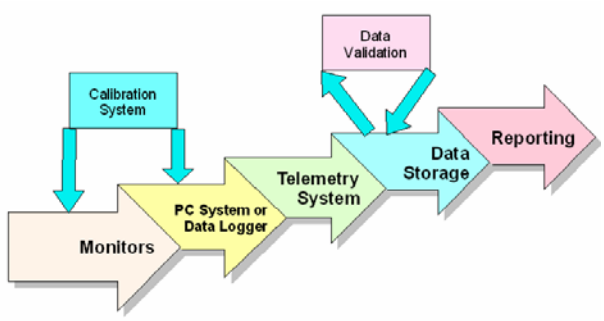
14.0 Data Acquisition and Information Management



Success of the Ambient Air Quality Program objectives relies on data and its correct interpretation. It is critical that data be available to users and that these data are:

- reliable;
- of known quality;
- easily accessible to a variety of users; and
- aggregated in a manner consistent with its prime use

In order to accomplish this activity, information must be collected and managed in a manner that protects and ensures its integrity.

Most of the data collected from the Ambient Air Monitoring Program will be collected through automated systems at various facilities. These systems must be effectively managed by using a set of guidelines and principles by which adherence will ensure data integrity. The EPA has a document entitled Good Automated Laboratory Practices (GALP)[1]. The GALP defines six data management principles:

1. DATA: The system must provide a method of assuring the integrity of all entered data. Communication, transfer, manipulation, and the storage/recall process all offer potential for data corruption. The demonstration of control necessitates the collection of evidence to prove that the system provides reasonable protection against data corruption.

2. FORMULAE: The formulas and decision algorithms employed by the system must be accurate and appropriate. Users cannot assume that the test or decision criteria are correct; those formulas must be inspected and verified.

3. AUDIT: An audit trail that tracks data entry and modification to the responsible individual is a critical element in the control process. The trail generally utilizes a password system or equivalent to identify the person or persons entering a data point, and generates a protected file logging all unusual events.

4. CHANGE: A consistent and appropriate change control procedure capable of tracking the system operation and application software is a critical element in the control process. All software changes should follow carefully planned procedures, including a pre-install test protocol and appropriate documentation update.

5. STANDARD OPERATING PROCEDURES (SOPs): Control of even the most carefully designed and implemented systems will be thwarted if appropriate procedures are not followed. The principles implies the development of clear directions and Standard Operating Procedures (SOPs); the training of all users; and the availability of appropriate user support documentation.

6. DISASTER: Consistent control of a system requires the development of alternative plans for system failure, disaster recovery, and unauthorized access. The control principle must extend to planning for reasonable unusual events and system stresses.

[1] http://www.epa.gov/irmpoli8/ciopolicy/2185.pdf

The principles listed above apply to both the local and central information management systems. The ambient pollutant data generated by gas analyzers or manual samplers must be captured, organized, and verified in order to be useful. The process of capturing the data is known as data acquisition. The organization of the data is known as data management. This section provides guidance in these areas, including identification of advanced equipment and procedures that are recommended for implementation. The recommended procedures rely on digital communication by the data acquisition system to collect a wider variety of information from the analyzers, to control instrument calibrations, and to allow for more routine, automated, and thorough data quality efforts. The section will discuss:

1. **Data acquisition-** collecting the raw data from the monitor/sampler, storing it for an appropriate interval, aggregating or reducing the data, and transferring this data to final storage in a local data base (monitoring organizations database)
2. **Data transfer**- preparing and moving data to external data bases such as AIRNow or the Air Quality System (AQS).
3. **Data management**- ensuring the integrity of the data collection systems

In response to guidelines issued by the Office of Management and Budget (OMB) under Section 515(a) of the Treasury and General Government Appropriations Act for Fiscal Year 2001 (Public Law 106-554; H.R. 5658), EPA developed the document titled *Guidelines for Ensuring and Maximizing the Quality, Objectivity, Utility, and Integrity of Information Disseminated by the Environmental Protection Agency*[2]. The Guideline contains EPA's policy and procedural guidance for ensuring and maximizing the quality of information it disseminates. The Guideline also incorporates the following performance goals:

- Disseminated information should adhere to a basic standard of quality, including objectivity, utility, and integrity.
- The principles of information quality should be integrated into each step of EPA's development of information, including creation, collection, maintenance, and dissemination.
- Administrative mechanisms for correction should be flexible, appropriate to the nature and timeliness of the disseminated information, and incorporated into EPA's information resources management and administrative practices.

EPA suggests monitoring organizations review this document since it is relevant to the ambient air information it generates and can help to ensure that data can withstand challenges to its quality.

