TECHNICAL MEMORANDUM

TO:	Madeleine Strum, Dave Shelow, Laurie Trinca, and Elizabeth Mannshardt Delivery Order Managers U.S. EPA Office of Air Quality Planning and Standards
FROM:	Regi Oommen, Karla Faught, Jaime Hauser, Bebhinn Do, and Steve Mendenhall, Eastern Research Group, Inc. (ERG)
DATE:	April 4, 2016
SUBJECT:	Phase X of the Compilation and Quality Assurance (QA) Summary Report for the Ambient Monitoring Archive for Hazardous Air Pollutants (HAPs)

1.0   INTRODUCTION

The purpose of this memorandum is to summarize improvements, modifications, and additional data incorporated into the development of EPA's Phase X Ambient Monitoring Archive (Archive). Under a prior Delivery Order, Eastern Research Group, Inc. (ERG) prepared Phase IX, which comprised of hazardous air pollutant (HAP) and non-HAP air toxics monitoring data, as well as criteria pollutant, greenhouse gas pollutants, and meteorological data, collected from numerous federal, state, local, and tribal agencies from 1973 to 2013.

ERG was tasked to develop Phase X by updating the Archive to 2014, incorporate additional data not in the Archive, and provide general maintenance/cleanup of the Phase IX Archive. All work was performed under EPA Contract No. EP-D-14-030, Delivery Orders 00-20 and 00-25 entitled "Report Development – Data Characterization."

2.0   BACKGROUND INFORMATION

EPA first developed a master HAP Archive in 2001 to consolidate HAP measurements that had been collected by various state and local agencies. At that time, there was no guidance or requirement that HAP data be submitted to EPA's Air Quality System (AQS). Thus, a concerted effort was made to gather these data, provide some level of quality assurance, and standardize the information for the development of a master database, which was called the Phase I Archive.

1

During that time, EPA also began implementing its Urban Air Toxics Strategy, which was finalized in 1999. In response, a number of EPA and state/local-sponsored ambient HAP monitoring initiatives began. As such, EPA regularly updated and appended the Archive to include new measurements. Over time, EPA began requiring that EPA-sponsored monitoring initiatives submit their data to AQS. Table 2-1 presents a summary of the HAP Archive's timeline.

Table 2-1. Summary of Prior HAP Archives

Phase	Year Completed	Coverage Years
I	2001	1990-2000
II	2003	1990-2001
III	2004	1990-2002
IV	2005	1990-2003
V	2007	1973-2005
VI	2009	1973-2008
VII	2013 (Feb)	1973-2010
VIII	2013 (Oct)	1973-2012
IX	2015	1973-2013

EPA previously developed the Phase IX Archive in January 2015, which contained 46 million HAP records from 1973 to 2013. The Phase IX Archive was the fourth successful update built upon the re-engineered system that was developed for Phase VI effort (Summer 2009). This re-engineering allowed EPA to simplify future updates. For example, data records were housed as their native sample durations from AQS, such as hourly measurements. Another update was the identification of non-detect data measurement records which were incorrectly substituted as one-half the method detection limit (MDL) value.

For the Phase X update, EPA requested that ERG:

- Retrieve 1990-2014 ambient HAP data from EPA's Air Quality Subsystem (AQS);

- Retrieve non-HAP species, criteria pollutant, greenhouse gas (GHG), and meteorological data from 1990-2014;

- Incorporate additional datasets, if available;

- Incorporate quality-assured National Air Toxics Trends Stations (NATTS) Network Assessment data collected;

- Perform general housekeeping/cleanup of the new data retrieved from AQS;

- Standardize all descriptions (pollutant names, sampling methodology, etc.) and data fields;

- Assign and QA "Sampling Frequency Code" based on sample dates;

- Assure each data record has a corresponding Method Detection Limit (MDL);

- Identify sample values which were incorrectly entered as ½ MDL;

- Identify sample values below MDL;

- Identify duplicate data reported in AQS from the reporting entity;

- Identify and maintain data records which have been invalidated;

- Standardize all reported concentrations to local conditions, where applicable; and

- Prepare flattened data files for posting to EPA's website.

The Ambient Monitoring Archive consists of five data types: 1) Group 1 consists of HAPs; 2) Group consists of Non-HAPs; 3) Group 3 consists of criteria pollutants; 4) Group 4 consists of GHGs; and 5) Group 5 consists of meteorological data. The focus of this memorandum is on the HAP records (Group 1).

