United States Environmental Protection Agency

Guide to Managing Air Quality Data

For community members & organizations looking to understand air quality data management options and decisions

This guide will lead you through the choices you'll need to make when beginning to manage air quality data. A glossary containing technical and subject-specific terms can be found at the end of the document.

Table of Contents

What is Data Management?

Data management is the process of collecting, storing, and using data securely, efficiently, and cost-effectively. It ensures your data are complete, correct, and accessible enough in your air monitoring project. Properly managing data takes time and planning, and good data management results in higher-quality data that helps you make sound decisions and achieve your desired outcome.

Benefits of Data Management

- **Saves your project time and money**
- **Increases data quality and credibility**
- **Enables problem detection & reduces data loss**
- **Makes data more usable & easier to share**

Functions of Data Management

Data management is an important component of air monitoring (see *Figure A*), consisting of a software/hardware system combined with quality assurance (QA) procedures. The three primary functions of data management are to 1) take in data from air monitoring devices, 2) store, process, and quality control (QC) data, and 3) distribute and share the data with others.

Figure A. *Components of an air monitoring project. Good data management helps connect air quality measurement device data with users and results.*

Four Steps to Managing Air Quality Data

This guide will lead you through the choices you'll need to make when beginning to manage air quality data. A glossary containing technical and subject-specific terms can be found at the end of the document.

1. Start Early

Start thinking about data management when you first consider monitoring air quality. There are many decisions to make, procedures to develop, and details to determine. Ultimately, your purpose for monitoring air quality will drive many of your decisions in selecting a data management system that's right for you. For more on determining an air monitoring purpose, check Chapter 3 of the EPA's *Enhanced Air Sensor Guidebook* (https://cfpub.epa.gov/si/si_public_record_report.cfm?Lab=CEMM&dirEntryId=356426).

As described in the *Enhanced Air Sensor Guidebook* and shown in the figure on the next page (see *Figure B*), there are five major steps to air monitoring. In each of these steps, data management is foundational. For example, in the *Plan* step, you'll want to identify a system to manage the data. In the *Setup* and *Collect* steps, you can use the data management system to review and compare measurements to ensure correct operations and data quality. Starting early will ensure that you collect useful data and that the purpose of the air monitoring is met.

