**Air Toxics Reporting Guide**
**Ambient Air Toxics Monitoring Pilot Study**
**October 2001**

## Introduction

The following information has been gathered by the Information Transfer and Program Integration Division (ITPID). Contacts for clarification of this information are Bill Frietsche, at 919/541-5451, and Mark Schmidt of the Monitoring and Quality Assurance Group, at 919/541-2416.

### *1) New fields resulting from the Air Toxics Data Management Workgroup discussions*

The data management work group has identified two additional fields needed to support the air toxics monitoring data efforts. These fields are MDL and Uncertainty. MDL (or Method Detection Limit) will be a value, with assumed units equal to the units of the reported value to which it is associated with. Uncertainty also carries the same units as the reported value. Please refer to the Addendum to this guide for specific details on submission of detection limits and uncertainty.

### *2) Additional codes needed in AQS for air toxics monitoring*

The data management approach for reporting air toxics monitoring data will not differ significantly from the current reporting for criteria pollutants. Different pollutants will be reported: a revised list of these air toxic parameter codes is included as Table 1.

In addition to using the appropriate parameter codes, appropriate sampling and analysis method codes must also be used. The method code table for AQS may not include all of the needed method codes for the toxic pollutants being reported. Additional method codes may be needed if new sampling and/or analysis methods are being used for these toxic pollutants. Please let us know if you need new method code values added to AQS. You can find the current list of method codes in the AIRS GeoCommon Subsystem. The table is called AQ COLL ANAL METHOD and is menu item 04 (Browse).

The work group reached consensus that the values for species for TSP and PM10 will be reported at standard temperature and pressure since that is the normal reporting procedure, and that values for species for PM2.5 will be reported at local temperature and pressure since that is how that data is currently reported. For both cases, the temperature and pressure will also be reported so that conversion to local conditions is possible. Also note that standard time will be used for reporting of sample collection and laboratory times.

Wind speed and direction are also needed for modeling efforts. In some cases these meteorological data may come from two different locations (e.g., the wind speed and direction from the airport, the

temperature and barometric pressure from a nearby PM2.5 monitoring station or the toxics monitoring station itself). AIRS (legacy and new) does not accommodate referencing the source of meteorological data from two different locations.

Our solution in this case is to reference the temperature and barometric pressure meteorological station identification through the "met site type" parameter (which gives the data manager 6 choices for type of site.) Then, for the "wind speed and direction" data, the data manager will include the origin (site identification) in the comment field. Modelers will have to make the extra step to search the comment field for the meteorological data needed.

The parameter, method, and interval codes to be used for temperature and barometric pressure will be the same as are currently used for data reporting for species for PM2.5. Windspeed and wind direction will be reported using existing codes as well.

***3) New vs. Old transaction formats***

The current AQS transactions ("old transaction format") are acceptable to submit air toxics data to EPA for a limited time. However, we recommend that you register all of your sites and monitors in the new system, and also submit your monitoring data as soon as possible. Data will only be converted as part of the new system startup until November 2001. The MDL and Uncertainty data will then be added to the new system using the new transaction formats after the new system is placed in production.

Regarding the transaction formats, we (the EPA workgroup members) are assuming that everyone has access to and knowledge of the AIRS manuals for coding and submitting transactions to the mainframe. AQ2 (the data coding manual) defines exactly the transactions, data types, and field lengths used to update each type of record in the system. AQ3 explains how to submit transactions sets to the mainframe and have them updated into the AQS database. These manuals can be downloaded from the AIRS TTN world wide web site.
`http://www.epa.gov/ttn/airs/aqs/`

There is also a manual (GC2) containing all of the code values (like parameter codes, method codes, agency and state codes, etc.). You may prefer to browse the appropriate tables in GeoCommon Subsystem to be sure you are seeing the most up to date table values.

See Table 2 for the new AQS transaction formats.