3.0   AMA DATA SOURCES

For the Phase X Archive, there were twelve primary data sources used. Information about each Data Source is presented in Sections 3.1 – 3.13.

3.1   AQS Data

AQS is EPA's official repository of ambient monitoring data. Users of AQS can upload and download data using standard or ad-hoc queries. Although not required for most air toxic programs, state and local agencies are encouraged to upload their ambient monitoring data to AQS. In contrast, data generated from EPA's National Air Toxics Trends System (NATTS) network, the Urban Air Toxics Monitoring Program (UATMP), and from community-scale air toxics monitoring grant sites are required to submit data to AQS. Data from 2014 were retrieved. Additionally, data from 1990-2013 were also retrieved to replace Phase IX database records since the Archive was last updated (January 2015). The retrieval date was November 1, 2015.

More information can be found at: https://www.epa.gov/aqs. Subsequent data pulls were performed in December 2015, January 2016, February 2016, and March 2016 as EPA was alerted to new data being added into AQS.

3.2   National Air Toxics Trends System (NATTS) Network Data Review

In Summer 2014, EPA prepared an interim status report on data reporting for the NATTS Network. As per the requirements of the NATTS Network, data must be submitted to AQS no later than 120 days after a calendar quarter. During this data review, a number of concentrations reported to AQS were identified as incorrect. Additionally, certain datasets were identified as missing from AQS, and were obtained from the NATTS Operators. The corrected and missing data obtained by EPA were formatted for inclusion into this Archive. EPA is currently updating the NATTS Network Assessment (to be completed Summer 2016), and again performed data review with particular focus on the 2011 through 2014 sampling years. Concentrations that were missing or suspicious were reported to the NATTS Operators, and the missing/incorrect data were updated for inclusion in the Archive. More information can be found at: https://www3.epa.gov/ttn/amtic/natts.html.

3.3   School Air Toxics, Phase 2

In 2009, as part of a new air toxics monitoring initiative, EPA, state and local air pollution control agencies monitored the outdoor air around schools for pollutants known as air toxics. EPA selected schools after evaluating a number of factors including results from an EPA computer modeling analysis, the mix of pollution sources near the schools, results from an analysis conducted for a recent newspaper series on air toxics at schools, and information from state and local air pollution agencies. Phase 1 Sampling took place in 2009-2010 in 59 schools, while Phase 2 Sampling in 2010-2011 took place in 22 schools. Nearly all of the data resides in AQS, with the exception of special VOC measurements taken at two schools: Enterprise High School in Enterprise, MS and Temple Elementary is Diboll, TX. These data were retrieved by EPA and formatted from inclusion into this Archive. More information can be found at: https://www3.epa.gov/air/sat/.

3.4   XAct Monitoring Data

U.S. EPA purchased XAct Monitoring Measurement Systems as a result of School Air Toxics Monitoring. The purpose of these continuous, multi-metal measurement systems is to help EPA, state, and local air agencies target and identify source characterization signatures of HAP metal-emitting facilities. The State of Oregon's Department of Environmental Quality (ODEQ) used XAct in a small 2011 study. Measurements data were sent by ODEQ to EPA and were processed for this Archive. After this study, EPA Region V conducted several monitoring campaigns from 2012 to 2014 in Illinois, Indiana, and Michigan using XAct for targeting of specific sources.

3.5   NADP Data

The National Acid Deposition Program consists of multiple deposition monitoring networks, such as: 1) the Atmospheric Integrated Research Monitoring Network (AIRMon); 2) the Ammonia Monitoring Network (AMON); 3) the Mercury Deposition Network (MDN); 4) the Atmospheric Mercury Network (AMNet); and the 5) National Trends Network (NTN). Data from the above networks were downloaded or sent to EPA via request, and processed from http://nadp.sws.uiuc.edu/data/.

3.6   Phase V Database

The Phase V Database consisted of over nine million daily concentration records for HAPs. Initial compilation of this air toxics archive began in the mid-1990s, consisting of datasets from a number of state and local agencies. Many of these datasets were eventually placed into AQS, or were subsequently deleted. A small portion of Phase V data records remain in the Archive, as they are not in EPA's AQS.