Five Steps Recommended for Planning Air Monitoring Projects

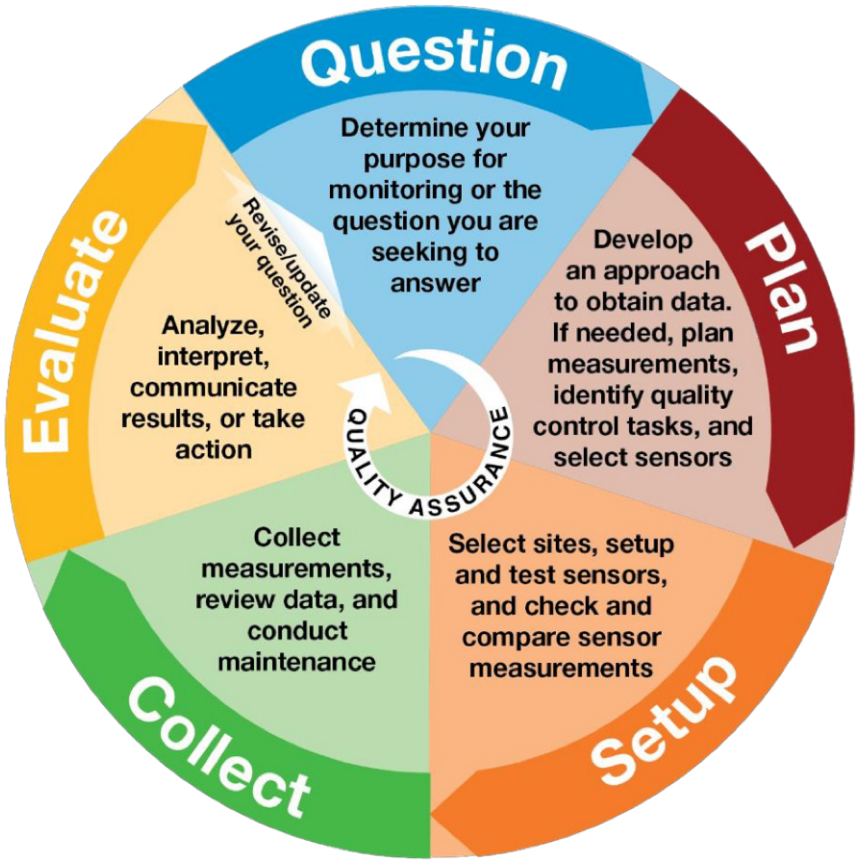

Figure B. *In each of these five air monitoring steps, data management is foundational.*

2. Choose a Data Management System

A Data Management System (DMS) is a collection of software and hardware that streamlines the operations of transferring, storing, processing, and distributing data. This section describes the three types of air quality DMSs and includes a checklist of questions on the next page to help you select the best DMS for your air monitoring work.

2a. Different Types of Data Management Systems

It may be relevant to use a combination of these DMSs. For example, a community using two different types of air monitoring devices may want to use a device vendor's DMS to ingest data and then build their own DMS to analyze and display the data.

	 Bundled with Hardware	 Buy	 Build
	Device Vendor DMS	**Third-Party Vendor DMS**	**Build Your Own DMS**
Description	Air monitoring companies and manufacturers often provide a DMS integrated with their air monitoring devices.	Companies provide third-party DMSs focused specifically on managing data from many different devices. These systems are often customized to meet specific air quality needs.	Develop your own DMS by writing software or using existing tools. This approach can range from a spreadsheet for several air monitoring devices to complex database and webpage systems. Be aware: building any DMS can be a complex and time-consuming endeavor.
Best for	Getting started with air monitoring, especially if you have a limited budget and only one type of air monitoring device.	Good for networks with different types of air monitoring devices in need of more advanced features such as: quality assurance features, data visualization, and data exporting.	An organization with the people, skills, and resources to develop a custom data management system.
Pros	• Integrated with the vendor's devices • Often bundled with purchase of a device • Usually quick to set up • Features tuned to their product • Customer support	• Works with lots of devices from different manufacturers • Typically has advanced features (quality assurance, maintenance, logging) • Customer support and training	• Customized to meet your needs • Can leverage existing open-source software and coding packages
Cons	• May only work with that company's device • May lack advanced features (e.g., quality control) • Customization is more challenging • Vendor may have access (or rights) to use your data • May have annual, recurring costs	• Cost may vary based on the size of the network: may be expensive • May have more features than you need • May have annual, recurring costs • Vendor may have access (or rights) to your data	• Depending on the complexity of your needs may take considerable time to develop • Requires software and data science skills to build, customize, and operate • May be costly to build and operate • If the system developer leaves your organization, maintaining and updating the system could become a challenge for others.

2b. What to Think About When Choosing a DMS

The following is a list of things to consider when you're determining what DMS to use. This list can be used to guide conversations with internal team members and/or external DMS providers you talk to about data management.

Size and Complexity of Air Monitoring Network. As the number of air monitoring sites and parameters (or pollutants) measured increases, so does the time it takes to collect, manage, and process data. The size of the network affects many things: people needed to maintain, automation of routine tasks, budget, etc. For larger networks, you'll want to have automated DMS features to streamline tasks and operate the network more efficiently.

Total Time of Air Monitoring Project. Consider the long-term cost of operating the DMS, where data will be stored and archived, and the potential staff or volunteer turnover and training that may be needed for a long-term air monitoring project.

User Skills. Community members will need training and technical skills to use a DMS fully. If using a vendor-provided system, check that they provide training. If you build your own DMS, consider the knowledge required to build, operate, and maintain it.

