

# EMISSIONS INVENTORY

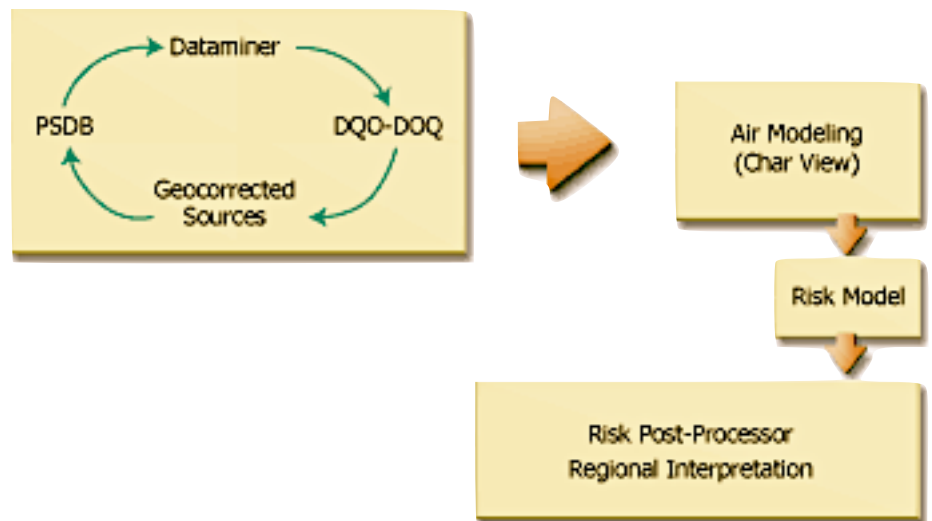
Lakes Environmental Software

EMISSIONS INVENTORY 2002



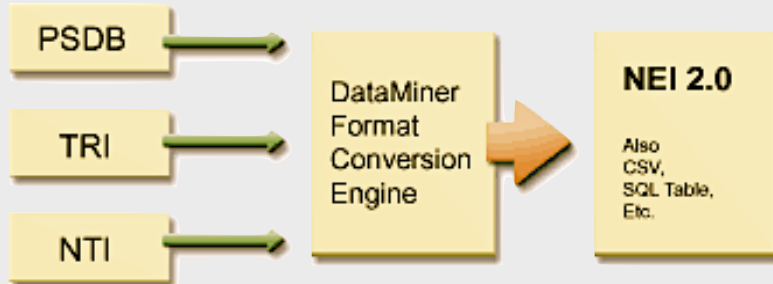
## PROJECT

Creating an emissions inventory is the first step in conducting a risk assessment. Our tools have evolved to correct source emissions reports and analyze data gaps. A complete description of these steps is presented to the right, employing the complex and large Texas Project Source Database inventory (PSDB).

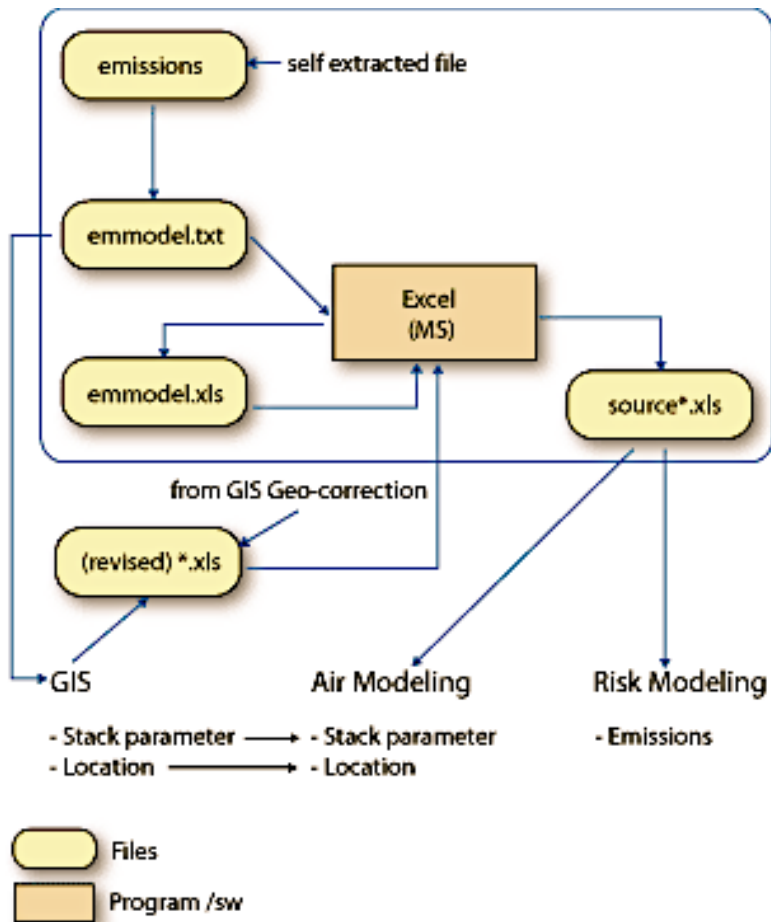


*Data quality steps within the risk assessment framework*

To accomplish our Data Quality Objectives (DQO) we needed powerful information technology tools. As an example, some of the complex steps required by emissions inventory data quality is presented in the Figure below.



