# SECTION 5
# DATA MANAGEMENT

## 5.0 INTRODUCTION

After air toxics data have been collected using the required sampling and analytical techniques described in earlier sections, the management of this information is key to the success of the NATTS mission. The integrity of the data collected and compiled in an acceptable management system is critical, as the data will be used to address human exposure to air toxics. Three elements concerning data management are discussed:

- data validation and reporting;
- data archiving; and
- data preparation for entry into EPA's AQS[1] data base.

## 5.1 DATA VALIDATION

The purpose of data validation is to detect and then verify or remove any data values that may not represent actual air quality conditions at the sampling station.[3] Validation of data is a key component of ensuring data quality. To help ensure data consistency, the techniques for validating data should be the same across all sites in the NATTS Program. In general, the data collected according to the specifications of advocated methodology are not automatically considered valid. To be validated, the data must be reviewed to confirm that sampling, analysis, and OA/QC were performed according to the specifications as presented in this TAD. Use of data validation techniques greatly reduces the risk of inconsistent/unacceptable data entering the EPA AQS data management system. Examples of validating field data include checks of monitoring equipment flow rates, sampling times, sample storage conditions, and hold times. If data have potentially been biased by "catastrophic releases" (such as a gasoline spill nearby), those data may be invalidated from the data set as they may artificially affect the data assessment and trend recognition. If a NATTS site is located in close proximity to a chemical manufacturing plant and a benzene leak occurs at the plant on a sampling day, the sample may show a very high ambient concentration of benzene. The resulting monitoring data should be flagged for later evaluation during the review process. The site logbook should also be reviewed for any unusual circumstances recorded during a sample collection period.

Laboratory data must be validated to confirm that the QC requirements for blanks, calibration curves, and regular calibration checks meet the method requirements as presented in this TAD.

The objectives for data validation should include the following:

- To produce a data base with values that are of known quality;
- To evaluate the internal, spatial, temporal, and physical consistency of the data and accept, correct, flag, and/or invalidate as appropriate; and
- To intercompare data to identify errors, biases, or outliers and accept, flag, correct, and/or invalidate as appropriate.

Typical elements of data validation include the following:

- Sample receipt. As soon as samples are received at the laboratory from the field, the COC documentation is checked to verify the sample identity and to invalidate samples with sampling anomalies;
- Sample analysis. The analyst preparing the sample verifies that all samples are prepared following individual method specifications;
- Post-Sample Analysis. After sample analysis, a reviewer verifies that all samples were prepared and analyzed within method-specified hold times;
- Preliminary data review. The instrumental analyst reviews the compound ID information obtained by the analytical instrumentation (GC/MS, HPLC, IC, ICP/MS);
- Second level review. A second reviewer verifies the analyst's determinations and prepares a quarterly report; and

- Third level review. The third reviewer reviews the quarterly report for a sample set, from sampling to analytical detection and quantitative analysis and final report.

Data validation should include the use of statistical analysis to determine invalid data. All statistical terms used in this section can be found in *Introduction to Probability and Statistics*[4]. As data are collected over a period of time, the statistics derived below can be compared against the historical data set. Preparation of a scatterplot and/or boxplot of NATTS Program site data is an effective way to visually determine potential outliers from the main body of data. Potential outlier data should be rigorously reviewed to determine whether contamination or operational errors occurred, which would invalidate the data.

Another statistical procedure for data evaluation includes determining the central tendency of the data set. There are four different ways to describe this central tendency:

- Arithmetic mean. The sum of the measured concentrations divided by the number of samples;
- Geometric mean. The result of multiplying the concentrations of samples with each other and taking the $n^{th}$ root of the number (n) of samples (e.g., for a data set with 20 concentrations, the $20^{th}$ root of the product would be taken);
- Median. The concentration value that represents the midpoint of the data set when arranged in order of magnitude (e.g., 50% of the data is greater than the median and 50% of the data is less than the median); and
- Mode. The concentration that has the highest frequency.