Budget. Factors affecting a DMS's cost generally include the number of features, customization needs, number of sites, length of air monitoring project, etc. You'll want to evaluate the costs of buying or using a DMS over the lifetime of your project. Some questions to consider:

- How long do you plan to use the DMS? Are there annual, recurring, or user fees?
- Does the cost of the DMS change as your network grows/shrinks?
- Does it cost to upgrade to a new (future) version?
- Does customization of the DMS cost extra? See the next section for what customization may include.
- Does the vendor charge to archive or retain the data after your project ends?
- Does training or technical support cost extra?
- If you're purchasing the DMS from a vendor, consider if you will own it forever or if it is a subscription service, and if you can modify or improve it.

Data Handling. How and where are data handled? Some questions to consider:

- Is the DMS hosted on your premises or offsite at a cloud computing facility? Do you have any hosting requirements or constraints?

- Where and how often will the data and metadata (i.e., supporting data) be archived and backed up?
- What metadata can be stored in the DMS? Can you customize these metadata fields?
- Who owns the data stored in the DMS? Some DMS providers may have agreements that give them full or partial ownership of data in their system.
- Similarly, who can use the data that are stored in the DMS? Is a data use agreement needed? For vendor and third-party DMSs, can the vendor use your data for other projects without your knowledge or permission?

3. Select Features for Data Management

Determine the tasks and steps to collect, process, view, and distribute your data. List them and have the DMS vendor demonstrate that their system meets your needs.

3a. Data Input

Receiving data from the monitoring device in the field and putting it into the DMS can be challenging. This process is often called data ingestion, and may involve data being sent to the DMS via Wi-Fi or cellular service on a frequent basis (e.g., every minute, hourly). In areas without cellular connection or Wi-Fi devices may have on-board memory (e.g., SD cards) that are then manually transferred from the device to the DMS.

Ask the vendor if the DMS can:

- ○ Ingest data from different devices? For example, your network may have air monitoring devices from two different manufacturers.
- ○ Detect problems and identify where data is missing?
- ○ Retrieve data in both real-time and from past time frames (historical data)?
- ○ Send problem (email or SMS) notification when air monitors fail to report data?
- ○ Provide summary information about the status of the monitoring network?

3b. Data Quality

Data quality is critical whether you're installing a couple of air monitoring devices or running a large network. A DMS and supporting procedures can help increase the level of data quality, increasing confidence in the accuracy of the data.

