# **Air Quality Data Processing Vocabulary with References**

## **Environmental Monitoring Terms**

**AQI (Air Quality Index)** - A standardized scale (0-500) that communicates daily air quality levels and associated health risks. Values are calculated from measurements of major air pollutants.  
 *Reference: U.S. EPA. (2018). Technical Assistance Document for the Reporting of Daily Air Quality – the Air Quality Index (AQI). EPA-454/B-18-007.*

**PM2.5 (Particulate Matter 2.5)** - Fine particles with diameters of 2.5 micrometers or smaller. These particles can penetrate deep into lungs and bloodstream, posing significant health risks.  
 *Reference: WHO. (2021). WHO global air quality guidelines: particulate matter (PM2.5 and PM10), ozone, nitrogen dioxide, sulfur dioxide and carbon monoxide. World Health Organization.*

**PM10 (Particulate Matter 10)** - Coarse particles with diameters of 10 micrometers or smaller, including dust, pollen, and mold. Less dangerous than PM2.5 but still capable of causing respiratory issues.  
 *Reference: U.S. EPA. (2019). Integrated Science Assessment (ISA) for Particulate Matter (Final Report). EPA/600/R-19/188.*

**IAQI (Individual Air Quality Index)** - Separate index values calculated for each measured pollutant, showing the air quality impact of specific contaminants.  
 *Reference: U.S. EPA. (2018). Technical Assistance Document for the Reporting of Daily Air Quality – the Air Quality Index (AQI). EPA-454/B-18-007.*

**Dominant Pollutant** - The pollutant with the highest individual index value, which determines the overall AQI for that location and time.  
 *Reference: U.S. EPA. (2018). Technical Assistance Document for the Reporting of Daily Air Quality – the Air Quality Index (AQI). EPA-454/B-18-007.*

**Ground-level Ozone (O3)** - A harmful air pollutant formed when nitrogen oxides and volatile organic compounds react in sunlight. Different from protective ozone in the upper atmosphere.  
 *Reference: U.S. EPA. (2020). Integrated Science Assessment (ISA) for Ozone and Related Photochemical Oxidants (Final Report). EPA/600/R-20/012.*

**Nitrogen Dioxide (NO2)** - A reddish-brown gas primarily produced by vehicle emissions and industrial processes. Indicator of traffic-related air pollution.  
 *Reference: U.S. EPA. (2016). Integrated Science Assessment (ISA) for Oxides of Nitrogen–Health Criteria (Final Report). EPA/600/R-15/068.*

**Sulfur Dioxide (SO2)** - A gas produced mainly by burning fossil fuels containing sulfur, particularly in power plants and industrial facilities.  
 *Reference: U.S. EPA. (2017). Integrated Science Assessment (ISA) for Sulfur Oxides–Health Criteria (Final Report). EPA/600/R-17/451.*

**Carbon Monoxide (CO)** - A colorless, odorless gas produced by incomplete combustion of carbon-containing materials, primarily vehicle exhaust.  
 *Reference: U.S. EPA. (2010). Integrated Science Assessment (ISA) for Carbon Monoxide (Final Report). EPA/600/R-09/019F.*

**UVI (UV Index)** - A measure of ultraviolet radiation intensity that affects human health, ranging from 0 (minimal risk) to 11+ (extreme risk).  
 *Reference: WHO. (2002). Global Solar UV Index: A Practical Guide. World Health Organization, World Meteorological Organization, United Nations Environment Programme, International Commission on Non-Ionizing Radiation Protection.*

## **Technical Data Terms**

**API (Application Programming Interface)** - A set of protocols and tools that allows different software applications to communicate and exchange data.  
 *Reference: Fielding, R. T. (2000). Architectural Styles and the Design of Network-based Software Architectures. Doctoral dissertation, University of California, Irvine.*

**JSON (JavaScript Object Notation)** - A lightweight, text-based data interchange format that's easy for humans to read and write, and easy for machines to parse and generate.  
 *Reference: Crockford, D. (2006). The application/json Media Type for JavaScript Object Notation (JSON). RFC 4627, Internet Engineering Task Force.*