Data analysts calculate these values to identify outliers for data validation. It is very important to proceed from the "big picture" to a closer view, proceeding from a month of data to a week and then to a day. This strategy is important in forming an overall understanding of the data. Another important factor is to inspect every species that is reported, even when low concentrations are expected. Data validation is critical because serious errors in data analysis, modeling results, and trends analysis can be caused by erroneous individual data values.

Sonoma Technology, Inc. (STI), has developed VOC data validation and analysis software named VOCDat.[5] The most current issuance is Version 2.61. STI provides this software at no cost to local, state, and regional agencies throughout the United States for use in preparing their VOC data for submittal to the U.S. EPA's data repository, AQS. VOCDat enables an analyst to screen VOC data for outliers and display data using time series, scatter, and fingerprint plots. [VOCDat software and its user guide can be downloaded free from ftp://ftp.sonomatech.com/public/vocdat/ .] VOCDat also displays other air quality parameters such as toxic compounds, $O_3$, $NO_x$, and meteorological measurements.

The three types of plots useful in data evaluation are discussed in more detail below.

— Time series plots. To take full advantage of time series plots, the time series of every species and species group should be plotted and inspected to identify outliers, calibration spikes, abrupt changes in concentrations, possible misidentification of peaks, and extended periods of unusually high or low concentrations. It is useful to plot species together which are primarily emitted by the same type of source (e.g., benzene and acetylene are both present in automobile exhaust), or to plot species together which are emitted from different sources.

— Scatter plots. In preparing scatter plots, several pairs of species or species groups such as benzene and toluene, benzene and acetylene, benzene and ethane, and other pairs should be plotted and inspected. Scatter plots are useful for comparing the relationship between species at one site or at a pair of sites.

— Fingerprint plots. Fingerprint plots show the concentration of each species in a sample (in chromatographic order) and help to identify unique characteristics of the samples. Fingerprint plots should be inspected quickly and fingerprints of samples that have been flagged (i.e., identified as suspect or invalid) should be inspected in time series or scatter plot analyses. Checking fingerprint plots one-by-one allows an analyst to observe diurnal changes in species or species groups quickly. The analyst should then inspect hours or days surrounding suspect and invalid data to see if there is any carryover effect.

Final data validation will be performed by EPA in the process of calculating trends in NATTS compounds concentrations. For each NATTS Program site, and on a national basis, final data validation techniques include calculating and assessing the following statistical parameters:

— The variance (or dispersion). The average of the square of the deviations of the measurements about their mean;

— The standard deviation. Equal to the positive square root of the variance; and

— The confidence interval. Uses the SD and size of the sample population, along with a t-value, to determine the statistical range in which the arithmetic mean may reside under a *normal* distribution.

Individual confidence intervals and coefficients of variation will be compared to the DQO coefficient of variation (i.e., ± 15%) as final validation for use of the data in trends analysis.

AQS reporting provides a regular summary of the results and observations made during NATTS Program monitoring. The data must be submitted to EPA's AQS data base quarterly.

## 5.2 DATA ARCHIVING

Data archiving is the backing up and storage of data that must be retained but not regularly accessed.[6] After data have been validated and reported to AQS, all records used to generate the data must be archived by the participating agency in a manner that is easily accessible and retrievable. These archived records should be stored for a period of no less than six years in a separate physical location from the laboratory or field site to minimize the potential for data loss.

## 5.3 EPA'S AIR QUALITY SYSTEM

EPA's AQS is a computer-based system for handling storage and retrieval of information pertaining to airborne pollutants. AQS is administered by the EPA, OAQPS, in Research Triangle Park, NC. AQS contains data from state and local agencies, tribes, and federal organizations, including descriptions of air monitoring sites and monitoring equipment, measured concentrations of air pollutants and related parameters, and calculated summary and statistical information. Reporting agencies submit air quality data as formatted transactions using EPA Central Data Exchange (CDX).

