



# City of Albuquerque Information Technology Services Division Data Management

## Data.cabq.gov Core Metadata Requirements

### Contact Information

*Who is the contact for this dataset? The contact will be the City employee who is accountable for the data provided in this dataset and can act as front-line support in the event of any questions about the data.*

<b>Name</b>	Fabian Macias
<b>Department/Division</b>	Albuquerque Environmental Health Department / Air Quality Monitoring and Assurance Sections
<b>Phone</b>	505-768-1969
<b>Email</b>	FMacias@cabq.gov

### What Does this Dataset Describe?

*What is the name of this dataset? How should a user identify this dataset in any communication with contact above? Provide a shorter description of the Dataset that can act as a one-line summary of the dataset when dealing with stakeholders. Provide a longer description of the data that can be readily understood by non-technical users.*

<b>Dataset Title</b>	Air Quality Data
<b>Short Description</b>	Air quality measurements are <i>polled</i> hourly from our air monitoring stations
<b>Full Non-Technical Description</b>	
The City of Albuquerque Environmental Health Department Air Quality Division measures ambient air quality conditions for the following criteria pollutants: Carbon monoxide (CO); Nitrogen dioxide (NO <sub>2</sub> ); Ozone (O <sub>3</sub> ); Coarse Particulate Matter (PM <sub>10</sub> ); Fine Particulate Matter (PM <sub>2.5</sub> ), Sulfur Dioxide (SO <sub>2</sub> ). Data provided is not quality assured and use of data is dependent on the end users knowledge.	

### How Should this Dataset be Cited?

*How should external sources refer to this dataset in publications or documentation? Often this will simply be the URL and the date retrieved.*

<http://data.cabq.gov/airquality/aqindex/daily/12071610.NM2>  
 12 year 07 month 16 day 10 hour .NM2

<http://data.cabq.gov/airquality/aqindex/history/12071623.NM2>  
 12 year 07 month 16 day 23 hour .NM2

Hourly files are placed in the daily directory between 01:15 and 23:15. All daily files will be deleted from the daily directory at 00:15. The 23<sup>rd</sup> hour file is placed in the history directory at 00:15.

### **Does the Dataset Reflect a Particular Time Period?**

*Provide any date restrictions that may affect the validity of the data. The table fields are defined as follows:*

<i>Field</i>	<i>Definition</i>
<i>Start Date</i>	<i>Start date of the time period within which this data falls. Format: MM/DD/YYYY HH:MM:SS.</i>
<i>End Date</i>	<i>End date of the time period within which this data falls. Format: MM/DD/YYYY HH:MM:SS.</i>
<i>Dataset Refresh Interval</i>	<i>Time period between Dataset refreshes. Format: “nn [seconds/minutes/hours/days/weeks/months/years]” or the word “Static” if never refreshed.</i>
<i>Data Expiration Date</i>	<i>Date after which the data must be considered stale and no longer of sufficient utility (fit-for-purpose). Format: MM/DD/YYYY HH:MM:SS.</i>
<i>Dataset Review Date</i>	<i>Date after which this dataset will be reviewed by the City for utility (fit-for-purpose) and usage. Format: MM/DD/YYYY HH:MM:SS.</i>
<i>Comments</i>	<i>Specific comments related to any time-specific features of this dataset.</i>

<b>Start Date</b>	08/22/12
<b>End Date</b>	On going

Dataset Refresh Interval	Hourly
Dataset Expiration Date	None.
Dataset Review Date	Data provided is not quality assured and use of data is dependent on the end users knowledge.
<b>Comments</b>	
Hourly files are placed in the daily directory between 01:15 and 23:15. All daily files will be deleted from the daily directory at 00:15. The 23 <sup>rd</sup> hour file is placed in the history directory at 00:15.	

### **Dataset Definition/Format**

*Provide a field-by-field breakdown and definition of each record. This section acts as the formal data dictionary for an individual record.*

Field Name	Format	Description
BEGIN_FILE		Indicates the beginning of the file
FORMATE_VERSION		Indicates the file format version being used. Currently we are using format version 2
AGENCY		Indicates the agency reporting the data. City of Albuquerque is NM2.
FILENAME		071610.NM2 07 month 16 day 10 hour .NM2 because State of NM is defined as NM1 in the EPA AirNow system.  Please note that the year is not part of the filename in the filename field.
DATA_VERSION		Contains the year, month, day, hour and minute the data was captured
TZONE		Time Zone for where the data is captured, monitoring network stations stay on standard time. MST, 7
BEGIN_GROUP		Indicates the beginning of the data block, repeated for each data block (repeating)
VARIABLE		Indicates the type of data being reported in the data block. OZONE, PM10, PM2.5, NO2, SO2
DATA_TYPE		POINT
MEASUREMENT_TYPE		SAMPLE