3.7   Sublette County, WY

Ambient HAP monitoring was conducted by the Wyoming Department of Environmental Protection (WY DEP). Fourteen monitoring sites were placed near oil and gas wells for a 1-year study from February 2009. Over 37,000 HAP concentrations were formatted for upload for the Archive. More information can be found at:
http://www.sublettewyo.com/documentcenter/view/438.

3.8   City of Ft. Worth, TX Natural Gas Air Quality Study

In 2010, the City of Ft. Worth, TX Department of Environmental Management (DEM) conducted a natural gas study within the city boundaries to characterize concentrations near natural gas wells. Under an agreement between DEM and EPA, the data from this study can be used by EPA for data analysis. During this two-month study, over 14,000 concentrations were generated at eight monitoring sites. More information can be found at: http://fortworthtexas.gov/uploadedFiles/Gas_Wells/AirQualityStudy_final.pdf.

3.9   MATES Data

The South Coast Air Quality Management District (SCAQMD) sponsored air quality data characterization studies called the Multiple Air Toxics Exposure Study (MATES). MATES-II and MATES-III data were obtained by EPA from SCAQMD. More information can be found at: http://www.aqmd.gov/home/library/air-quality-data-studies/health-studies.

3.10   IMPROVE Data

The IMPROVE Program implements long term monitoring of the visibility in National Parks and Wilderness Areas. Agencies involved in this program include: federal (NOAA; Fish and Wildlife Service; EPA, etc.), regional (NESCAUM, MARAMA, etc.), and state (Arizona). Part of tracking changes in visibility is the measurement of air toxic metal species. Sixty pollutant species and meteorological parameters are monitored in the IMPROVE Program across 180 locations every 3 to 4 days. Before data is released to the public, all IMPROVE data must go through a 4-level QA process. Data from 1990-2013 were retrieved from the IMPROVE website (http://views.cira.colostate.edu/fed/DataWizard/Default.aspx). IMPROVE data for 2014 was conformed to be uploaded into AQS, and were not retrieved from the IMPROVE website.

3.11   Integrated Atmospheric Deposition Network (IADN) Data

The Integrated Atmospheric Deposition Network (IADN) has been in operation since 1990 under the guidance of an implementation plan signed in that year. IADN has been designed with one Master Station on each of the five Great Lakes, supplemented by a number of Satellite Stations to provide more spatial detail for deposition. The Master Stations allow the complete range of measurements made in the Network, enabling total atmospheric loading to be

determined for Semivolatile Organic Compounds (SVOCs) and trace metals. Satellite Stations only collect a portion of the measurements made at the Master Stations. IADN also estimates gas exchange of the SVOCs with the lake surfaces by using the air concentration measurements of the SVOCs at these sites in combination with water concentration measurements of the same chemicals made by other programs. U.S. data from 1991-2008 for the organic, PAH, and PCB compounds were retrieved from the IADN website (http://ec.gc.ca/data_donnees/STB-AQRD/Toxics/IADN/).

3.12   Minnesota Air Toxics Data

The Minnesota Pollution Control Agency (MNPCA) oversees a large network of air toxics monitoring stations across the state. While all of the data were uploaded to AQS, EPA was alerted about data reporting issues that occurred when reporting to AQS, such as truncation of concentrations and method detection limits. As such, MNPCA submitted their entire dataset from 2002-2014 to EPA for replacement of the data records that were in AQS for the Phase X Archive. A total of 53 sites and 51 pollutants for 664,960 records were incorporated into the Archive.

3.13   Baltimore Inner Harbor Monitoring Study

The Maryland Department of the Environment and U.S. EPA Region III oversaw a special hexavalent chromium monitoring study at six sites in the Baltimore Inner Harbor beginning in 2014. The study focused on establishing baseline air quality concentrations. A total of 1,048 records were incorporated into the Archive.

3.14   Blend/Merging of the Data

All data were uploaded into Microsoft SQL Server for pre-processing and setting data field conventions. Microsoft SQL Server is capable of handling large amounts of data, and provides a robust platform for manipulating data for QA purposes. For example, IMPROVE data are also uploaded into AQS, but EPA found discrepancies in the reported concentrations and data completeness. Thus, all IMPROVE data uploaded into AQS were "invalidated" during the blend/merge process.

SQL Server also offers the ability to create primary key constraints on tables to ensure no duplication of records. In total, there are over 1.75 billion records in the Archive. Of that, over 49 million HAP records were in the blended master database.