Twenty types of transactions are used to provide data and control information for updating the AQS data base, with detailed instructions available for coding individual transactions.[7] Four general types of values are used to code air quality transactions: codes, dates, numeric data, and alphanumeric data. Each of these values must be entered on transactions exactly as they are stored in the AQS tables. The 20 AQS transaction formats contain certain fields in common, as well as unique fields:

- The transaction type specifies which batch transaction is being processed by the batch load software and determines which tables and columns will be updated with the data in the delimited fields;
- The action code indicates the data manipulation action to be performed by the transaction;
- The state code identifies one of the 50 states, U.S. territories, Washington, DC, or foreign countries;
- The county code identifies a county or equivalent geopolitical entity such as parish, independent city, or Tribal entity. For foreign countries, the county code identifies the geopolitical equivalent to U.S. states, such as Mexican states or Canadian provinces;
- The site ID is a numeric code that uniquely identifies each air monitoring site within a county. Site numbers are not assigned continuously or in any particular

order. Local organizations are free to allocate site numbers in any way they choose as long as there is no duplication within a county;

- A set of three site transactions is used to update site information:
  - Type AA (Basic Site Information)
  - Type AB (Site Street Information)
  - Type AC (Site Open Path Information).

- A set of 11 transactions is used to update monitor information in the site file:
  - Type MA (Basic Monitor Information)
  - Type MB (Monitor Sampling Periods)
  - Type MC (Monitor Type Information)
  - Type MD (Monitor Agency Role)
  - Type ME (Monitoring Objective Information)
  - Type MF (Monitor Sampling Schedule)
  - Type MG (Monitor Street Description)
  - Type MH (Monitor Obstruction Information)
  - Type MI (Monitor Regulatory Compliance)
  - Type MJ (Monitor Collocation Period)
  - Type MK (Monitor Protocol).

- Raw Data Transactions
  - Type RC (Composite Raw Data)
  - Type RD (Hourly, Daily, and Subhourly Raw Data).
  - Type RB (Field or Trip Blank Data).

- Accuracy/Precision Transactions
  - Type RA (Accuracy Data)
  - Type RP (Precision Data)

- Annual Summary Data (Transaction Type RS).

### 5.3.1 *Air Toxics Flagging and Reporting for EPA's AQS Database*

The guidance presented in this section explains and outlines how to report and flag air toxics data collected by Regional, State, Local and Tribal agencies that report their data to EPA's AQS database.

### 5.3.1.1 Clarification of Terminology

There are a plethora of scientific terms and acronyms that are used for defining the lowest level that can be detected a given piece of instrumentation. Some (*not all*) of the terms and acronyms that are related to the quantification and detection of instrument sensitivity and reporting of data are presented below. After careful review of all terms available, it was decided to simplify the schema. It is realized that some laboratory and data analysts will wonder why terms they are most comfortable with are not included. The explanation given here is that, in most instances, these terms define the same concept, but have different monikers. Therefore, terms that are utilized most often are included in this document. Generally, two distinct classes of terms exist: quantitation limits and detection limits. The definitions associated with these two classes are presented below:

**Quantitation Limits (QL)—**The lowest level at which the entire analytical system must give a recognizable signal and acceptable calibration point for the analyte. It is equivalent to the concentration of the lowest calibration standard, assuming that all method-specific sample weights, volumes, and cleanup procedures have been employed.

**Detection Limits (DL) —**Minimum concentration of an analyte that can be measured above instrument background. The DL is an estimate of concentrations at which one can be fairly certain that the compound is present. Concentrations below this limit may not be detected. Concentrations above this limit are almost certainly detected.