14.1 Data Acquisition

Data acquisition technology is advancing and ever changing. Computer systems are now available in most air quality instruments. This has changed data acquisition in a profound way; most data is available in an instantaneous digital format from the instrument. This can be a powerful tool to quickly recognize and mitigate data quality problems. These digital systems should increase data capture and reporting. On the other hand, this increase in instantaneous data can be overwhelming if the monitoring organization is not prepared. The timely reporting of high quality, highly time-resolved ambient monitoring data will require a coordinated effort to ensure data management systems are meeting desired performance needs. These data management systems will need to provide validated data, to the extent possible, in near real time to multiple clients within minutes from the end of a sample period. Data management systems used

[2] http://www.epa.gov/quality/informationguidelines/documents/EPA_InfoQualityGuidelines.pdf

in ambient air monitoring will need to provide efficient processing and validation of data, and provide appropriate communication of that data in a format appropriate and available for multiple users. As an example, improved data management systems from all NCore continuous monitors can provide near real-time, high quality, hourly data during episodes. This will allow technical and policy staff to better understand the exposure and interactions of air pollutants in the atmosphere of most interest. This section provides information on Data Acquisition Systems (DAS), a term used for systems that collect, store, summarize, report, print, calculate or transfer data. The transfer is usually from an analog or digital format to a digital medium. This section will also discuss limitations of data collected with DAS.

14.1.1 Automated Data Acquisition Requirements

DAS have been available to air quality professionals since the early 1980s. The first systems were single and multi-channel systems that collected data on magnetic media. This media was usually hand transferred to a central location or laboratory for downloading to a central computer. With the advent of digital data transfer from the stations to a central location, the need to hand transfer data has diminished. However, errors in data reporting can occur with digital data. For DAS, there are two sources of error between the instrument (sensor) and the recording device: 1) the output signal from the sensor, and 2) the errors in recording by the data logger. For DAS that collect digital meta and reported data, these are not issues. Digital transfer of data does not suffer from the same problems as digital to analog transfer. When one digital device sends digital signals, the data is sent in data package streams that are coded then decoded at the receiving end. This digital transfer does not suffer from signal degradation. Most automated data acquisition systems support the acquisition of QC data like zero, one point QC and span data. One way to ensure that the QC data are correctly merged with the ambient readings is to code the QC values directly into the data set at the location corresponding to the time of the checks, replacing the normal hourly reading that is lost anyway because of the check. These data can be marked or flagged to differentiate it from ambient data and later deleted from the final routine data report printout. When QC data is acquired automatically by a data acquisition system for direct computer processing, the system must be sufficiently sophisticated to:

- ensure that the QC data is never inadvertently reported as ambient measurements,
- ignore transient data during the stabilization period before the analyzer has reached a stable QC response (this period may vary considerably from one analyzer to another),
- average the stable QC readings over some appropriate time period so that the readings obtained accurately represents the analyzer's QC response,
- ignore ambient readings for an appropriate period of time immediately following a QC reading until the analyzer response has restabilized to the ambient-level concentration.

14.1.2 Instrument to Data logger

Figure 14.1 shows the basic transfer of data from the instrument to the final product; a hard copy report, or data transfer to a central computer. Most continuous monitors have the ability to output data in at least two ways*:* analog output and an RS232 digital port. Some instrumentation may now be including USB, Ethernet and firewire capability. The instrument has a voltage potential that generally is a DC voltage. This voltage varies directly with the concentration collected. Most instruments' output is a DC voltage in the 0-1 or 0-5 volts range. The following provide a brief summary of the analog (A) or digital (D) steps

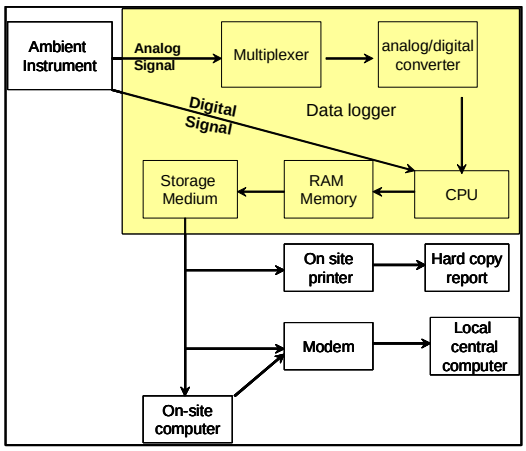


Figure 14.1 DAS data flow

- (A) the voltage is measured by the multiplexer which allows voltages from many instruments to be read at the same time.
- (A) the multiplexer sends a signal to the a/d converter which changes the analog voltage to a low amperage digital signal.
- (A) the a/d converter send signals to the central processing unit (cpu) that directs the digital electronic signals to a display or to the random access memory (ram) which stores the short-term data until the end of a pre-defined time period.
- (A/D) the cpu then shunts the data from the ram to the storage medium which can be magnetic tape, computer hard-drive or computer diskette.
- (A/D) the computer storage medium can be accessed remotely or at the monitoring location.