4.0   QA FIXES

After a preliminary assessment of the Phase IX database, the following errors and issues were identified and corrected:

- **Non-detects**. Non-detects are to be reported in AQS as zeroes, with the appropriate flag of "ND" populated. However, several sample concentration values in AQS were actually surrogate values which equated to ½ MDL. The concentrations for these records were changed to 0, and the SAMPLE_VALUE_FLAG field was populated with "ND". The following approach was used to identify these records:

    o   Step 1: Identify all records in which the concentration is one-half MDL.

    o   Step 2: By site code, pollutant, and year, summarize counts of sample dates, sample values, non-detect flags, one-half MDLs, and below MDL flags.

    o   Step 3: Identify site code, pollutant, and year combinations in which all of the below MDL flag counts is equal to the count of one-half MDLs.

    o   Step 4: For the records in Step 3, if the count of below MDL flags are equal to the counts of one-half MDL records AND if non-detects are not reported, then it was marked as being an incorrectly substituted record for non-detects.

- **Negative Concentrations**. A small number of concentrations were negative. These were converted to zero, and flagged accordingly as "ND" in the SAMPLE_VALUE_FLAG data field and as "NEG" in AQS_QUALIFIER_08 data field.

- **Method Code Fix**. Method codes were incorrect for a small number of concentration records.

- **Invalidated Data**. Through the NATTS Network Assessment, a small number of concentrations were invalidated. These concentrations were converted to null, and flagged accordingly as "AM" in the AQS_NULL_DATA_CODE data field and as "INV" in AQS_QUALIFIER_07 data field. Similarly, the State of Kentucky has recently invalidated all VOC measurements analyzed by their laboratory since 1995 due to laboratory error ("AR" code).

- **Duplicate Data**. Some agencies report concentrations of metals in both standard and local conditions for the same measurement. For these duplicates, the local condition

concentration was retained, while the standard concentration was retained, but invalidated.

- **Revised Concentrations**. Through the NATTS Network Assessment and the Urban Air Toxics Monitoring Program, small sets of data that were mistakenly entered into AQS were corrected. Additionally, outlier concentrations were identified, and in some cases, revised data were sent to EPA.

- **Sampling Frequency Code**. EPA developed a routine to calculate sampling code frequency based on the submitted sample days and days measured between samples.

- **Inconsistency of Coding**. EPA evaluated AQS coding of the following Qualifier Codes:

    o MD: This Qualifier Code is used to designate reported concentrations between the Method Detection Limit (MDL) and the Instrument Detection Limit (IDL). Concentration records were deemed "inconsistent" if they were assigned "MD", but the reported values were greater than or equal to the MDL. As such, the flag was removed.

    o MS: This Qualifier Code is used to designate reported concentrations that are substituted with ½ MDL. Concentration records were deemed "inconsistent" if they were assigned "MS", but the reported values were not equal to the ½ MDL. As such, the flag was removed.

    o ND: This Qualifier Code is used to designate reported concentrations as "no value detected". Concentration records were deemed "inconsistent" if they were assigned "ND" but the reported values were greater than zero. As such, the flag was removed.

    o PQ: This Qualifier Code is used to designate reported concentrations between the Practical Quantitation Limit (PQL) and the MDL. Concentration records were deemed "inconsistent" if they were assigned "PQ", but the reported values were less or equal to the MDL. As such, the flag was removed.

    o SQ: This Qualifier Code is used to designate reported concentrations between the Sample Quantitation Limit (PQL) and the MDL. Concentration records were deemed "inconsistent" if they were assigned "SQ", but the reported values were greater than the MDL. As such, the flag was removed.

Additionally, five Qualifier fields were populated as a result of quality assuring and compiling the database:

- AQS_QUALIFER_06: This field is reserved for data records which were identified as duplicates and were invalidated. For example, duplicates were identified if a concentration record was reported as both a local condition and a standard condition. While the parameter codes may be different, they are the same. As such, the local

condition record was retained and the standard condition was invalidated. Accordingly, "DUP" was assigned to the AQS_QUALIFIER_06 field to quickly identify these records as being invalidated.

- AQS_QUALIFER_07: This field is reserved for data records in which the sample value was invalidated as a result of the NATTS Network Assessment or through discussions with the Data Owners (e.g., the state agency). Accordingly, "INV" was assigned to the AQS_QUALIFIER_07 field to quickly identify these records as being invalidated. Additionally, "INV_MDL" was assigned to identify records in which the alternate MDL value entered was incorrect.