## Table 1. Pilot Study Core Pollutants

**Pollutant Codes for AIRS Reporting**

***ACETALDEHYDE***

43503 ACETALDEHYDE AKA ACETIC ALDEHYDE

***BENZENE***

45201 BENZENE

***BERYLLIUM***

12105 BERYLLIUM (TSP)
82105 BERYLLIUM (PM10)
88105 BERYLLIUM PM2.5 LC

***BUTADIENE***

43218 1,3-BUTADIENE

***CADMIUM***

12110 CADMIUM (TSP)
82110 CADMIUM (PM10)
88110 CADMIUM PM2.5 LC

***CARBON TETRACHLORIDE***

43804 CARBON TETRACHLORIDE

***CHLOROFORM***

43803 CHLOROFORM

***CHROMIUM***

12112 CHROMIUM (TSP)
82112 CHROMIUM (PM10)
88112 CHROMIUM PM2.5 LC

***DICHLOROPROPANE***

43829 1,2-DICHLOROPROPANE AKA PROPYLENE

***FORMALDEHYDE***

43502 FORMALDEHYDE AKA-OXYMETHYLENE

***LEAD***

12128 LEAD (TSP)
82128 LEAD (PM10)
88128 LEAD PM2.5 LC

***MANGANESE***

12132 MANGANESE (TSP)
82132 MANGANESE (PM10)
88132 MANGANESE PM2.5 LC

***METHYLENE CHLORIDE***

43802 METHYLENE CHLORIDE AKA DICHLOROMET

***NICKEL***

12136 NICKEL (TSP)
82136 NICKEL (PM10)
88136 NICKEL PM2.5 LC

***TETRACHLOROETHYLENE***

43817 PERCHLOROETHYLENE AKA-TETRACHLOROE

***TRICHLOROETHYLENE***

43824 TRICHLOROETHYLENE

***VINYL CHLORIDE***

43860 VINYL CHLORIDE

# Data Input Formats for the Re-engineered Air Quality Subsystem

This document provides the new format to process batch transactions for the re-engineered Air Quality Subsystem. The first field of all input formats will be the format type. This will instruct the system what type of data is being sent to the database. Valid transaction types are listed below. The second field will be the type of affect the particular transaction will have on the database. These will either be an "Insert" (I), "Update" (U), or "Delete" (D). Please note that not all data is able to be provided through a batch process. Descriptions on how to provide data through an on-line mode will be made available in the coming months.

All transactions will be pipe delimited ("|"). A delimiter should not follow the final field in a record, so there will always be one less delimiter than fields for the given transaction type. It is essential that the proper number of delimiters be provided for a given transaction. So even if you need to only update one column on the database with "Monitor Basic" information, you still need to have 27 delimiters in that row. If you need to report a given field is null, place two delimiters back to back ("||"). Textual fields should not be enclosed with single or double quotation marks. The choice to use delimited versus positional is based on:

**a)** **Change in the monitor-id:** The monitor id will now consist of a 2-digit POC code. Instead of having to worry about adding a "0-padding" in front of or behind the current POC, you will be able to use the same id as now.

**b)** **Potential changes in record lengths:** There is always the potential of record lengths changing. We hope by using a delimited format, this will minimize the impact of any future changes in the structure of the data input formats or new entries in the associated reference tables. The choice of the pipe delimiter was to help ensure that the symbol would not be inadvertently used in free-format text fields.

Other Notation: R - Field is required for any action
R(n) - Field is Required for action 'n' (R(I) means required for an Insert Action for example)
X(n) - Conditionally Required for action 'n'