**Sample Quantitation Limit (SQL) —**Also known as a Practical Quantitation Limit (PQL), SQL is the lowest concentration of an analyte that can be reliably measured within specified limits of precision and accuracy during routine laboratory operating conditions. The Agency has used the SQL to estimate or evaluate the minimum concentration at which most laboratories can be expected to reliably measure a specific chemical contaminant during day-to-day analyses of drinking water samples. Normally, the SQL is determined as a multiplier of the method detection limit (eg., 3.18 times) and is considered the lowest concentration that can be accurately measured, as opposed to just detected. *Report the values at or between the SQL and MDL using the "SQ" QA flag.*

**Method Detection Limit (MDL)—**EPA definition: the minimum concentration of a substance that can be measured and reported with 99% confidence that the analyte concentration is greater than zero and is determined from analysis of a sample in a given matrix containing the analyte (Part 136, App. B). Determined by taking a minimum of seven aliquots of the sample (in case of air sample analysis we are using individual canister samples) to be used to calculate the method detection limit and process each through the entire analytical method. Make all computations according to the defined method with the final results in the method reporting units ($\mu g/m^3$). The EPA Compendium Methods state that MDLs must be determined and reported. **EPA policy dictates that all data, *to include values <u>below</u> MDL,* shall be reported to AQS.** *Data values at or below the MDL must be flagged with an "MD" flag. <u>Do not report 1/2 MDL or any integer of the MDL, only report the actual measured value and the "MD" QA flag.</u>*

**Non-Detected Compounds—**If a reporting agency cannot reliably detect a compound, then report 0 and the "ND" flag.

A summary of flags associated with quantitation and detection limits is presented in Table 5.3-1.

**Table 5.3-1. Summary of QL/DL Flags and their Application**

| If Concentration is | Value to Report | Flag Applied |
|---|---|---|
| > SQL | Value | None |
| > MDL ≤ SQL | Value | SQ |
| ≤ MDL | Value | MD |
| Not Detected | 0 | ND |

#### 5.3.1.2 Data Flagging

There are clear and established situations when flags should be applied to ambient air toxics data for the NATTS program. The flags fall into four categories: quantification and detection flags, laboratory flags, chain-of-custody flags, and field maintenance and operation flags. There are also flags that must be used when an agency submits collocated or duplicate data for the calculation of precision. Most of these flags currently exist in EPA's AQS database. Those flags that are new to AQS, and detailed in this document, are scripted in **bold.** (Note: AQS can accommodate up to 10 flags per record.)

#### 5.3.1.3 Types of Flags and Hierarchy

For air toxics data submittals there are two qualifier flag types: Null and QA.

- **Null** flag**—**This type of flag should be assigned when a scheduled sample is not usable (e.g., canister leaked, canister damaged in shipment, etc.).
- **QA Qualifier** flag**—**These flags are used to denote there was a procedural or quality assurance issue that happened that could possibly affect the uncertainty or concentration of the value. (Note: the flags for the quantitation and detection limits are QA qualified flags.)

AQS flags can be used in combination if they are non-Null flags. If a Null flag is used, no other flag is necessary, nor should another flag be used since no data are reported.

Among the new flags are blank issue identifier flags (as noted in Table 5.3-3). These should be used if reported blank values are above those set by the individual laboratories SOPs or QAPP. If high blank values are associated with samples, it is important that the values be reported but appropriately flagged. Do not invalidate values due to high blank values.

**Table 5.3-2. Quantitation and Detection Flags**

| Qualifier Type | Qualifier Type Desc | Qualifier Code | Qualifier Description |
|---|---|---|---|
| QA | Quality Assurance Qualifier | **SQ** | Values between SQL and MDL |
| QA | Quality Assurance Qualifier | **MD** | Values less than MDL |
| QA | Quality Assurance Qualifier | **ND** | No value detected |