- AQS_QUALIFER_08: This field is reserved for data records in which the Collection Frequency Code was not populated, or if the value entered was suspected to be incorrect. Accordingly, "CF" was assigned to the AQS_QUALIFIER_08 field to quickly identify these records. The following "CF" codes were developed:

  o CF-N: Calculation frequency codes which were null, and were populated by EPA.

  o CF-I-9: Calculation frequency codes incorrectly entered as "9", which is Random.

  o CF-I-7: Calculation frequency codes incorrectly entered a "7", which is every 12 days.

  o CF-I-6: Calculation frequency codes incorrectly entered a "6", which is every 6 days.

  o CF-I-5: Calculation frequency codes incorrectly entered a "5", which is every 5 days.

  o CF-I-4: Calculation frequency codes incorrectly entered a "4", which is every 4 days.

  o CF-I-3: Calculation frequency codes incorrectly entered a "3", which is every 3 days.

  o CF-I-2: Calculation frequency codes incorrectly entered a "2", which is every 2 days.

  o CF-I-1: Calculation frequency codes incorrectly entered a "1", which is every day.

- AQS_QUALIFER_09: This field is reserved for data records in which the sample value was suspected to be populated with ½ MDL or in which the pollutant code equals 43505, which is "Acrolein – Unverified". Accordingly, "SM" and "QV" were assigned, respectively, to the AQS_QUALIFIER_09 field to quickly identify these records. For the "QV" data records, results of a recent short-term laboratory study have raised questions about the consistency and reliability of monitoring results of acrolein. Because of the uncertain accuracy of acrolein measurements, OAQPS has

changed the name of the existing acrolein parameter code in AQS (43505) to "Acrolein - Unverified" to indicate the current level of uncertainty that exists with the data already reported to AQS. Correspondingly, a new parameter code (43509) has been created in AQS for "Acrolein - Verified." Whether or not all or a subset of existing data remain in the unverified parameter code, or are re-categorized as verified and moved / reported to this new parameter code, is a choice over which each owning agency has complete discretion. Until such time as agencies evaluate their acrolein monitoring procedures and the quality of reported data, EPA recommends that already-reported data remain in the unverified method code.[1]

- AQS_QUALIFER_10: This field is reserved for data records in which the reported sample value was negative. Accordingly, "NEG" was assigned to the AQS_QUALIFIER_10 field to quickly identify these records.

5.0   DATABASE STRUCTURE

The Phase X database is designed in a relational format structure. In the relational format, the data codes from the dictionary tables are linked as foreign keys to the ambient monitoring archive table. Relational tables ensure data integrity and provide more scalability.

5.1   Ambient Monitoring Archive

The raw ambient monitoring data are housed in the Ambient Monitoring Archive data table. The data fields in the Phase X raw table are presented in Table 5-1. Primary key fields are denoted by a "*". By setting specific fields as primary keys, data records are prevented from being entered twice.

Table 5-1. Ambient Monitoring Archive Data Input Fields

Data Field	Data Description
*AMA_SITE_CODE	Ambient Monitoring Archive (AMA) Site Code
*AQS_POC	Parameter Occurrence Code
*SAMPLE_DATE	Date Sample was taken
*SAMPLE_START_TIME	Time at which sample began
*AQS_PARAMETER_CODE	Air Quality Subsystem (AQS) Pollutant Code
HAP_FLAG	Flag to identify if HAP record
NON_HAP_FLAG	Flag to identify if Non-HAP record
CRITERIA_POLL_FLAG	Flag to identify if criteria pollutant record
GHG_POLL_FLAG	Flag to identify if GHG record
MET_DATA_FLAG	Flag to identify if meteorological data record

---

[1] Found at: "Data Quality Evaluation Guidelines for Ambient Air Acrolein Measurements. OAQPS. December 17, 2010. Internet address: http://www.epa.gov/ttn/amtic/files/ambient/airtox/20101217acroleindataqualityeval.pdf

Table 5-1. Ambient Monitoring Archive Data Input Fields (Continued)

