checks, the associated criteria for any recommended actions on questionable or duplicative data, and what other QA/QC checks are performed on all data types, by type of data.*

- Quality assurance routines such as gap filling for missing sites, counties, SCCs or pollutants are performed, consistent with the version and typing restrictions enforced by the EDMS. For example, if a submittal does not include a particular county for point sources, that county's information is forwarded from a previous year, or previous version of the inventory. If an area source submittal does not include certain pollutants for an SCC, those pollutants are brought forward from a previous version of that year's inventory.
- Remediation routines, such as stack augmentation procedures, or Particulate Matter (PM) augmentation procedures are also performed.

- Null non-mandatory data, such as seasonal percentages are filled in with defaults as required or normalized.
- Gap-filling procedures are subject to the restrictions and guidance of WRAP policy.
- Automated data validation QA/QC routines in the EDMS can either be turned on or turned off manually by the DBA. If the QA/QC procedures are turned on, the QA procedures run as a process following the import of data. To date, the QA/QC procedures have been turned off in the EDMS during data import to expedite the loading of the initial inventory data.

List of QA/QC data validation Checks:-

- Setting record types to the appropriate record type for the table.
- Resetting of submittal flags to RA/RD where the submittal flags were submitted as A/D incorrectly.
- Conversion of units provided to EPA standard units (such as changing TONS to TON).
- Recalculating stack information where derived data has been incorrectly calculated.
- Resetting stack information to standards where values provided is out of range.
- Converting UTM coordinate to latitude/longitude coordinates.
- Multiply X coordinates provided in latitude/longitude coordinates by -1 where provided as a positive value.
- Resetting coordinates where diagnosed as outside of county boundaries.
- Normalization of seasonal throughputs to add to 100%.
- Conversion of pollutants to standard criteria pollutants (NO2 to NOX, for example).
- If MAX NAMEPLATE CAPACITY negative, take absolute value.
- If FACTOR NUMERIC VALUE negative, take absolute value.
- If HEAT CONTENT negative, take absolute value.
- If EM RELIABILITY INDICATOR negative, take absolute value.
- If ASH CONTENT negative, take absolute value.
- If ACTUAL THROUGHPUT negative, take absolute value.
- If HORIZONTAL AREA FUGITIVE negative, take absolute value.
- If RELEASE HEIGHT FUGITIVE negative, take absolute value.
- If RELIABILITY INDICATOR negative, take absolute value.
- If STACK FENCELINE DISTANCE negative, take absolute value.
- If SULFUR CONTENT negative, take absolute value.
- If ACTUAL THROUGHPUT negative, take absolute value.
- Normalizing control efficiencies to percentages where they have been supplied as decimals.

4. *What steps are followed when a data owner/provider submits revised data?*

    - When a data owner submits revised data, it can be loaded in to EDMS directly by using the upload tool under Data Management.
    - If data is send directly to the DBA, then Pechan requests permission from Tom Moore before loading the data into EDMS.
    - The same steps are then followed as outlined in question 2, above.

5. *What are the steps taken to create each released emission inventory version? What internal documentation is prepared? I also wonder whether the Inventory Description Document (IDD) is enough detail for external users?*

    - After discussion with and approval by the Steering Committee, the DBA creates a new inventory section in the EDMS that will contain the submittals for that inventory type. An Inventory ID is created with a brief description of the inventory
    - Once the inventory has passed the QA/QC routines and is in the production database, the data providers review the inventory for completeness. At the completion of the data provider's review of the data, the WRAP Technical Coordinator grants permission to release it as a publicly accessible inventory.
    - Pechan is designing a new IDD will have detailed information, user notes, comments, references and data summaries. Pechan will be consulting with Tom Moore on this re-design. The new IDD will be developed for the version 2.0 release of the 2002 inventory.