**Table 5.3-3. Laboratory Generated Flags**

| Qualifier Type | Qualifier Type Desc | Qualifier Code | Qualifier Description |
|---|---|---|---|
| NULL | Null Data Qualifier | AR | General lab error |
| NULL | Null Data Qualifier | AS | Poor quality assurance results |
| NULL | Null Data Qualifier | BH | Interference / co-elution |
| QA | Quality Assurance Qualifier | **FB** | Field blank value above acceptable limit |
| QA | Quality Assurance Qualifier | **TB** | Trip blank value above acceptable limit |
| QA | Quality Assurance Qualifier | **LB** | Lab blank value above acceptable limit |
| QA | Quality Assurance Qualifier | **LJ** | Analyte identified; reported value estimated |
| QA | Quality Assurance Qualifier | **LK** | Analyte identified; reported value may be biased high |
| QA | Quality Assurance Qualifier | **LL** | Analyte identified; reported value may be biased low |
| QA | Quality Assurance Qualifier | **EH** | Estimated; exceeds upper range |
| QA | Quality Assurance Qualifier | **CC** | Clean canister residue |
| QA | Quality Assurance Qualifier | 7 | Below lowest calibration level |

**Table 5.3-4. Chain-of-Custody Flags**

| Qualifier Type | Qualifier Type Desc | Qualifier Code | Qualifier Desc |
|---|---|---|---|
| NULL | Null Data Qualifier | **MC** | Module end cap missing |
| NULL | Null Data Qualifier | **TS** | Holding time or transport temperature out of spec |
| NULL | Null Data Qualifier | AF | Scheduled but not collected |
| NULL | Null Data Qualifier | AG | Sample time out of limits |
| NULL | Null Data Qualifier | AJ | Filter damage |
| NULL | Null Data Qualifier | AK | Filter or sample leak |
| NULL | Null Data Qualifier | AL | Voided by operator |
| NULL | Null Data Qualifier | AM | Miscellaneous void |
| NULL | Null Data Qualifier | AQ | Collection error |
| NULL | Null Data Qualifier | **FI** | Filter Inspection flag |

**Table 5.3-5. Field Operations and Maintenance Flags**

| Qualifier Type | Qualifier Type Desc | Qualifier Code | Qualifier Desc |
|---|---|---|---|
| NULL | Null Data Qualifier | AA | Sample pressure out of limits |
| NULL | Null Data Qualifier | AB | Technician unavailable |
| NULL | Null Data Qualifier | AC | Construction repairs in the area |
| NULL | Null Data Qualifier | AD | Shelter storm damage |
| NULL | Null Data Qualifier | AE | Shelter temperature out of specification |
| NULL | Null Data Qualifier | AH | Sample flow rate out of limits |
| NULL | Null Data Qualifier | AI | Insufficient data to make calculation |
| NULL | Null Data Qualifier | AN | Machine malfunction |
| NULL | Null Data Qualifier | AO | Bad weather |
| NULL | Null Data Qualifier | AP | Vandalism |
| NULL | Null Data Qualifier | AT | Calibration |
| NULL | Null Data Qualifier | AU | Monitoring waived |
| NULL | Null Data Qualifier | AV | Power failure |
| NULL | Null Data Qualifier | AW | Wildfire damage |
| NULL | Null Data Qualifier | AX | Precision check performed |
| NULL | Null Data Qualifier | AY | QC Control points (Zero /Span) |
| NULL | Null Data Qualifier | AZ | QC audit |
| NULL | Null Data Qualifier | BA | Maintenance / routine repairs |
| NULL | Null Data Qualifier | BB | Unable to reach site |
| NULL | Null Data Qualifier | BC | Multipoint calibration |
| NULL | Null Data Qualifier | BD | Auto calibration |
| NULL | Null Data Qualifier | BE | Building site repair |
| NULL | Null Data Qualifier | BF | Precision, zero or span performed |
| NULL | Null Data Qualifier | BI | Lost or damaged in transit |
| NULL | Null Data Qualifier | BJ | Operator Error |
| NULL | Null Data Qualifier | BK | Site computer/data logger down |
| QA | Quality Assurance Qualifier | 2 | Operational Deviation |
| QA | Quality Assurance Qualifier | 3 | Field Issue |
| QA | Quality Assurance Qualifier | V | Validated value |
| QA | Quality Assurance Qualifier | W | Flow rate average out of specs. |
| QA | Quality Assurance Qualifier | X | Filter temperature difference out of spec. |
| QA | Quality Assurance Qualifier | **HT** | Sample pick-up hold time exceeded; data questionable |

### 5.3.1.4 Reporting Units

For the NATTS program, air toxics data are to be reported in engineering units of $\mu g/m^3$. With the exception of VOCs, data must be reported in local conditions. Data should not be reported in standard conditions (standard conditions are defined as data adjusted to 760 mm Hg and 298.2$^o$ Kelvin).