Data Field	Data Description
AQS_METHOD_CODE	Sampling Method Code
AQS_UNIT_CODE	Unit of Measure Code
AQS_SAMPLE_DURATION_CODE	Sample Duration Code
AQS_NULL_DATA_CODE	Data Qualifier code for null sample values
AQS_MONITOR_PROTOCOL_ID	AQS Protocol ID for precision and accuracy records
AQS_QUALIFIER_01	Data Qualifier code field (reserved for reporting agency)
AQS_QUALIFIER_02	Data Qualifier code field (reserved for reporting agency)
AQS_QUALIFIER_03	Data Qualifier code field (reserved for reporting agency)
AQS_QUALIFIER_04	Data Qualifier code field (reserved for reporting agency)
AQS_QUALIFIER_05	Data Qualifier code field (reserved for reporting agency)
AQS_QUALIFIER_06	Data Qualifier code field (reserved for EPA QA – See Section 4.0)
AQS_QUALIFIER_07	Data Qualifier code field (reserved for EPA QA – See Section 4.0)
AQS_QUALIFIER_08	Data Qualifier code field (reserved for EPA QA – See Section 4.0)
AQS_QUALIFIER_09	Data Qualifier code field (reserved for EPA QA – See Section 4.0)
AQS_QUALIFIER_10	Data Qualifier code field (reserved for EPA QA – See Section 4.0)
ALTERNATE_MDL	Method detection limit (MDL), in native units, if entered by Entity
UNCERTAINTY	Estimate of uncertainty surrounding the data , if available
AQS_SAMPLING_FREQUENCY_CODE	Code identifying how often the measurements were collected
SAMPLE_VALUE_REPORTED	Reported concentration value (in native units, where possible)
SAMPLE_VALUE_ADJ	Adjusted concentration value
DATA_SOURCE	Identifies the data source for the data record
COMMENT	Reserved for comments
SAMPLE_VALUE_STD	Concentration value standardized to $\mu g/m^3$
MDL_STD	MDL standardized to $\mu g/m^3$
MDL_TYPE	Identifies the source of the standardized MDL
SAMPLE_VALUE_FLAG	Identifies if the concentration record is a non-detect (Flag = "ND")
AQS_FEDERAL_MDL	Default Federal MDL value
AQS_FEDERAL_MDL_UNIT_CODE	Default Federal MDL value engineering Unit of Measure Code
BELOW_MDL_FLAG	Identifies if the non-zero sample value is less than the MDL (Flag = Y")
TRANSACTION_DATE	Date in which the data record entered the Archive
CONV_FLAG	Flag to identify concentration records which need to be converted (Flag = 1) to local conditions using local temperature and pressure.
TEMP_STD	Ambient temperature standardized to degrees Celsius for concentration records which need converted to local conditions.
TEMP_STD_UNITS	Units for the standardized temperature.
TEMP_SOURCE	Source of data for the ambient temperature
PRESS_STD	Ambient pressure standardized to mmHg for concentration records which need converted to local conditions.
PRESS_STD_UNITS	Units for the standardized pressure.
PRESS_SOURCE	Source of data for the barometric pressure
SAMPLE_VALUE_STD_LC	Concentration value standardized to $\mu g/m^3$, local conditions
LC_TYPE	Flag to indicate if the SAMPLE_VALUE_STD_LC is local "L" or standard "S".

* = primary key field

Sample values populated with a 0 indicate a non-detect, and a corresponding "ND" flag is populated in the SAMPLING_VALUE_FLAG field. Similarly, sample values with no data (or

null) indicate that the sample or the pollutant concentration was invalidated by the responsible agency or EPA for any number of reasons.

To translate the data in the Ambient Monitoring Archive, EPA developed nine data dictionary tables. These dictionaries are critical in properly describing and standardizing the raw data, and are needed for conducting accurate data analyses. AQS data dictionaries were initially retrieved from EPA's website, and provided the necessary information for AQS-submitted data. When data elements were not in the AQS data dictionaries, they were subsequently added. The nine data dictionaries are presented in Sections 5.2 through 5.11 below.

5.2   Site Information

Table 5-2 presents data fields for the HAP monitoring sites in the AMA_SITE_INFORMATION data table. The "AMA_SITE_CODE" field is the only primary key field in this data dictionary table (denoted by a "*").