**Nested JSON** - JSON structures that contain objects within objects or arrays within arrays, creating hierarchical data relationships.  
 *Reference: Bray, T. (Ed.). (2017). The JavaScript Object Notation (JSON) Data Interchange Format. RFC 8259, Internet Engineering Task Force.*

**Payload** - The actual data content transmitted in an API response, excluding headers and metadata.  
 *Reference: Fielding, R., Gettys, J., Mogul, J., Frystyk, H., Masinter, L., Leach, P., & Berners-Lee, T. (1999). Hypertext Transfer Protocol -- HTTP/1.1. RFC 2616, Internet Engineering Task Force.*

**Schema** - The structure and organization of data, defining what fields are present and their data types.  
 *Reference: Wright, A., Andrews, H., Hutton, B., & Dennis, G. (2022). JSON Schema: A Media Type for Describing JSON Documents. Internet-Draft, Internet Engineering Task Force.*

**Time-series Data** - Data points indexed in chronological order, such as pollution measurements taken at regular intervals over time.  
 *Reference: Box, G. E., Jenkins, G. M., Reinsel, G. C., & Ljung, G. M. (2015). Time Series Analysis: Forecasting and Control. John Wiley & Sons.*

**Telemetry** - The automatic collection and transmission of data from remote sensors or monitoring devices.  
 *Reference: IEEE. (2017). IEEE Standard for Internet of Things (IoT) - Data-Link Layer Specifications. IEEE 802.15.4e-2012.*

## **Software Engineering Terms**

**Data Normalization** - The process of organizing and standardizing data to eliminate redundancy and ensure consistency across different sources.  
 *Reference: Codd, E. F. (1970). A relational model of data for large shared data banks. Communications of the ACM, 13(6), 377-387.*

**Semantic Enhancement** - Adding meaning and context to raw data by applying domain knowledge and interpretive frameworks.  
 *Reference: Berners-Lee, T., Hendler, J., & Lassila, O. (2001). The semantic web. Scientific American, 284(5), 34-43.*

**Abstraction Layer** - A software component that hides complex implementation details and provides a simplified interface for data access.  
 *Reference: Parnas, D. L. (1972). On the criteria to be used in decomposing systems into modules. Communications of the ACM, 15(12), 1053-1058.*

**Modular Architecture** - A design approach that organizes software into separate, interchangeable components that can be developed and maintained independently.  
 *Reference: Baldwin, C. Y., & Clark, K. B. (2000). Design Rules: The Power of Modularity. MIT Press.*

**Graceful Degradation** - A system's ability to continue operating with reduced functionality when some components fail or are unavailable.  
 *Reference: Laprie, J. C. (1985). Dependable computing and fault tolerance: concepts and terminology. In Proceedings 15th International Symposium on Fault-Tolerant Computing (pp. 2-11). IEEE.*

**Source Attribution** - The practice of maintaining references to original data sources to ensure traceability and credibility.  
 *Reference: ISO/IEC 25012:2008. Software engineering -- Software product Quality Requirements and Evaluation (SQuaRE) -- Data quality model.*

**Error Handling** - Programming techniques that manage and respond to unexpected conditions or invalid data without causing system failures.  
 *Reference: Goodenough, J. B. (1975). Exception handling: issues and a proposed notation. Communications of the ACM, 18(12), 683-696.*

## **Data Analysis Terms**

**Pandas** - A Python library providing data structures and analysis tools for handling structured data, particularly useful for time-series analysis.  
 *Reference: McKinney, W. (2010). Data structures for statistical computing in Python. In Proceedings of the 9th Python in Science Conference (Vol. 445, pp. 51-56).*

**DataFrame** - A two-dimensional data structure in pandas, similar to a spreadsheet, with labeled rows and columns.  
 *Reference: McKinney, W. (2017). Python for Data Analysis: Data Wrangling with Pandas, NumPy, and IPython. O'Reilly Media.*

**Trend Analysis** - The practice of collecting and analyzing data over time to identify patterns, directions, or changes.  
 *Reference: Kendall, M. G. (1975). Rank Correlation Methods. Charles Griffin & Company.*

**Forecast Data** - Predicted future values based on current conditions and historical patterns.  
 *Reference: Hyndman, R. J., & Athanasopoulos, G. (2018). Forecasting: Principles and Practice. OTexts.*