The data transfer may occur via modem to a central computer storage area or printed out as hard copy. In some instances, the data may be transferred from one storage medium (i.e. hard drive to a diskette, tape, or CD) to another storage medium. The use of a data logging device to automate data handling from a continuous sensor is not a strict guarantee against recording errors. Internal validity checks are necessary to avoid serious data recording errors.

Analog Versus Digital DAS -

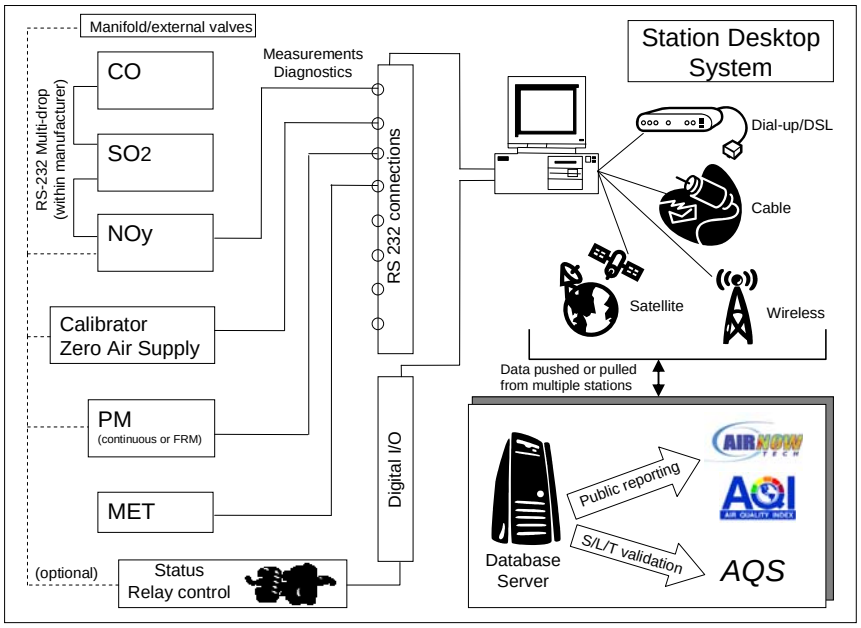


Figure 14.2 Flow of data from gas analyzers to final reporting

Most analyzers built within the last 15 years have the capability (RS232 ports) to transfer digital signals, yet many monitoring organizations currently perform data acquisition of automated monitors by recording an analog output from each gas analyzer using an electronic data logger. As explained above, the analog readings are converted and stored in digital memory in the data logger for subsequent automatic retrieval by a remote data management system. This approach can reliably capture the monitoring data, but does not allow complete control of monitoring operations, and the recorded analog signals are subject to noise that limits the detection of low concentrations. Furthermore, with the analog data acquisition approach, the data review process is typically labor-intensive and not highly automated. For these reasons, EPA

encourages the adoption of digital data acquisition methods. In that regard, the common analog data acquisition approach often does not fully utilize the capabilities of the electronic data logger. Many data loggers have the capability to acquire data in digital form and to control some aspects of calibrations and analyzer operation, but these capabilities are not utilized in typical analog data acquisition approaches.

Digital data acquisition reduces noise in the recording of gas monitoring data, thereby improving sensitivity. It also records and controls the instrument settings, internal diagnostics, and programmed activities of monitoring and calibration equipment. Such data acquisition systems also typically provide automated data quality assessment as part of the data acquisition process.

It may be cost-effective for monitoring organizations to adopt digital data acquisition and calibration control simply by more fully exploiting the capabilities of their existing electronic data loggers. For example, many gas analyzers are capable of being calibrated under remote control. The opportunity to reduce travel and personnel costs through automated calibrations is a strong motivator for monitoring organizations to make greater use of the capabilities of their existing data acquisition systems. The NCore multi-pollutant sites are taking advantage of the newer DAS technologies. Details of these systems can be found in the technical assistance document for this program[3].

Figure 14.2 illustrates the recommended digital data acquisition approach for the NCore sites. It presents the data flow from the gas monitors, through a local digital data acquisition system, to final reporting of the data in various public databases. This schematic shows several of the key capabilities of the recommended approach. A basic capability is the acquisition of digital data from multiple analyzers and other devices, thereby reducing noise and minimizing the effort needed in data processing. Another capability is two-way communication, so that the data acquisition system can interrogate and/or control the local analyzers, calibration systems, and even sample inlet systems, as well as receive data from the analyzers. Data transfer to a central location is also illustrated, with several possible means of that transfer shown. Monitoring organizations are urged to take advantage of the latest technology in this part of the data acquisition process, as even technologies such as satellite data communication are now well established, commercially available, and inexpensive to implement for monitoring operations.