Table 5-2. Site Information Data Fields

Data Field	Data Description
*AMA_SITE_CODE	Site Identifier made up of STATE_FIPS, COUNTY_FIPS, and LOCAL_SITE_ID
STATE_FIPS	State Code
COUNTY_FIPS	County Code
STATE_COUNTY_FIPS	Combination of the State and County FIPS
COUNTY_NAME	County Name
LOCAL_SITE_ID	Local Site Identifier
SITE_NAME	Name of Site, if available
CENSUS_TRACT_ID_2000	U.S. Census Tract Identifier for Year 2000
CENSUS_TRACT_ID_2010	U.S. Census Tract Identifier for Year 2010
CENSUS_TRACT_POPULATION_2000	U.S. Census Tract population for Year 2000
CENSUS_TRACT_POPULATION_2010	U.S. Census Tract population for Year 2010
ADDRESS	Monitoring Site Address
CITY	Monitoring Site City
STATE_ABBR	Monitoring Site State Abbreviation
ZIP_CODE	Monitoring Site Zip Code
EPA_REGION	EPA Region Designation
SUPPORT_AGENCY_CODE	Code for the Support Agency
SUPPORT_AGENCY	Support Agency Name
NATTS_SITE_FLAG	Identifies the site as a NATTS Site
UATMP_SITE_FLAG	Identifies the site as a UATMP Site
PAMS_SITE_FLAG	Identifies the site as a PAMS Site

Table 5-2. Site Information Data Fields (Continued)

Data Field	Data Description
IMPROVE_SITE_FLAG	Identifies the site as an IMPROVE Site
CASTNET_SITE_FLAG	Identifies the site as an CASTNET Site
PM_SUPERSITES_SITE_FLAG	Identifies the site as an PM Supersites Site
PILOT_SITE_FLAG	Identifies the site as an EPA Pilot site
POST_KATRINA_SITE_FLAG	Identifies the site as a Post-Katrina UATMP site
CSATAMP_SITE_CYCLE_FLAG	Identifies the site as a Community-Scale Air Toxics Monitoring site
CANDIDATE_NCORE_SITE_FLAG	Identifies the site as a potential NCORE monitoring site
SCHOOL_AIR_TOXICS_SITE_FLAG	Identifies the site as a School Air Toxics monitoring site
BP_OIL_SPILL_SITE_FLAG	Identifies the site as a BP Oil Spill monitoring site
LEAD_NAAQS_SITE_FLAG	Identifies the site as a Lead NAAQS monitoring site
MONITOR_LATITUDE	Vertical coordinates of the monitoring site
MONITOR_LONGITUDE	Horizontal coordinates of the monitoring site
UTM_NORTHING	Universal Transverse Mercator Projection Y-coordinate value
UTM_EASTING	Universal Transverse Mercator Projection X-coordinate value
UTM_ZONE	Zone for the UTM coordinates
ELEVATION	Elevation of the monitoring site, in meters
LOCATION_TYPE	Type of location
LAND_USE	Use of land
CBSA_NAME	Core-Based Statistical Area name
CBSA_TYPE	CBSA type (metropolitan or micropolitan)
URBAN_AREA_NAME	Shortened MSA name
MONITOR_TRAFFIC_COUNT	Traffic passing by the monitoring site
TRAFFIC_COUNT_YEAR	Year of traffic count
RFG_MANDATED_AREA_FLAG	Indicates the site is in a reformulated gasoline Mandated regulated area
RFG_OPT_IN_AREA_FLAG	Indicates the site is in a reformulated gasoline Opt-In regulated area
RFG_OPT_OUT_AREA_FLAG	Indicates the site is in a reformulated gasoline Opt-Out regulated area
WINTER_OXYGENATED_AREA_FLAG	Indicates the site is in a Winter Oxygenated regulation area
CLOSEST_NWS_STATION	Closest National Weather Service (NWS) station
CLOSEST_NWS_STATION_WBAN	Closest National Weather Service (NWS) station identifier
CLOSEST_NWS_STATION_DISTANCE_MILES	Distance between the monitoring site and the closest NWS station

Table 5-2. Site Information Data Fields (Continued)

Data Field	Data Description
CLOSEST_NWS_STATION_BEARING_FROM_EAST	Bearing angle from the east of the monitoring site and the closest NWS station
SECOND_CLOSEST_NWS_STATION	Second closest National Weather Service (NWS) station
SECOND_CLOSEST_NWS_STATION_WBAN	Second closest National Weather Service (NWS) station identifier
SECOND_CLOSEST_NWS_STATION_DISTANCE_MILES	Distance between the monitoring site and the second closest NWS station
SECOND_CLOSEST_NWS_STATION_BEARING_FROM_EAST	Bearing angle from the east of the monitoring site and the second closest NWS station
COMMENT	General comment

* = primary key field

A number of useful metadata are provided, related to site location, demographic/population activities, and regulatory applicability. A total of 2,719 records are in this data dictionary.