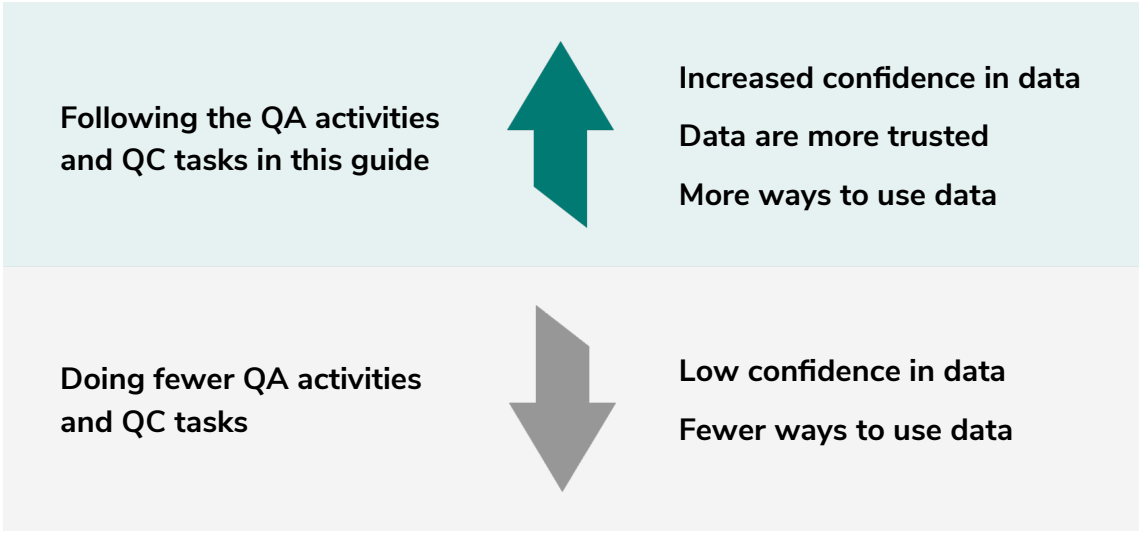

Figure C. *More quality assurance and quality control increases confidence in your data.*

Ask the vendor if the DMS has the following:

- ○ Automatic quality control (QC) checks? If so, how do they work? (i.e., automated data flags, customizable data ranges)
- ○ Tools to help users evaluate data quality?
- ○ Customizable QC screening to detect problems in the data?
- ○ Manual QC screening to adjust or flag questionable data?
- ○ Methods to store both the quality controlled and raw data?
- ○ Methods to flag data (i.e., quality indicators)?
- ○ Tools to correct or calibrate data? How do those tools work?
- ○ Ability to lock data to prevent adjustments or restrict user editing permissions?
- ○ Ability to tag data with notes to capture device maintenance records?
- ○ IT/Tecnhical support services? Is there a fee?

3c. Data Visualization

Viewing your data using graphs, tables, and displays will be needed to monitor your network, understand your data, and share and communicate it. Most DMSs include data graphing and visualization features.

Ask the DMS vendor:

- What types of displays are included with the system?
- How are these displays customizable?
- How do you access data visualizations? Is there a data visualization website?
- Can you save these visualizations?

3d. Data Sharing

Several possible methods exist to distribute data:

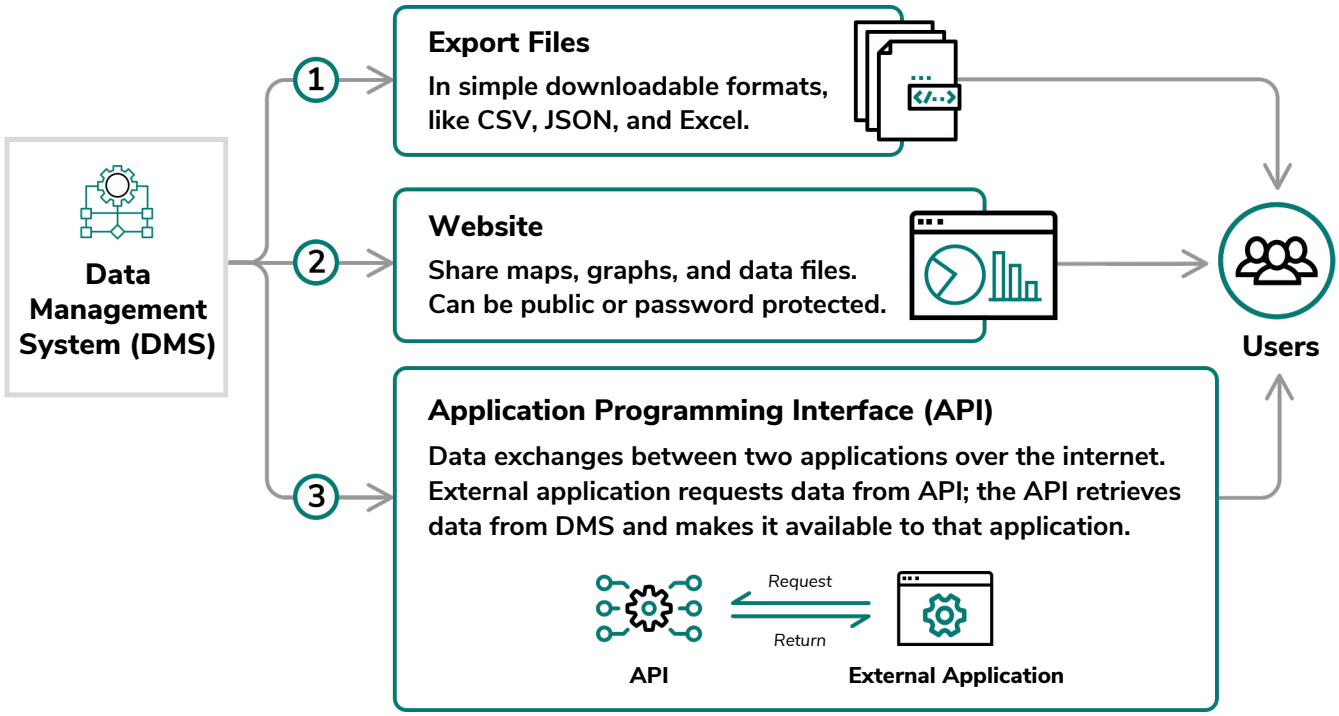

Figure D. *Schematic showing three different ways to share air quality data from a DMS: files, website, and API.*

Data-sharing questions to consider when selecting a DMS:

- What features does the DMS include for distributing data? Is there an additional charge to use these?
- How do you control who can access the data? API? Website?
- What file formats can data download as? (CSV, JSON, Excel)
- What supporting information is available to use the data (metadata, documentation)?
- Is there a website for sharing data?
- How are metadata (i.e., information about the data) shared?
- Are there any licensing, use, or ownership concerns (if the DMS provider makes claims on the data)?

4. Procedures for Setting Up & Operating

Before placing air monitoring devices to the field, you'll want to set up the DMS and supporting procedures to ensure reliable operations and higher-quality data. These procedures, sometimes called standard operating procedures (SOPs), should include the who, when, and how of achieving different data management tasks.

There are many procedures related to data management, and they are generally grouped into three tasks: setting up the network, operating and managing it, and sharing data. The following pages provide some checks for these three tasks.

Setup Checks

Do these checks to ensure the device and DMS are correctly set up and configured.

Operational Data Checks

Do these checks on a regular basis to ensure the quality of the data being collected.

Data Sharing Checks

Do these checks when sharing data with others.

Figure E. *Types of checks to perform when carrying out different data management tasks to ensure high-quality data.*

4a. Set Up Checks

As soon as an air monitoring device is running and starts sending data to a DMS, ensure everything on the device and within the DMS is correctly set up and configured. Establish a procedure to check the following things every time a new air monitoring device comes online.

Checks every time a new air monitoring device comes online:

- ○ Ensure that the DMS ingests data and the data are complete (no missing pollutants or skipped time periods).
- ○ Review the measurement values for reasonableness (e.g., realistic values). Compare with nearby measurements.
- ○ Check that the timestamp is correct and reported in the right time zone. The timestamp is the time when data are tagged by an instrument and stored in the DMS. Air quality data is typically reported in standard time (timestamps do not change with daylight savings).
- ○ Confirm that the monitor site and location information (latitude, longitude, elevation) are correct in the DMS.
- ○ Confirm that all parameter names and associated units are correct within the DMS (e.g. Fine particulate matter is labeled PM2.5 and the units are in µg/m3).
- ○ Confirm that any adjustments, corrections, or calibrations to the data are working properly.
- ○ Record other metadata in the DMS for each monitoring site, including the site address, device details, photos of the location, nearby sources, etc.
- ○ Make sure other useful data looks correct. Your device may also report diagnostic data, or data on how the hardware is performing. This can include measurements like internal box temperature, what speed air is being pumped through the instrument, or what percent an internal battery is charged. These data can be very useful for troubleshooting problems.
- ○ Confirm that archives or backups of the DMS are occurring and complete.