Depending on the monitoring objective, it may be important that data are reported in formats of immediate use in public data bases such as AQS[4], and the multi-monitoring organization AIRNow[5] sites. An advantage of DAS software is the ability to facilitate the assembly, formatting and reporting of monitoring data to these databases.

Digital data acquisition systems such as those in Figure 14.2 offer a great advantage over analog systems in the tracking of calibration data, because of the ability to control and record the internal readings of gas analyzers and calibration systems. That is, a digital data acquisition system not only can record the analyzer's output readings, but can schedule and direct the performance of analyzer calibrations, and record calibrator settings and status. Thus, flagging of calibration data to distinguish them from ambient monitoring data is conducted automatically during data acquisition with no additional effort or post-analysis. These capabilities greatly reduce the time and effort needed to organize and quantify calibration results.

---

[3] Version 4 of the Technical Assistance Document for Precursor Gas Measurements in the NCore Multi-pollutant Monitoring Network. http://www.epa.gov/ttn/amtic/pretecdoc.html
[4] http://www.epa.gov/ttn/airs/airsaqs/aqsweb/
[5] http://airnow.gov/

14.1.3 DAS Quality Assurance/Quality Control

Quality assurance aspects of the DAS deal with whether the system is being operated within defined guidelines. Usually, this means that each value that is collected on the DAS is the same value that is generated from the analyzer and reported to the Air Quality System (AQS) data base. This usually is accomplished by calibrations, data trail audits and performance audits.

Calibration- In the case where analog signals from monitoring equipment are recorded by the DAS, the calibration of a DAS is similar to the approach used for calibration of a strip chart recorder. To calibrate the DAS, known voltages are supplied to each of the input channels and the corresponding measured response of the DAS is recorded. Specific calibration procedures in the DAS owner's manual should be followed when performing such DAS calibrations. For DAS that receive digital data from the instruments, a full scale check (the instrument is in a mode and the output is at the full scale of the instrument) should be performed to see if the data received digitally is the same as the display of the instrument. The DAS should be calibrated at least once per year. Appendix G provides a simple approach for calibration of the DAS.

In addition, gas analyzers typically have an option to set output voltages to full scale or to ramp the analog output voltages supplied by the analyzer over the full output range. Such a function can be used to check the analog recording process from the analyzer through the DAS.

Data Trail Audit- The data trail audit consists of following a value or values collected by the DAS to the central data collection site and then eventually to AQS. A person other than the normal station operator should perform this duty. The following procedure should be followed:

- A data point should be collected from the DAS (usually an hourly value or another aggregated value reported to AQS) and be checked on the DAS storage medium against the hard copy report. Also if strip chart recorders are used, a random number of hourly values should be compared to the data collected by the DAS. This audit should be completed on a regular defined frequency and for every pollutant reported.
- From the central computer, the auditor checks to see if this hourly value is the same.

The above actions should be completed well in advance of data submittal to AQS. If the data has been submitted to AQS, then the AQS data base should be checked and modified as necessary per the appropriate AQS procedures.

Whether a monitoring organization is transferring the data from an instrument via an on-site DAS or transferring the data digitally, the data trail audit should be performed on a routine basis.

Performance Audit- The performance audit consists of challenging the instrument and DAS to a known audit source gas and observing the final response. The response should correspond to the value of the audit source gas. Section 15 discusses these performance audits.

Initialization Errors

All data acquisition systems must be initialized. The initialization consists of an operator "setting up" the parameters so that the voltages produced by the instruments can be read, scaled correctly and reported in the correct units. Errors in initializations can create problems when the data is collected and reported. Read the analyzer manufacturer's literature before parameters are collected. If the manufacturer does not

state how these parameters are collected, request this information. The following should be performed when setting up the initializations:

- Check the full scale outputs of each parameter.
- Calibrations should be followed after each initialization (each channel of a DAS should be calibrated independently). Appendix G provides an example of a DAS calibration technique.
- Review the instantaneous data stream, if possible, to see if the DAS is collecting the data correctly.
- Save the initializations to a storage medium; if the DAS does not have this capability, print out the initialization and store it at the central computer location and at the monitoring location.
- Check to see if the flagging routines are performed correctly; data that are collected during calibrations and down time should be flagged correctly.
- Check the DAS for excessive noise (variability in signal). Noisy data that are outside of the normal background are a concern. Noisy data can be caused by improperly connected leads to the multiplexer, noisy AC power, or a bad multiplexer. Refer to the owner's manual for help on noisy data.
- Check to see that the average times are correct. Some DAS consider 45 minutes to be a valid hour, while others consider 48 minutes. Agency guidelines should be referred to before setting up averaging times.

