Database Administrator (DBA) activities documentation—WRAP-EDMS—July 7, 2006 Pechan Draft

- 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, then the EDMS DBA requests permission from Tom Moore. Once permission is granted, the user is approved as a Data Owner with specific rights to upload inventories into EDMS.

- 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.

- 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?

Data owner/provider prepares the file in NIF format and uploads the inventory in a ZIP format by selecting the inventory to which the data needs to be applied. Once the file is uploaded, a record is created to check the status of the file under Data Management section of EDMS. There are 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. 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.

  - 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.

  - After QA routines and diagnoses of the staging database are complete, the data are moved to the production database. 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.

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

- 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 PM 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.

List of QA/QC 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.

- 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 asks Tom Moore for permission before loading the data into EDMS.
- QA/QC checks are applied manually, and if the data is approved by the Steering Committee, the DBA loads the data directly into the EDMS. A separate log is generated for the dataset.

- What are the steps taken to create each released emission inventory version? What internal documentation is prepared? I also wonder whether the IDD is enough detail for external users?

- The DBA designates a new inventory after discussion with the Steering Committee.
- Once the inventory has passed the QA/QC routines and is in production database, Pechan sends a request to Tom Moore for the permission to release it to public. An Inventory ID is created with brief description.
- Pechan is designing a new section in which the IDD will have detailed information, user notes, comments, references and data summaries. Pechan will consult with Tom Moore on this re-design. The sample IDD section will be developed for the version 2.0 release of the 2002 inventory.

- 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 public, then the DBA extracts the data using the EDMS web interface. If the user needs specific data that is not available to public or can not be obtained using the web interface, then Pechan sends a request to Tom Moore for permission. Once approved, the DBA extracts the data for the user.

- 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 discrepancy, an error file is generated for the data administrator to review and manually remove the data (or unlock the data depending upon the results of the review). Data in this set of tables is not available for users to view. Data can be submitted only to a current, unlocked inventory version.
- If there is a system function problem, then Pechan investigates the problem and provides instructions to the development team to fix the problem. Once the problem is fixed, it is first tested on the development site before posting on the production site.

- 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).

- 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.

  Email support actions are kept in a log, and the DBA has access to it. The web based requests also go to the DBA via email thru the EDMS email server.