**Converting Multiple Emissions Inventory Data to NIF 2.0**



**Sample steps of emissions inventory data quality assessment**

Existing solutions were tested and created enormous problems and delays in our projects. The major challenge was the explosive memory requirements for complex environmental emissions inventory cross-referencing and data mining. The following Figure to the left presents a sample cross-referencing using Data Miner developed by *Lakes Environmental*.

AC_LAST_24	AC_NEARBY	CE_ACCOUNT	CE_LONGTERM_ALLOW	CE_ACTUAL	CE_SIPSET_EMS
ATOPNAPETR COENRALS INC	PORT ARTHUR	J200504	0	12640	0
ATOPNAPETR COENRALS INC	PORT ARTHUR	J200501	0	9780	0
ATOPNAPETR COENRALS INC	PORT ARTHUR	J200501	0	9251	0
ATOPNAPETR COENRALS INC	PORT ARTHUR	J200501	0	8200	0
ATOPNAPETR COENRALS INC	PORT ARTHUR	J200504	0	12743	0
ATOPNAPETR COENRALS INC	PORT ARTHUR	J200501	0	2306	0
ATOPNAPETR COENRALS INC	PORT ARTHUR	J200501	0	4585	0
ATOPNAPETR COENRALS INC	PORT ARTHUR	J200501	0	2864	0
ATOPNAPETR COENRALS INC	PORT ARTHUR	J200501	0	17949	0
ATOPNAPETR COENRALS INC	PORT ARTHUR	J200501	0	2306	0
ATOPNAPETR COENRALS INC	PORT ARTHUR	J200504	0	11125	0
ATOPNAPETR COENRALS INC	PORT ARTHUR	J200501	0	2082	0
ATOPNAPETR COENRALS INC	PORT ARTHUR	J200504	0	2835	0
ATOPNAPETR COENRALS INC	PORT ARTHUR	J200501	0	2408	0
ATOPNAPETR COENRALS INC	PORT ARTHUR	J200504	0	14734	0
ATOPNAPETR COENRALS INC	PORT ARTHUR	J200501	0	2408	0

Search by: AC\_LAST\_24  
 Filter: (CE\_ACCOUNT = 'ATOPNAPETR COENRALS INC') AND (CE\_MAP = 1) AND (CE\_ACTUAL > 0) AND (CE\_ACTUAL < CE\_LONGTERM\_ALLOW)

## Cross Referencing Data into Views

Initial data mining tests on a large database, the 3Gbytes (text-only) Texas Point Source Database (PSDB), indicated the size of the problem. Due to existing SQL database solutions, such as Oracle and MS-SQL, searches on the PSDB database would create “views” over **1 Terabyte** in size. Regulators in Texas were using large Unix workstations to enable analysis of the data. Note that making reports out of database tables does not cause this problem. It is only evident when one tries to cross reference information, which is referred to as “Mining” the data.

Few system analysts are aware that the Views in SQL store data in a linear architecture, to facilitate cross-referencing. This way, even small queries are expanded from a master reference relational structure to a linear one. Computer resources were mysteriously vanishing with our evaluations. We tried in vain to add more RAM and hard disk.



After consultation with DBMS vendors, such as InterBase and Oracle, we learned that Data Mining software was being developed by independent parties to address the problem. Since we could not wait any longer for these systems we approached the problem in a unique way.

Prototype software was implemented, where users would create a Data Mining Expression using visual tools to link data and arguments. Subsequently, we would create our own SQL queries and parse the results into

a linked-tree structure. Once all the data was collected, we would cross-reference the data and re-store it in a new relational table structure. Using this approach, the response time to large and complex data mining expressions was fairly rapid. Memory requirements were never over the original database size.

The Figure below presents actual test results with the Texas Point Source Database (PSDB). Note that our solution implementation is referred to as the “Dataminer”.



*Memory requirements between Lakes Environmental's Dataminer and other advanced solutions*

**Our solution was so well received that it has been quickly adopted by various branches of the US EPA.**