14.1.4 Data Logger to Database

Once data are on the data logger at the ambient air monitoring station, they need to be sent to servers where they can be summarized and disseminated to data users. In most cases this will occur by using a server at the office of the monitoring organization. The conventional way to get data from the monitoring stations has been to poll each of the stations individually. With more widespread availability of the internet, pushing data from monitoring sites on a regular basis will be especially effective in mapping and public reporting of data. Note, in some cases it is possible to report data directly from a monitor to a database without the use of a station data logger. This solution is acceptable so long as the monitor is capable of data storage for periods when telemetry is off-line.

Data transfer is usually accomplished in three ways: hard copy printout, downloading data from internal storage medium to external storage medium, or digital transfer via the telephone lines, internet, satellite or other advanced means of communication. Due to the desire for real time data for the Air Quality Index (AQI) and other related needs, monitoring organizations should plan to upgrade to digital data acquisition and communication systems.

Hard copy report- Most DAS have the ability to create a hard copy report. Usually, this report is in tabular format showing 1 minute, 5 minute or hourly averages. Monitoring organization are encouraged to keep hard copy printouts for several reasons*:*

- they can be reviewed by the station operators prior to and/or during site visits to ascertain the quality of the data;
- they can be compared against the historical data stored on the DAS at the site for validation;
- notes can be made on the hard copy reports for later review by data review staff; and
- they create a "back-up" to the electronically based data.

NOTE: It is strongly recommended that monitoring organizations create an electronic back-up of their data on a defined schedule. The frequency of the back-ups and any other associated

information should be reflected in their Quality Assurance Project Plan (QAPP) and Standard Operating Procedures (SOP).

External Storage- This term refers to storing and transferring the data on diskettes or CD. Many new generation DAS are computer platforms. The newer generation computers generally have the ability to download data to CD or zip drive. If remote access via telephone is not an option, then data can be hand transferred to a central office for downloading and data review.

Digital Transfer- All new generation DAS allow access to the computer via the telephone and modem. These systems allow fast and effective ways to download data to a central location. The EPA recommends using these systems for the following reasons:

- in case of malfunction of an ambient instrument, the appropriate staff at the central location can begin to diagnose problems and decide a course of action;
- down loading the data allows the station operators, data processing team, and/or data validators to get a head start on reviewing the data; and
- when pollution levels are high or forecasted to be high, digital transfer allows the pollution forecaster the ability to remotely check trends and ensure proper operation of instruments prior to and during an event.

14.1.5 DAS Data Review

The data review is an ongoing process that is performed by the station operators (SO) and the data processing team (DP). At a minimum a cursory review is performed daily, preferably in the morning to provide a status of the data and instrument performance at monitoring sites. Detailed analysis can be extremely difficult for the data processing team when reviewing the raw data without the notations, notes and calibration information that the station operators provide for the group. The typical review process for the station operator and data reviewer(s) include:

- (SO) Review of zero, span, one point QC verification information, the hourly data, and any flags that could effect data and record any information on the daily summaries that might be vital to proper review of the data.
- (SO) Transfer strip charts both analog and digital information, daily summaries, monthly maintenance sheets, graphic displays of meta data and site log notes to the central location for a secondary and more thorough review.
- (SO) At the central location, review the data, marking any notations of invalidations and provide electronic strip charts, meta data charts, daily summaries, site notes, and monthly maintenance sheets for ready access by the data processing staff.
- (DP) Review zero, span and one point QC verifications, station notes, and monthly maintenance sheets for the month; check a percentage of all zero, span and one point verifications. Compare a defined number of hand reduced and/or strip chart readings to electronic data points generated by the DAS. If significant differences are observed, determine what corrective action steps are required.

Outliers

Outliers are "measurements that are extremely large or small relative to the rest of the data and are suspected of misrepresenting the population from which they were collected" (EPAQA/G9R)[6]. When reviewing data, some potential outliers will be obvious such as, spikes in concentrations, data remaining the same for hours, or a sudden drop in concentration but still in the normal range of observed data. Many of these outlier checks can be automated and provide efficient real-time checks of data. Outliers do not necessarily indicate the data is invalid; they serve to alert the station operator and/or data reviewers there may be a problem. In fact, the rule of thumb for outliers should be that the data be considered valid until there is an explanation for why the data should be invalidated. At some point it may be necessary to exclude outliers from instantaneous reporting to the AIRNow network and/or AQI reporting until further investigation has occurred. EPA Guidance Documents[7] *Guidance on Environmental Data Verification and Validation* (EPA QA/G8) and *Guidance for Data Quality Assessment – a Reviewers Guide* (EPA QA/G9R) provide insight on outlier and data reviews in general.