| Transaction Types: | | Number of Fields | Number of Delimiters |
|---|---|---|---|
| AA | Basic Site Information | 42 | 41 |
| AB | Site Street Information | 12 | 11 |
| AC | Site Open Path Information | 13 | 12 |
| MA | Basic Monitor Information | 28 | 27 |
| MB | Monitor Sampling Periods | 9 | 8 |
| MC | Monitor Type Information | 10 | 9 |
| MD | Monitor Agency Role | 11 | 10 |
| ME | Monitoring Objective Information | 11 | 10 |
| MF | Monitor Sampling Schedule | 22 | 21 |
| MG | Monitor Street Description | 9 | 8 |
| MH | Monitor Obstruction Information | 11 | 10 |
| MI | Monitor Regulatory Compliance | 10 | 9 |
| MJ | Monitor Collocation Period | 11 | 10 |
| MK | Monitor Protocol | 11 | 10 |
| RC | Composite Raw Data | 25 | 24 |
| RD | Hourly, Daily, Sub-Hourly Raw Data | 28 | 27 |
| RA | Accuracy Data | 32 | 31 |
| RP | Precision Data | 18 | 17 |
| RS | Annual Summary Data | 38 | 37 |

***Please Note***

The intent of the "Formatting Rules" described on the following pages are not intended to illustrate the complete validation procedures that a particular piece of data will be subject to. It is merely stating the format of the field or foreign key constraint as defined in the database. Future publications will explain all data validations in greater detail.

**R-2. Hourly, Daily, and Sub Hourly Raw Data (RD)**

| Field Name | Formatting Rule |
|---|---|
| Transaction Type [R] | Must exist within Reference Table (RD = Raw Data Type) |
| Action Code [R] | Must = I, U, or D |
| State Code [R] | Must exist within Reference Table |
| County Code [R] | Must exist within Reference Table |
| Site ID [R] | Must exist within SITE table |
| Parameter [R] | Must exist within Reference Table |
| POC [R] | Must exist within MONITOR table |
| Sample Duration [R(I,U)] | Must exist within Reference Table |
| Unit [R(I,U)] | Must exist within Reference Table |
| Method [R(I,U)] | Must exist within Reference Table |
| Date [R] | YYYYMMDD format |
| Start Time [R] | hh:mm format |
| Sample Value [X (I,U)] | Number - 5.5 format |
| Null Data Code [X (I,U)] | Must exist within Reference Table, if valued |
| Sampling Frequency | Must exist within Reference Table, if valued |
| MP_ID | Must exist within Monitor Protocols Table for the Monitor |
| Qualifier-1 | Must exist within Reference Table, if valued |
| Qualifier-2 | Must exist within Reference Table, if valued |
| Qualifier-3 | Must exist within Reference Table, if valued |
| Qualifier-4 | Must exist within Reference Table, if valued |
| Qualifier-5 | Must exist within Reference Table, if valued |
| Qualifier-6 | Must exist within Reference Table, if valued |
| Qualifier-7 | Must exist within Reference Table, if valued |
| Qualifier-8 | Must exist within Reference Table, if valued |
| Qualifier-9 | Must exist within Reference Table, if valued |
| Qualifier-10 | Must exist within Reference Table, if valued |
| Method Detectable Limit | Number – 5.5 format |
| Uncertainty | Number – 5.5 format |

**ADDENDUM -- AIRS Reporting of Detection Limit and Uncertainty Information for the Toxics Pilot Monitoring Study**

**10/1/01**

This document describes the mechanism for submitting detection limit and uncertainty information to AIRS for the toxics pilot monitoring study. General data reporting information for the toxics pilot is covered in the Air Toxics Reporting Guide.

## Background Notes

1. This addendum only addresses data entry *into the new AIRS system using the new input formats*. The new system and input formats will be put into production in late fall. Agencies can start data entry earlier using the legacy system and legacy formats, however, no guidance is provided for that process. Reporting organizations were given the option to delay their data entry until the new system is operational. If they exercise the option, EMAD will acquire available raw data in a prescribed alternative format (e.g., spreadsheet) in the interim. The new AIRS input file formats were provided as an attachment to the Air Toxics Reporting Guide; they can also be found at http://www.epa.gov/ttn/airs/aqs/reeng/index.html.