CHARACTERISTIC		OBSERVED
START_DTG		Start Date Group, Indicates the date and start time of the data in this group. yyyyymmddhhhh
END_DTG		End Date Group, Indicates the data and end time for the data in this group. yyyyymmddhhhh
INTERVAL		Indicates the interval of the reading in minutes.
START_REF		Start hour
NUMSTEPS		Number of readings that are included in this data group. Each hour additional readings are added to the file so at the end of 10 hours there will be 10 readings through the 24-hours of the day.
AVG_TIME		The average time between readings.
UNITS		The unit of measurement used in this data group. PPB = Parts Per Billion UG/M3 = micrograms per metric cubed PPM = Parts Per Million
STATION		The number of the station that is reporting the data in this data group.
BEGIN_DATA		Indicates the beginning of the data being reported (repeating)
Date Row		There is one row of data for each station reporting data.
Site Name		Name of the site where the reading is taken
EPA Site Code		The EPA Site number where the reading is taken
Reading		The reading, there will be the same number of readings as there are NUMSTEPS. They are reported in order with the 1 <sup>st</sup> hourly value first and the 10 <sup>th</sup> hourly value last through the 24-hours of the day.
Quality of Data Row		There is one quality row for each data row.
Site Name		Name of the site where the reading is taken
EPA Site Code		The EPA Site number where the reading is taken
Quality Indicator		For each reading in the row above there is an indicated as to whether the reading is a “G”ood reading or a “B”ad reading. This lets the AirNow system know if the data is questionable. IF “B”

		data then the corresponding reading in the row above is not used in the indexing process. All “B” data should be ignored.
END_DATA		Indicates the end of the data being reported (repeating)
END_GROUP		Indicates the end of the data block, repeated for each data block (repeating)
END_FILE		Indicates the end of the file

### ***Dataset Technical Description***

*Provide a technical description of the dataset. This should be a complete technical description aimed at developers and expert users who need to understand the scope, strengths and limitations of the dataset.*

This data is provided in a comma delimited text format with repeating sections. File data and Group data are in a paired, field, data combination where the actual readings and quality of the readings in on individual lines with data elements separated by a comma.

See example below through the 10:00 hour. Note data is 24/7.

BEGIN\_FILE

FORMAT\_VERSION,2  
AGENCY,NM2  
FILENAME,071610.NM2  
DATA\_VERSION,201207161004  
TZONE,MST,7

BEGIN\_GROUP

VARIABLE,OZONE  
DATA\_TYPE,POINT  
MEASUREMENT\_TYPE,SAMPLE  
CHARACTERISTIC,OBSERVED  
START\_DTG,201207160000  
END\_DTG,201207160959  
INTERVAL,60  
START\_REF,0  
NUMSTEPS,10  
AVG\_TIME,60  
UNITS,PPB  
STATIONS,7

Site Name

EPA Site Code

1<sup>st</sup> Hourly Value

10<sup>th</sup> Hourly Value

BEGIN\_DATA

TRAMWAY ,350011012,43,39,52,58,59,55,52,59,62,64  
TRAMWAY ,350011012, G, G, G, G, G, G, G, G, G, G  
NORTH VALLEY ,350011013,35,32,36,45,43,32,35,51,55,55  
NORTH VALLEY ,350011013, G, G, G, G, G, G, G, G, G, G  
DEL NORTE HIGH SCH ,350010023,40,44,39,50,48,38,35,50,58,62  
DEL NORTE HIGH SCH ,350010023, G, G, G, G, G, G, G, G, G, G

.f

F8

END\_DATA

END\_GROUP

BEGIN\_GROUP

VARIABLE,PM10  
DATA\_TYPE,POINT  
MEASUREMENT\_TYPE,SAMPLE  
CHARACTERISTIC,OBSERVED  
START\_DTG,201207160000  
END\_DTG,201207160959  
INTERVAL,60  
START\_REF,0  
NUMSTEPS,10  
AVG\_TIME,60  
UNITS,UG/M3  
STATIONS,3

BEGIN\_DATA

NORTH VALLEY ,350011013,22.9,21.9,12.1,11.6,13.7,23.3,22.5,12.3,8.5,11.3  
NORTH VALLEY ,350011013, G, G, G, G, G, G, G, G, G, G

END\_DATA

END\_GROUP

END\_FILE

## **Dataset Assumptions**

*What technical and business assumptions are implied in the creation of this dataset? Examples could include the way in which a salary figure was calculated or data that was omitted for a specific reason.*

Not all data is good. Bad data should be ignored.