14.2 Data Transfer – Public Reporting

The area of public reporting for air monitoring data may provide the largest number of users of data. This area has been growing rapidly in the last few years as a result of the increased availability of air quality reporting, especially for ozone and $PM_{2.5}$. For public reporting of the AQI, the AIRNow web site will remain the EPA's primary medium for distribution of air monitoring data. The additional continuous monitoring parameters collected from NCore will also be reported to AIRNow. These parameters are expected to be made publicly available for sharing throughout technical user communities. However, they are not expected to be widely distributed through AIRNow as products for public consumption.

This section will discuss the transfer of data from the monitoring organization to two major data repositories: 1) AIRNow for near real-time reporting of monitoring data, and 2) AQS for long term storage of validated data.

14.2.1 Real-time Data Reporting for AIRNow and NCore

One of the most important emerging uses of ambient monitoring data has been public reporting of the Air Quality Index (AQI). This effort has expanded on EPA's AIRNow web site from regionally-based near real-time ozone mapping products color coded to the AQI, to a national multi-pollutant mapping, forecasting, and data handling system of real-time data. Since ozone and $PM_{2.5}$ drive the highest reporting of the AQI in most areas, these two pollutants are the only two parameters currently publicly reported from AIRNow. While other pollutants such as CO, SO_2, NO_2, and PM_{10} may not drive the AQI, they are still important for forecasters and other data users to understand for model evaluation and tracking of air pollution episodes. Therefore, the NAAMS seeks the following goals:

- Share all continuous O_3, $PM_{2.5}$ and PM_{10} data, where available, across the nation;
- For NCore sites, share all gaseous CO, SO_2, NO and NOy data and base meteorological measurements across the nation.

[6] http://www.epa.gov/quality1/qs-docs/g9r-final.pdf
[7] http://www.epa.gov/quality1/qa_docs.html

This program allows for short term non-validated data to be collected by a centrally located computer that displays the data in near real time data formats such as tables and contour maps. In addition, EPA, in conjunction with the monitoring organizations, developed the National Ambient Air Monitoring Strategy (NAAMS) which includes the development of the NCore network. This section will discuss the needs of real time data acquisition for the deployment of AIRNow and the NAAMS**.**

Reporting Intervals

Currently, hourly averages are the reporting interval for continuous particulate and gaseous data. These are the reporting intervals for both AQS (AQS supports a variety of reporting intervals) and to AIR**N**ow for AQI purposes**.** These reporting intervals will meet most of the multiple objectives of NCore for supporting health effects studies, AQI reporting, trends, NAAQS attainment decisions, and accountability of control strategies. However, with these objectives also comes the desire for data at finer time resolutions: 5 minute averages for gaseous pollutants and sub-hourly averages for certain particulate matter monitors. Examples of this need for finer time resolution of data include, but are not limited to: tracking air pollution episodes, providing data for exposure studies, model evaluation, and evaluating shorter averaging periods for potential changes to the NAAQS. Monitoring organizations generally have the hardware and software necessary to log and report this data. The challenge to obtaining and reporting the data is the current communication packages used, such as conventional telephone modem polling. One widely available solution to this would be the use of internet connectivity, allowing data at individual monitoring sites to be pushed to a central server rather than being polled. Monitoring organizations should begin to investigate the possibilities of using this media.

With this generation of data having a shorter averaging interval, the challenge becomes validation of all the data. The historical perception has been that each criteria pollutant measurement needs to be verified and validated manually. With the amount of data generated, this would be a time-consuming task. To provide a nationally consistent approach for the reporting interval of data, the NCore networks will take a tiered approach to data reporting. At the top tier, hourly data intervals will remain the standard for data reporting. Long term, the NCore networks will be capable of providing at least 5 minute intervals for those methods that have acceptable data quality at those averaging periods. For QA/QC purposes such as zero/span and one-point QC, monitoring organizations should be capable of assessing data on at least a 1-minute interval.