1. Partly in anticipation of the emergence of toxics data reporting, two new AIRS fields were created: raw data MDL and raw data Uncertainty. Ongoing discussions have arrived at the conclusion that these fields are *currently not necessary* for the pilot study. Both fields, however, are still being implemented in the new AIRS since they are needed in the short-term for other programs (e.g., IMPROVE) and *will eventually* (longer term) be utilized on a national scope by programs such as toxics and particulate speciation. States may wish to utilize these fields sooner for their own uses. Detection limits will still be recorded in the AIRS database for the toxics study but via a different technique (i.e., '*monitor* alternate MDL's). Although raw data 'uncertainty' information has potential value for data analyses, the specific field definition has not yet been decided. Once it is agreed upon what should be in the uncertainty field (precision, bias, or some combination) and sufficient precision and bias data have been collected (likely 12 months plus), then hopefully, reporting organizations can calculate and populate the field. AIRS may eventually be able to automatically fill this field based on other entered information. For now, data users can calculate uncertainty however desired, by using the raw data in the precision and accuracy transaction records. (See precision and accuracy reporting details below.)

## Detection Limits

All data for the pilot study will be reported without 'screening' or 'censoring' the data below detection

or reporting limits. That is, the actual quantifiable data value will be reported for cases below detection limits, *not '0', not half of detection limit, nor non-reporting (missing)*. Detection limits will be computed according to the guidance provided in the Pilot City Air Toxics Measurement Summary (http://www.epa.gov/ttn/amtic/files/ambient/airtox/toxics2a.pdf). That memo describes several types of detection limits but prioritizes the use of 'Method Detection Limits' (MDL's). MDL's will be reported at the *monitor* level in AIRS; see details below. In addition to reporting MDL values, toxics pilot agencies will also flag all raw data that are below the 'lowest calibration level' (LCL; also referred to as the 'minimum reporting level'). For this program, the LCL is determined to be 3 times the MDL. Data below the LCL will be flagged with a '7' qualifier.

There are 3 levels of Method Detection Limit (MDL) available in the new AIRS:

1. The Monitor x Method x Sample Duration x Units x Date-time or '*Raw Data*' level. Raw data MDL's can be loaded to AIRS via RD (raw data) transactions.
2. The Monitor x Method x Sample Duration x Units or '*Monitor*' level. Monitor level MDLs can be added to AIRS either interactively or via MK (monitor minimum detectable) transactions.
3. The Parameter x Method x Sample Duration x Units or '*Federal*' level. Federal level MDLs can only be added to AIRS (or modified) by the AIRS Group in ITPID. [An Air Toxics Reporting Guide attachment showed the Federal level MDL's for the core target Parameter x Method x Sample Duration x Units combinations. Federal level MDL's for other combinations can be viewed with an AIRS browse. AIRS records should already exist for every Parameter x Method x Sample Duration x Units combination used in the toxics pilot. If a new or different combination is being used or considered, the agency should contact Mark Schmidt (919-541-2416) or Joann Rice (919-541-3372) for assistance in getting the code combination and corresponding Federal MDL added to AIRS .]