**Temporal Integration** - The process of combining time-based data from different sources or time periods into coherent analytical datasets.  
 *Reference: Chen, L., Özsu, M. T., & Oria, V. (2005). Robust and fast similarity search for moving object trajectories. In Proceedings of the 2005 ACM SIGMOD international conference on Management of data (pp. 491-502).*

**Statistical Aggregation** - Mathematical operations (like average, minimum, maximum) applied to groups of data points to summarize information.  
 *Reference: Gray, J., Chaudhuri, S., Bosworth, A., Layman, A., Reichart, D., Venkatrao, M., ... & Pirahesh, H. (1997). Data cube: A relational aggregation operator generalizing group-by, cross-tab, and sub-totals. Data Mining and Knowledge Discovery, 1(1), 29-53.*

## **Regulatory and Health Terms**

**EPA (Environmental Protection Agency)** - The U.S. federal agency responsible for environmental protection and regulation, including air quality standards.  
 *Reference: U.S. EPA. (2023). About EPA. Retrieved from* [*https://www.epa.gov/aboutepa*](https://www.epa.gov/aboutepa)

**Health Categories** - EPA-defined classifications that translate AQI values into health risk levels:

* Good (0-50): Air quality is satisfactory
* Moderate (51-100): Acceptable for most people
* Unhealthy for Sensitive Groups (101-150): May affect sensitive individuals
* Unhealthy (151-200): Everyone may experience health effects
* Very Unhealthy (201-300): Health warnings for everyone
* Hazardous (301-500): Emergency conditions

*Reference: U.S. EPA. (2018). AirNow - Air Quality Index (AQI) Basics. Retrieved from* [*https://www.airnow.gov/aqi/aqi-basics/*](https://www.airnow.gov/aqi/aqi-basics/)

**Sensitive Groups** - Populations at higher risk from air pollution, including children, elderly adults, people with asthma, heart disease, or lung disease.  
 *Reference: U.S. EPA. (2019). Integrated Science Assessment (ISA) for Particulate Matter (Final Report). EPA/600/R-19/188.*

**Regulatory Framework** - The set of rules, standards, and guidelines established by government agencies to monitor and control environmental conditions.  
 *Reference: Clean Air Act, 42 U.S.C. §7401 et seq. (1970). United States Code.*

**Threshold Values** - Specific numerical limits that trigger health warnings or regulatory actions when exceeded.  
 *Reference: U.S. EPA. (2015). National Ambient Air Quality Standards (NAAQS) Table. 40 CFR Part 50.*

## **System Architecture Terms**

**Production Environment** - The live system where software runs for actual users, as opposed to development or testing environments.  
 *Reference: Fowler, M., & Foemmel, M. (2006). Continuous integration. Thought-Works. Retrieved from* [*https://martinfowler.com/articles/continuousIntegration.html*](https://martinfowler.com/articles/continuousIntegration.html)

**API Integration** - The process of connecting different software systems through their APIs to share data and functionality.  
 *Reference: Richardson, L., & Ruby, S. (2008). RESTful Web Services. O'Reilly Media.*

**Data Pipeline** - A series of processing steps that move data from source systems through transformations to final destinations.  
 *Reference: Kimball, R., & Ross, M. (2013). The Data Warehouse Toolkit: The Definitive Guide to Dimensional Modeling. John Wiley & Sons.*

**Extensible Design** - Software architecture that can be easily modified or expanded to accommodate new requirements without major restructuring.  
 *Reference: Gamma, E., Helm, R., Johnson, R., & Vlissides, J. (1995). Design Patterns: Elements of Reusable Object-Oriented Software. Addison-Wesley.*

**Multi-vendor Support** - The ability of a system to work with products or services from different suppliers or providers.  
 *Reference: IEEE. (2017). IEEE Standard for System, Software, and Hardware Verification and Validation. IEEE 1012-2016.*

**Real-time Processing** - The ability to process and respond to data immediately as it arrives, without significant delay.  
 *Reference: Stankovic, J. A. (1988). Misconceptions about real-time computing: a serious problem for next-generation systems. Computer, 21(10), 10-19.*