With procedures for these checks, you'll be confident that the air monitoring devices are working correctly.

4b. Operational Data Checks

Now that data are flowing from the air monitoring device to the DMS, establish operational (i.e., day-to-day) procedures.

Checks for day-to-day operation:

- ○ Conduct daily data checks (e.g., review plots and tables of data) to confirm that all sites report valid data for all measured parameters. If data are not received, determine what actions are necessary and who can fix the problem.
- ○ Apply automatic quality control checks to detect problems and identify questionable data.
- ○ Conduct monthly and quarterly data checks: perform an in-depth review and analyze the data to identify trends and other data features over time.
- ○ Perform quality assurance and quality control steps in the DMS to improve data quality. Document any changes or adjustments to the data.
- ○ Correct and calibrate the data. Store the calibration data. See Section 3.6 of the the EPA's Enhanced Air Sensor Guidebook for more information on calibration.
- ○ Log maintenance activities and document other network activities in the DMS (e.g., changes in site conditions and device hardware).
- ○ Maintain the DMS: Keep the system up-to-date and virus-free. Make sure there is sufficient storage space.

Once you have confidence in the data's completeness and quality, you may want to distribute and share it with others.

4c. Data Sharing Checks

Create procedures to check your data prior to sharing with others. Example checks can be found below.

Checks when sharing data:

- ○ Share the best quality (quality-controlled and calibrated) data when possible.
- ○ Utilize the features in the DMS to share data, which typically includes exporting data files, hosting a data display website, or distributing data via an API.
- ○ Provide documentation about the data. A common practice is to provide an additional file explaining device type, locations, data format, corrections or calibrations applied to the data, quality control information, etc.
- ○ Provide data users with attribution information (name of community member or organization who took the data, any rights or usage restrictions, how to reference data).
- ○ Provide contact information and notification procedures for when things go wrong.

Glossary

This glossary covers terms you're likely to encounter while managing air quality data and speaking with DMS providers or local air quality regulators.

A

Application Programming Interface (API)
Software that exchanges data between two applications over the internet. External applications request or "call" the API; the API retrieves the data from the DMS and makes it available to that application.

C

Calibration
A procedure for checking and adjusting an instrument's settings so that the measurements produced are comparable to a certified standard value. (Source: https://www3.epa.gov/ttnamti1/files/ambient/pm25/qa/vol2sec12.pdf)

Collocation
The process by which an air monitoring device and a reference instrument are operated at the same time and place under real world conditions. The siting criteria (e.g., proximity and height of the air monitoring device and the reference monitor) should follow procedures outlined in 40 CFR Part 58 as closely as possible. For example, air monitoring devices should be placed within 20 meters horizontal of the reference instrument, positioned such that the sample air inlets for the air monitoring devices are within a height of ± 1 meter vertically of the sample air inlets of the reference instrument, and placed as far as possible from any obstructions (e.g., trees, walls) to minimize spatial and wind turbulence effects on sample collection. (Source: https://www.epa.gov/air-sensor-toolbox/air-sensor-performance-targets-and-testing-protocols)

Correction
The adjustments made to measurement data to more closely match the data collected by a reference monitor.

CSV file
A comma separated value file is a delimited text file that separates values using commas. CSV files are commonly used to share air quality data because they are easily opened in software like Microsoft excel and coding applications like Python, R, or MatLab.

D

Data archive
A separate storage location for data that is no longer being actively used.

Data management
The process of collecting, storing, and using data securely, efficiently, and cost-effectively.

Data Management System (DMS)
A collection of procedures, software, and hardware needed to acquire, process, and distribute data.

Data Quality Objectives (DQO)
Not discussed in this worksheet. Quantitative acceptance criteria for the quality and quantity of data to be collected, relative to the ultimate use of the data. (Source: https://www.epa.gov/sites/default/files/2015-06/documents/g5-final.pdf)

J

JSON file
A JavaScript object notation (JSON) is a text-based file format used to store and transfer data. JSON file formats are most commonly used to display data on a website or web-based application.