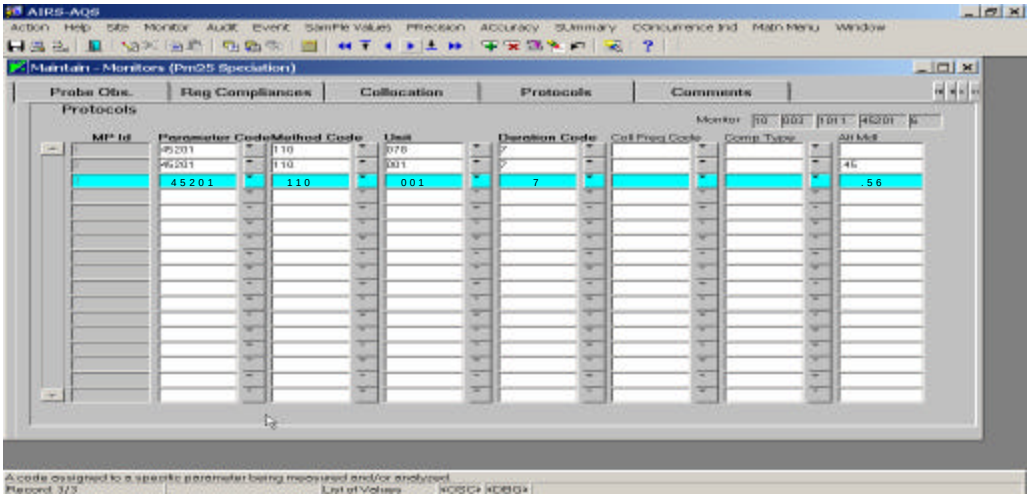
The MDL associated with specific raw data will be determined by the above hierarchy: If a raw data MDL is present, it will be the applicable MDL. If a Raw Data MDL is not present but a Monitor one is present and the raw data are linked to that specific limit (more details are provided below on how the link is specified with an Alternate MDL ID), then that is the applicable MDL. If a Raw Data MDL is not present and the raw data are not 'linked' to an Alternate MDL, then the Federal MDL is the default. The toxics pilot will utilize Monitor MDL's; reliance on only the Federal MDL is not an acceptable option. [Note: States can use Raw Data MDL's in lieu of Monitor MDL's but that process is not described here.]

Pilot study labs will determine MDL's prior to sampling and then update them a minimum of once a year. For AIRS entry, agencies will enter the original detection limits at the 'Monitor' level and then update when appropriate. If the user opts to add the Monitor MDL's to the system interactively, they will enter them on the Maintain - Monitors - Protocol panel. All existing 'Monitor Protocol' records (valid combinations of Monitor x Method x Sample Duration x Units x Alternate MDL) are shown on this panel. Monitor Protocol (MP) ID's, which are unique integers used to identify these combinations, are also shown. Users will enter the new Alt MDL value on the first blank row below the existing MP entries. Once the user cursors to the blank row, it will become highlighted and the new MP ID will

appear. After the user enters the new Alt MDL (and the applicable Method Code, Unit code, Duration code, and Parameter code) they should Save the transaction. If the user opts to add the Monitor MDL to the system with an MK batch transaction, they will specify the MP ID in the Alternate MDL ID field. [Note: MP ID and Alternate MDL ID are one and the same.]. Before creating the MK transaction, the user should peruse the Maintain - Monitors - Protocol panel to view existing protocol records and determine what MP ID to assign. Both methods of entry (interactive or MK batch transaction) will achieve the same result. The raw data reported (with RD transactions) during the period where the MDL is applicable will have to be reported with the corresponding MP ID; the MP ID will be entered in the Alternate MDL ID field of the RD transaction. If MDL's are updated once a year, then the same value will appear in the Alternate MDL ID field (for that lab-parameter) the entire year. If the ID field is left blank, the raw data will be associated with the Federal MDL! Some batch coding and interactive entry examples are shown below.

Example 1 - Interactive Entry of Monitor MDL

- Here is a Maintain - Monitors - Protocol panel.




- User enters Parameter Code (45201), Method Code (110), Unit Code (001), Duration Code (7), and Alt Mdl value (.56) on the first blank row and carries insert with a Save (F8). In subsequent raw data transactions, the associated MP ID (3) must be specified in the ALT MDL ID field.

Example 2 ~ Batch Insert of Monitor MDL:

*This accomplishes the same result as Example 1.*

- Insert the monitor MDL with an MK transaction:

```
MK|I|10|003|1011|45201|6|3|7|001|110|||.56
                           ^              ^
                  Alternate MDL ID   Minimum Detect Value
```

Example 3 ~ Submission of raw data with link to Monitor MDL

- Insert the raw data record with RD transaction. The MP ID (Alternate MDL ID) links the raw data to the proper Monitor MDL.

```
                                                                  Qualifier
                                                                      ^
RD|I|10|003|1011|45201|6|7|001|110|20000101|00:00|1.35|||3|7|||||||||||
                                                        ^       ^
                                                Concentration  Alternate MDL ID
```

- Note that because the concentration is below the LCL [1.35 < (3 * .56)], a '7' qualifier was assigned.