The data represent the best information available. However, some values may be absent due to incomplete reporting, and some values subsequently may be changed due to quality assurance activities. The database is updated hourly.

Readers and data users are cautioned not to infer a qualitative ranking order of geographic areas based on AirData maps, charts, or reports. Air pollution levels measured in the vicinity of a particular monitoring site may not be representative of the prevailing air quality of a county or urban area. Pollutants emitted from a particular source may have little impact on the immediate geographic area, and the amount of pollutants emitted does not indicate whether the source is complying with applicable regulations.

## **Who Produced the Dataset?**

*Which department in the City produced this dataset? Note that this might not always be the data owner. An example of this could be a dataset that ITSD produced on behalf of EHD who owned the data.*

Air Quality Monitoring and Assurance Section, Albuquerque Environmental Health Department as required by the US EPA.

## **Who Owns the Data?**

*Where did this data originate? Who owns the data used in this dataset? Note that this might not always be the dataset producer. An example of this could be a dataset that ITSD produced on behalf of EHD who owned the data.*

Air Quality Monitoring and Assurance Section, Albuquerque Environmental Health Department as required by the US EPA.

## **Why was the Dataset Created?**

*All datasets should have an explicit reason for existence and should, somehow, have value to someone. What is the perceived value that this dataset will bring?*

Air Quality Monitoring and Assurance Section, Albuquerque Environmental Health Department as required by the US EPA for the purposes of reporting data to the EPA AirNow data management center.

### **How was the Dataset Created?**

*How was this dataset produced? Was it a manual process? An automated process? What were the main IT systems involved in producing this dataset?*

This data is extracted from the Air Quality system and automatically FTP'd to data.cabq.gov by the Environmental Health Department.

Not a manual process.

Agilaire EDAS System.

### **What Similar or Related Data Should the User be Aware of?**

*Are there any other datasets available that may contain related or similar information? Might there be situations in which these other datasets might be a better alternative?*

US EPA AirNow Contractor System via Sonoma Technology

### **How Reliable are the Data?**

*Are there any concerns about overall data reliability? Are there any data problems that the user needs to be aware of? Are there any constraints with data accuracy? What levels of confidence with this dataset could the user reasonably assume?*

Data is very reliable and comes from the same source that is sent to AirNow.

### **How Well Have the Observations Been Checked?**

*What quality assurance steps have been performed? Sometimes, a third-party verification/audit process may also be required. If so, provide the name of the third-party who performed the verification.*

Data is not verified by the department prior to submittal. Each reading has a "G"ood or "B"ad indicator that indicates if the reading is reliable and therefore is ok to use.

## ***Are there Legal Restrictions on the Access or Use of the Data?***

*Are there any specific legal or compliance restrictions for this data? How might this affect the way in which end users might access and use this data?*

The data represent the best information available. However, some values may be absent due to incomplete reporting, and some values subsequently may be changed due to quality assurance activities. The database is updated hourly.

Readers and data users are cautioned not to infer a qualitative ranking order of geographic areas based on AirData maps, charts, or reports. Air pollution levels measured in the vicinity of a particular monitoring site may not be representative of the prevailing air quality of a county or urban area. Pollutants emitted from a particular source may have little impact on the immediate geographic area, and the amount of pollutants emitted does not indicate whether the source is complying with applicable regulations.

## ***Legal Disclaimer***

The data represent the best information available. However, some values may be absent due to incomplete reporting, and some values subsequently may be changed due to quality assurance activities. The database is updated hourly.

Readers and data users are cautioned not to infer a qualitative ranking order of geographic areas based on AirData maps, charts, or reports. Air pollution levels measured in the vicinity of a particular monitoring site may not be representative of the prevailing air quality of a county or urban area. Pollutants emitted from a particular source may have little impact on the immediate geographic area, and the amount of pollutants emitted does not indicate whether the source is complying with applicable regulations.

The City's standard copyright, disclaimers and legal statements may be found at <http://www.cabq.gov/about/legal>. The City data policy governing data.cabq.gov may be found at <http://data.cabq.gov/policy/>.