With instantaneous data going to external websites, monitoring organizations operating their own websites containing the same local and/or regional data should add a statement about the quality of data being displayed at the site. This cautionary statement will notify the public that posted data has not been fully quality assured and discrepancies may occur. For an example, the AIRNow Website makes the statement

> *"Although some preliminary data quality assessments are performed, the data as such are not fully verified and validated through the quality assurance procedures monitoring organizations use to officially submit and certify data on the EPA AQS(Air Quality System). Therefore, data are used on the AIRNow Web site only for the purpose of reporting the AQI. Information on the AIRNow web site is not used to formulate or support regulation, guidance or any other Agency decision or position."*

14.2.2 Reporting Frequency and Lag Time for Reporting Data

Continuous monitoring data that are being shared in near real-time from NCore monitoring stations are to be reported each hour. Data should be reported as soon as practical after the end of each hour. For the near term, the goal is to report data within twenty minutes past the end of each hour. This will provide enough time for data processing and additional validation at the Data Management Center (DMC); generation of reports and maps; distribution of those products to a variety of stakeholders and web sites; and still allow enough time for staff review before the end of the hour. This is an important goal to support reporting of air pollution episodes on news media programs by the top of the hour. The long term goal is to report all data within five minutes after the end of an hour. This will further enhance NCore's ability to deliver timely data within a reasonable time period that takes advantage of existing commercially available technology.

14.3 Data Transfer-Reporting to External Data Bases

Today, the need for the ambient air monitoring data reaches outside the monitoring community. In addition to the traditional needs of the data, determination of NAAQS compliance and the daily AQI report, a health researcher or modeler may want a very detailed accounting of the available data in the shortest time intervals possible. Atmospheric scientists typically desire data in a relatively unprocessed yet comprehensive form with adequate descriptions (meta data) to allow for further processing for comparability to other data sets. These needs increase the demands for the data and require multiple reports of the information.

14.3.1 AQS Reporting

All ambient air monitoring data will eventually be transferred and stored in AQS. The current system, implemented in early 2002, has much more functionality than the previous main-frame system. As stated in 40 CFR Part 58.16[8], the monitoring organization shall report all ambient air monitoring and associated quality assurance data and information specified by the AQS Users Guide into the AQS format. The data is to be submitted electronically and on a specified quarterly basis. Since changes in reporting requirement occur, monitoring organization should review CFR for the specifics of this requirement.

The AQS manuals are located at the AQS Website[9]. This site contains the old AIRS/AQS manuals as well as the new AQS Manuals. The AQS Data Coding Manual replaces the previous Volume II and provides coding instructions, edits performed, and system error messages. The AQS User Guide replaces the former Volume III and describes the procedures for data entry. Both manuals will be updated as needed and the new versions will be available at the web site. Table 14-1 provides the units and the number of decimal places that, at a minimum, are required for reporting to AQS for the criteria pollutants. These decimal places are used for comparison to the NAAQS and are displayed in AQS summary reports. However, monitoring organizations can report data up to 5 values to the right of the decimal (beyond five AQS will truncate). Within the five values to the right of the decimal place, AQS will round to the minimum displayed in Table 14-1. Reported values will remain in raw data files.

[8] http://www.access.gpo.gov/nara/cfr/cfr-table-search.html
[9] http://www.epa.gov/ttn/airs/airsaqs/manuals/

Table 14-1 AQS Data Reporting Requirements

Pollutant	Decimal Places	Example	Units
$PM_{2.5}$	1	10.2	$\mu g/m^3$
PM_{10}	1	26.2	$\mu g/m^3$
$PM_{10-2.5}$	1	10.2	$\mu g/m^3$
Lead	1	1.5	$\mu g/m^3$
SO_2	2	0.03	ppm
NO_2	3	0.053	ppm
CO	1	2.0	ppm
O_3	3	0.108	ppm
PAMS (VOCs)	2	6.23	ppb-carbon

14.3.2 Standard Format for Reporting to AQS

AQS allows flexibility in reporting formats. The formats previously used by AQS can be used for raw data (hourly, daily, or composite) and for reporting precision and bias data. The system also has new report formats for this data as well as formats for registering new sites and monitors. These new formats are defined in the AQS Data Coding Manual. Work is also in progress to define an Extensible Markup Language (XML) schema for AQS to allow for that reporting format as well. Use of XML as a data format is consistent with EPA and Federal guidelines towards better data integration and sharing.

14.3.3 Annual Certification of Data

The annual data certification is also stored in AQS. The monitoring organization is required to certify the data (by formal letter) for a calendar year (Jan 1-Dec 31) by July 1 through the year 2009. Beginning in 2010 the annual data certification letter is due by May 1. See 40 CFR Part 58.15 for details. This certification requires the monitoring organization to review the air quality data and precision/bias data for completeness and validity and to submit a certification letter to the Regional Office. The certification letter and accompanying reports are reviewed and if the results of the review are consistent with the criteria for certification, the certification flag is set in the AQS database. After certification is complete, any updates to the data will cause the critical review process to identify that the certified data has been changed and the certification flag will be dropped.