## Precision Information

Two indicators of precision are 1) collocated samples (duplicates) which measures the precision of the entire system and 2) replicate analyses of an individual sample in the laboratory which measures the laboratory component of precision. For the pilot program, it was agreed that replicate laboratory analyses will be done on both the primary and duplicate samples. Thus, for a particular monitor-day, 4 concentration values will be generated: the primary sampler base value (which is reported in the raw data record), the collocated (duplicate) sampler base value, the primary sampler replicate analysis value, and the duplicate sampler replicate analysis value. Agencies will submit the latter 3 (called 'precision information') via the RP (precision) transactions. [Agencies will not have to create monitor records nor submit 'raw data' transactions for the collocated samplers; collocated data only have to be submitted in the RP transactions.] 'Precision ID' will differentiate the duplicates, replicates, and duplicate-replicates as follows:

- Use Precision ID = '1' for the duplicate
- Use Precision ID = '2' for the replicate
- Use Precision ID = '3' for the duplicate-replicate

[Note: In the legacy AIRS / legacy input formats, this would correspond to coding (w/ Transaction '9') the duplicate as 'Collocated 1', the replicate as 'Collocated 2', and the duplicate-replicate as 'Collocated 3'. ]

The duplicate, replicate, and duplicate-replicate values will be reported in the 'Indicated Value' field of the RP transactions (on 3 separate rows) and the corresponding method code in the 'Indicated Method' field. The primary sampler values, which also are submitted with RD transactions, will be repeated in the 'Test Value' field (on all 3 rows) along with the corresponding method code ('Test Method').

Example ~ Adding duplicate, replicate, and duplicate-replicate information

C Duplicate information reported with Precision ID = 1

```
RP|I|10|003|1011|45201|6|1|7|001|110|20000101|1.35|110|1.25|||
                          ^       ^            ^    ^    ^
                  Precision ID  Prim. Meth.  Prim. Value  Dup. Meth (l) & Value(r)
```

C Replicate information reported with Precision ID = 2

```
RP|I|10|003|1011|45201|6|2|7|001|110|20000101|1.35|110|1.36|||
                          ^       ^            ^    ^    ^
                  Precision ID  Prim. Meth.  Prim. Value  Rep. Meth (l) & Value (r)
```

C Duplicate-replicate information reported with Precision ID = 3

```
RP|I|10|003|1011|45201|6|3|7|001|110|20000101|1.35|110|1.33|||
                          ^       ^            ^    ^    ^
                  Precision ID  Prim. Meth.  Prim. Value  Dup-Rep. Meth (l) & Value (r)
```

**Accuracy / Bias Information**

An indication of bias (for the VOC's) is the use of audit samples where a sample of known concentration is analyzed and the measurement value obtained by the laboratory is used to estimate bias. Since we don=t have a national audit program for toxics at this time (one is currently in development), we will have to rely on laboratory 'round-robin' inter-comparison samples to estimate bias across the group. [In general, the 'round-robin' plan calls for multiple canisters pulled from the same base can to be sent to different labs, analyzed there, and the results centrally compiled.] The tentative plan is to compute an average of the individual round-robin results and use this average as an indicator for 'truth' or the Actual Value. The individual round-robin results will be the Indicated Values. After the round-robins, OAQPS will compute the parameter averages, create RA (accuracy) transaction files, and forward the agency-specific portion of those files to the agency for upload. OAQPS could just disseminate the averages to the agencies, but the extra step (creation of the transactions) will help facilitate the process. For the metals, flow rate checks will provide an indication of bias. Results of the flow rate checks will also be submitted with RA (accuracy) transactions.

**Need Help?**

If you have detection limit, precision, or accuracy data reporting questions, please contact Mark Schmidt (919-541-2416), Jake Summers (919-541-5695), or Bill Frietsche (919-541-5451)