### 5.3.1.5 Flagging for Collocated, Duplicate and Replicate Data

The addition of duplicate/collocated replicate split samples is new and noted in **bold.** Additional information on reporting precision data can be found in EPA's AQS Data Coding Manual.

**Precision Information—**Method precision is determined by qualifying the variability associated with sample collection, and the variability associated with sample analysis. Collection precision determination is addressed through collection and analysis of collocated samples or duplicates samples. Analytical precision is addressed through the replicate analyses of these collocated or duplicate samples. The definitions of these terms are:

- Collocated Sample Collection **–** Collocated samples are samples collected simultaneously using two completely separate sampling systems, and then analyzing the samples and comparing the results obtained. This approach provides information on "Inter-system" variability.

- Duplicate Sample Collection **–** Duplicate samples are samples collected simultaneously using one collection system (i.e., two separate samples through the same sampling system at the same time), and then analyzing the samples and comparing the results obtained. This simultaneous collection is typically accomplished by teeing the line from the from the flow control device to the canisters, and then doubling the collection flow rate. This approach provides information on "Intra-system" variability.

- Replicate Analysis **–** Replicate analyses is the analysis of one discrete sample multiple times. These are also known as "split" sample analyses. This approach provides information on "Analytical" variability.

Precision ID will differentiate the duplicates, replicates, and primary-replicate duplicate / collocated-replicates as follows:

1) Use Precision ID = '1' for the duplicate or collocated;
2) Use Precision ID = '2' for the primary replicate;
3) Use Precision ID = '**31**' for the duplicate/collocated-replicate sample #1;
4) Use Precision ID = '**32**' for the duplicate/collocated-replicate sample #2.

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'). Below is an example scenario that outlines the coding required.

Example: Adding duplicate/collocated, replicate, and duplicate-replicate information

Duplicate/collocated 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)
```

Primary 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)
```

Duplicate/collocated-replicate #1 information reported with Precision ID = **31**

```
RP|I|10|003|1011|45201|6|31|7|001|110|20000101|1.35|110|1.33|||
                         ^          ^              ^   ^    ^
                  Precision ID   Prim. Meth.    Prim. Value  Dup-Rep.  Meth (l) & Value (r)
```

Duplicate/collocated-replicate #2 information reported with Precision ID = **32**

```
RP|I|10|003|1011|45201|6|32|7|001|110|20000101|1.35|110|1.33|||
                         ^          ^              ^   ^    ^
                  Precision ID  Prim. Meth.     Prim. Value  Dup-Rep.  Meth (l) & Value (r)
```

### *5.3.2 Data Entry into AQS*

EPA's AQS[1] contains ambient air pollution data collected by EPA, state, local, and tribal air pollution control agencies from thousands of monitoring stations. AQS also contains meteorological data, descriptive information about each monitoring station (including its geographic location and its operator), and data QA/QC information. AQS users rely upon the system data to assess air quality, assist in attainment/nonattainment designations, perform modeling for permit review analysis, and other air quality management functions. With quarterly reporting of data to AQS, the NATTS Program will use data gathered in AQS to assess trends in air quality data.

The AQS data base is EPA's data management repository for NATTS Program network data, which contains validated measurements of ambient concentrations of air pollutants and associated meteorological data. As with other types of EPA ambient air monitoring programs (i.e., criteria pollutants, PAMS, etc.), NATTS Program data must be prepared and entered into AQS. Data preparation and entry is the responsibility of each participating agency.