14.3.4 Summary of Desired Performance for Information Transfer Systems

To define the needed performance criteria of a state-of-the art information technology system, a table of needs has been developed. This table provides performance needs for an optimal information technology system, but is not intended to address what the individual components should look like. For instance, once low level validated data for a specific time period are ready to leave the monitoring station, a number of telemetry systems may actually accomplish moving those data. By identifying the needed performance criteria of moving data, rather than the actual system to move it, monitoring organizations may be free to identify the most optimal system for their network. Table 14-2 summarizes the performance elements of the data management systems used to log, transfer, validate, and report data from NCore ambient air monitoring stations.

Table 14-2 NCore (Level 2 and 3) Information Technology Performance Needs

Performance Element	Performance Criteria	Notes
Sample Periods	5 minutes (long term goal), and 1 hour data (current standard)	5 minutes and 1 hour data to support exposure, mapping and modeling. 1 hour data for Air Quality Index reporting and NAAQS. Sample period may need to be higher for certain pollutant measurement systems depending on method sample period and measurement precision when averaging small time periods.
Data Delivery	Near Term goal - Within 20 minutes nationally each hour Long term goal - Within 5 minutes nationally each hour	As monitoring organizations migrate to new telemetry systems the goal will be to report data within 5 minutes. This should be easily obtained with broadband pushing of data to a central server.
Low Level Validation	- Last automated zero and QC check acceptable - Range check acceptable - Shelter parameters acceptable -Instrument parameters acceptable	Other validation should be applied as available: - site to site checks - rate of change -lack of change.
Data Availability	- all QC data, operator notes, calibrations, and pollutant data within network - Low level validated pollutant data externally	Create log of all monitoring related activities internally. Allow only validated data to leave monitoring organization network.
Types of monitoring data to disseminate-externally	-continuous and semi-continuous pollutant data -accompanying meteorological data	Associated manual method supporting data (for instance FRM ambient Temperature) should be collected but not reported externally.
Additional data for internal tracking	Status of ancillary equipment such as shelter temperature, power surges, zero air system, calibration system	
Relevant site information	Latitude, longitude, altitude, land use category, scale of representativeness, pictures and map of area	Other site information may be necessary.
Remote calibration	Ability to initiate automated calibrations on regular schedule or as needed	
Reviewing calibration	- allow for 1 minute data as part of electronic calibration log	
Initialization of manual collection method	Need to be able to remotely initiate these or have them set at an action level from a specific monitor	
Reporting Format	Short Term - Maintain “Obs” file format and pipe delimited formats for AIRNow and AQS reporting, respectively Near Term -XML	Need to coordinate development of XML schema with multiple stakeholders. XML is an open format that will be able to be read by most applications.

14.4 Data Management

Managing the data collected is just as important as correctly collecting the data. The amount of data collected will continue to grow based on the needs of the data users. Previous sections have confirmed this statement providing a glimpse of the potential data users and the uses. Generally, data is to be retained for a period of 3 years from the date the grantee submits its final expenditure report unless otherwise noted in the funding agreement. Refer to 40 CFR Part 31.42. With electronic records and electronic media, this information can be stored and managed with less use of space than with the conventional paper records. However, even with today’s technology there will be some paper records and those need to be managed in an orderly manner. The manner in which a monitoring organization manages its data is documented in its QMP and QAPP.

All information collected in any ambient air monitoring program should be organized in a logical and systematic manner. There is no one best way to organize a system. How a monitoring organization

organizes its information is required to be discussed in its QMP (QA/R-2)[10] and QAPP (QA/R-5)[11]. Monitoring organizations should consult EPA's records management webpage[12] for other useful information when beginning to plan or revise how its data records are stored.

This information should be reviewed not only by those in a monitoring organization responsible for overall data management but also by the monitoring organization's Systems or Network Administrator. The latter person(s) can provide helpful information in designing the overall data management system according to today's industry standards. Remember, the data has to be of known quality, reliable and defensible. In order for monitoring organizations to continue to meet those objectives, many sources of information need to be reviewed.

Section 5 presented guidance on documentation and records. This information can be helpful in managing ambient air monitoring data. In addition, the EPA Office of Environmental Information (OEI) has a website[13] that provides information management policies and guidance. As an example the document *Good Automated Laboratory Practices,* described earlier in this document, is posted on the OEI website and can be very useful in developing information management systems.

[10] http://www.epa.gov/quality1/qs-docs/r2-final.pdf
[11] http://www.epa.gov/quality1/qs-docs/r5-final.pdf
[12] http://www.epa.gov/records/
[13] http://www.epa.gov/irmpoli8/policies.htm