M

Metadata
A dataset or document that describes and gives context to an air quality (or other) data set. When you download or request an air quality data set it will often come with two files: the actual data, and a metadata file that describes the actual data. It may include a written description of the data set, the name and contact of the community member or organization who took the data, the time period in which the data was taken, the time zone the data is in, the location of the air quality monitor(s) if that is not already in the actual data, nearby sources, and other relevant details.

Q

Quality Assurance (QA)
Planned steps performed to manage a project and collect, assess, and review data to ensure that measurements meet the data quality needed for the monitoring objective. An example QA activity is developing a plan for air monitoring. (Source: https://www.epa.gov/sites/default/files/2015-06/documents/g5-final.pdf)

Quality Assurance Project Plan (QAPP)
Not discussed in this worksheet. A plan that describes the activities of a monitoring project involved with the acquisition of environmental information whether generated from direct measurements activities, collected from other sources, or compiled from computerized databases and information systems. The QAPP documents the results of a project's technical planning process, providing in one place a clear, concise, and complete plan for the environmental data operation and its quality objectives and identifying key project personnel. The QAPP communicates the specifications for implementation of the project design to all parties and ensures that the quality objectives are achieved for the project. (Source: https://www.epa.gov/sites/default/files/2015-06/documents/g5-final.pdf)

Quality Control (QC)
Steps performed to limit error from instruments or in measurements during a project. Examples of QC activities include collocation, correction of data, maintenance, automatic data checks, and data review. (Source: https://www.epa.gov/sites/default/files/2015-06/documents/g5-final.pdf)

R

Raw data
Data that has not been edited, processed, or quality controlled.

Real-time data
Data that is delivered to a system immediately after it was collected.

S

Standard Operating Procedures (SOPs)
A set of written instructions that detail the one-time and repetitive activities to be conducted or followed within an organization. An SOP provides individuals with the information to perform a job properly, which facilitates consistent conformance to technical and quality system requirements and supports data quality. (Source: https://www.epa.gov/sites/production/files/2015-06/documents/g6-final.pdf)

Acknowledgements

Developed by Ron Evans (EPA Office of Air Quality Planning and Standards) and TD Enviro staff. We appreciate the assistance of several EPA staff and others who provided reviews of this document.

Disclaimer: *This guide is a resource for community air monitoring and does not necessarily reflect U.S. EPA policies.*

Resources

EPA's Enhanced Air Sensor Guidebook

> Guide for users in planning and collecting air quality measurements using air sensors. See Section 3.7 that covers collecting, quality controlling, and managing data.
>
> **Link**: https://cfpub.epa.gov/si/si_public_record_report.cfm?Lab=CEMM&dirEntryId=356426

U.S. EPA's Air Sensor Toolbox

> Information and resources for topics related to air sensors; includes links to other organizations and resources that sensor users may find helpful.
>
> **Link**: https://www.epa.gov/air-sensor-toolbox

South Coast Air Quality Management District

> Guidebook for community organizations that covers planning for monitoring using sensors; sensor deployment, use, and maintenance; and data handling, interpretation, and communication. See the section on Understanding Your Data.
>
> **Link**: http://www.aqmd.gov/aq-spec/special-projects/star-grant

U.S. EPA's Handbook for Citizen Science Quality Assurance and Documentation

> Handbook that covers common expectations for quality assurance and documentation and best management practices for organizations that train and use volunteers in the collection of environmental data.
>
> **Link**: https://www.epa.gov/participatory-science/quality-assurance-handbook-and-toolkit-participatory-science-projects

U.S. EPA Ambient Monitoring Technology Information Center

> Contains technical information regarding ambient air monitoring programs, including the networks of state and local air monitoring stations, monitoring methods, and quality assurance/quality control procedures.
>
> **Link**: https://www.epa.gov/amtic