To enter data into AQS, the user must connect to the AQS website and also to the Central Data Exchange (CDX) website using a web browser to facilitate transferring input files. The user must be registered and have a valid password (i.e., which are the same for both AQS and CDX). Specific details of these procedures are available at:

http://www.epa.gov/tn/airs/airsaqs/manuals/manuals.htm.

If required, additional assistance is available by calling the AQS help line at (866) 411-4372.

### *5.3.3 VOCDat*

VOCDat is a software package developed to display VOC data, to perform QC tasks on the data, to allow an analyst to begin exploratory data analysis, and to prepare data for entry into EPA's AQS data base. The most current issuance is version 2.61. VOCDat can be used for data collected on an hourly basis with automated gas chromatography systems or on other sampling intervals (e.g., 3-hour, 24-hour) with canisters. One of the goals underlying the development of this software was to enable states to rapidly validate and submit their data. The software and user manual[5] are available from STI at ftp://ftp.sonomatech.com/public/vocdat/ .

### 5.3.4 Aethalometer Data Management

The Washington University Air Quality Laboratory (WUAQL) has extensive experience using the dual channel version of the aethalometer, as well as processing and interpreting resultant data, and provides user support for the Magee Scientific Aethalometer (available at http://www.seas.wustl.edu/user/jrturner/aethalometer/). Note that the official source for information on the Magee Scientific Aethalometer is http://www.mageesci.com.

In collaboration with George Allen of Northeast States for Coordinated Air Use Management (NESCAUM), WUAQL developed a software package, *The Aethalometer Data Masher*, which processes raw 5-minute data into hourly averages and automatically performs a suite of data validation and formatting steps. The output of the Data Masher can easily be converted into AQS format using VOCDat (developed by Sonoma Technology, Inc.). The current public release version of the Data Masher software is Version 4.2g, dated November 20, 2003 (the version number and the date are displayed on the software welcome page with the executable file name). Contact Jay Turner at jrturner@seas.wustl.edu to obtain a copy of the freeware package and to be registered for automatic notification of future updates.

The WUAQL aethalometer Data Masher is a 32-bit program written in Microsoft Visual Basic 4.0. The Data Masher requires the following dynamic link library (DLL): “VB40032.DLL”. If an attempt to start the Data Masher program produces an error message stating that this dynamic link library cannot be found, the DLL can be downloaded from numerous web sites such as Freewareweb.com.

## Section 5
## References and Resources

1. *Air Quality Subsystem (AQS)*; U.S. Environmental Protection Agency: Research Triangle Park, NC, Last update: November 2001. Available at: http://www.epa.gov/airs/aqs.html.

2. Jones, Kim. *XML: The Future of Data*; White paper. 2000. Available at: http://www.forecross.com/white-papers/xml_future_of_data.htm.

3. *Quality Assurance Handbook for Air Pollution Measurement Systems: Volume II. Ambient Air Specific Methods*; EPA-600/4-77-027a; U.S. Environmental Protection Agency: Environmental Monitoring Systems Laboratory, Research Triangle Park, NC, July 1984.

4. Mendenhall, William. *Introduction to Probability and Statistics*, 4th Edition; Wadsworth Publishing: Belmont, CA, 1975.

5. Main, H.; Prouty, J. *VOCDat User's Guide*; STI-999434-2015-UG; Sonoma Technology, Inc. Available at: http://www.sonomatech.com.

6. *White Paper: Storage—Are You Backed Up?*; Sun Microsystems; 2002. Available at: http://www.sun.com/storage/white-papers/backup-article.html.

7. *AQS Data Coding Manual (AQ2)*, Version 1.1; U.S. Environmental Protection Agency: Office of Air Quality Planning and Standards, Information Transfer and Program Integration Division, Information Management Group, Research Triangle Park, NC, August 9, 2002.

8. Carter, W.P.L. Development of Ozone Reactivity Scales for Volatile Organic Compounds. *J. Air Waste Manage. Assoc.* **1994**, *44*, 881-899.