6. *What are the steps followed by the DBA when data retrievals and summaries are requested?*

    - When a DBA receives a request for data retrievals, if the data is in the production database and available to the public, then the DBA extracts the data using the EDMS web interface. If the user needs specific data that is not available to the public or can not be obtained using the web interface, then Pechan sends a request to Tom Moore for permission to proceed. Once approved, the DBA extracts the data for the user using manual scripts written specifically for the request.

7. *What are the steps followed by the DBA when a data content and/or system function problem or error is reported?*

    - If there is a data discrepancy in an imported file, an error file is generated during the import process for the data administrator to analyze. Errors may also be found from analysis of inventory data by one or more users. Based on the error type and impact, different approaches may be developed and discussed with the WRAP Steering Committee before a solution is derived. While data is accepted into the EDMS "as is," the data errors are transmitted to the data providers. The data providers must then determine if they have the resources to rectify the data errors. A replacement file can always be submitted to rectify problems-- in other cases specific database scripts can be written to make changes. A record of all known data errors and fixes will be maintained and kept on-line as an addendum to the IDD.

- If there is a system function problem, then Pechan first manually logs the problem in trouble ticket software, Bugzilla and then investigates the problem, providing instructions to the development team to fix the problem. System functions and messages that are helpful in problem resolution are also manually logged into Bugzilla. The DBA communicates with the WRAP Technical Coordinator and any user that reports being affected by the problem and provides an estimated timeframe for a solution. Once the problem is repaired by the development team, it is installed and regression tested on the development site. The tests and their results are logged into Bugzilla. Once the test results deem the fix successful, the fix is moved to the production site. The WRAP Technical Coordinator and involved users are notified via e-mail that the fix has been made public.

8. *What logs are kept for each of the activities listed above?*

- The Logging/Notes tables track application activity. The Log and Log Detail tables track application activity associated with inventory QA/QC procedures. The Notes table tracks comments associated at multiple levels of the application, from the transfer table and inventory level, to notes on individual tables or records. While this information is stored in a single physical table, the presentation of the data is logically separated within EDMS.
- A separate log is kept and tracked on Bugzilla, which is web based software where every bug or enhancement is entered and tracked according to its status (new, resolved, fixed & reopened).

9. *What logs are kept for user contact and user support?*

Pechan provides web-based, e-mail-based, and telephone support for the EDMS through the duration of the contract and keeps a log of every customer support action taken. These logs are maintained by the DBA and summarized on a monthly basis and posted on the project website.

E-mail support actions are kept in a log, and the DBA has access to it. The web based requests also go to the DBA via e-mail thru the EDMS e-mail server. The support actions in e-mail formats are kept as a log and are maintained by the DBA. There is no log kept for the telephonic support provided to the users;

however, most telephone support calls are followed-up with an e-mail to the person with the support request.

10. *Do we get a report before gap filling?*

Any gap-filling performed is shown in the Inventory Description Document. AS mentioned earlier, this document will be more detailed in the future; however, the current version does indicate where inventories were gap-filled.

11. *Do we fill the tribal gap data?*

To date, no tribal data has been gap-filled. It is possible using the existing gap-fill procedures to gap-fill tribal data, if such data exists in the NIF format. It is our understanding that the states and tribes have initiated a process to coordinate future responsibilities within each state for data development that should eliminate most needs for gap filling.

The DBA's role in resolving tribal data overlaps is to analyze the imports and identify for possible data overlaps. Parties of the data overlaps are contacted as well as Tom Moore and informed of the situation. The data providers then need to come to some consensus on how data will reside in the EDMS. The DBA provides options and then carries out the solution arrived at via the data providers. This is a manual process that is tailored to each specific case.

12. *Distinguish between manual and automatic changes.*

At this point, the EDMS does not have any automatic changes for either the software or the data. All changes to either are performed manually.

13. *Do we give the data report to states for review?*

The data providers are provided a time window to review data before it is made available to the public. The data providers use the ad-hoc query capabilities of the system to assist them with their analysis. A specific report for states to review data is not generated from the EDMS.

14. *How many tribes will be updated for 2005?*

This question will be answered when we start accepting data for 2005. Perhaps ITEP can provide a response to this question.