5.3   Monitor Information

Table 5-3 presents data fields for the monitors situated at the monitoring sites in the AMA_MONITOR_INFORMATION data table. A MONITOR_CODE is composed of the AMA_SITE_CODE, AQS_POC, and AQS_PARAMETER_CODE. These three fields, as well as YEAR represent the primary key fields (denoted by a "*"). This data dictionary table includes information about the monitor objective and monitor type, as well as the Program in which the data were collected. The Program information is useful in identifying which data were collected under EPA programs, such as the National Air Toxics Trends System, Urban Air Toxics Monitoring Program, Photochemical Assessment Monitoring Sites, and the IMPROVE network. A total of 374,213 records are in this data dictionary.

Table 5-3. Monitor Information Data Fields

Data Field	Data Description
*AMA_SITE_CODE	Site Identifier made up of STATE_FIPS, COUNTY_FIPS, and LOCAL_SITE_ID
*AQS_POC	Parameter Occurrence Code
*AQS_PARAMETER_CODE	AQS Pollutant Identifier
*SAMPLE_YEAR	Year of Sampling
MIN_DATE	Start date of measurements for the Sample Year
MAX_DATE	End date of measurements for the Sample Year

Table 5-3. Monitor Information Data Fields (Continued)

Data Field	Data Description
MONITOR_CODE	Site Identifier made up of AMA_SITE_CODE, AQS_POC, and AQS_PARAMETER_CODE
PROGRAM	Program associated with each monitor, if available
MONITOR_OBJECTIVE	Sampling Objective of the Monitor
MONITOR_TYPE	Type of Monitor
MONITOR_DESIGNATION	Indicates whether the monitor is the primary, secondary, or not determined
COUNT_RECORD	Number of AMA HAP Records
COUNT_CONCENTRATION	Number of AMA HAP Concentrations

* = primary key field

5.4 Pollutant Information

Table 5-4 presents data fields for a comprehensive list of pollutants listed in the AMA_POLLUTANT_CODES_DICTIONARY. This data table includes HAPs, non-HAPs, GHG pollutants, criteria pollutants, and meteorological data. The "AQS_PARAMETER_CODE" is the only primary key field in this data dictionary (denoted by a "*"). This data dictionary table includes physical, method profile, and pollutant grouping information. A total of 1,205 records are in the master data dictionary, of which 504 are HAPs.

Table 5-4. Pollutant Information Data Fields

Data Field	Data Description
REPORTED	Flag to identify if parameter code is to be reported in the Output file
*AQS_PARAMETER_CODE	AQS Pollutant Identifier
AQS_PARAMETER_NAME	Pollutant or Parameter Name
POLLUTANT_CASNUM	Pollutant CAS Number, if available
NEI_POLLUTANT_ID	National Emissions Inventory Pollutant Code
POLLUTANT_TYPE	Pollutant Grouping Type
REPORTING_PARAMETER_NAME	Reported Parameter Name
REPORTING_CATEGORY_NAME	Reported Pollutant Grouping Name
NUM_CARBON	Number of carbons
MOLECULAR_WEIGHT	Molecular weight of pollutant
NATTS_MQO_CORE_HAP	Designated as a priority EPA hazardous air pollutant (HAP)
URBAN_33_POLL_FLAG	Designated as an urban-33 pollutant
HAP_FLAG	Indicates pollutant is a HAP
CAP_FLAG	Indicates pollutant is a criteria air pollutant
GHG_FLAG	Indicates pollutant is a greenhouse gas air pollutant
NON_HAP_FLAG	Indicates pollutant is a non-HAP
MET_DATA_FLAG	Indicates parameter is meteorological data
TO15_FLAG	Indicates pollutant is a TO-15 compound
TO11A_FLAG	Indicates pollutant is a TO-11A compound
IO3_5_FLAG	Indicates pollutant is an IO3.5 compound
TO13_FLAG	Indicates pollutant is a TO-13A compound

Table 5-4. Pollutant Information Data Fields (Continued)

Data Field	Data Description
8270C_FLAG	Indicates pollutant is a 8270 compound
SNMOC_FLAG	Indicates pollutant is a SNMOC compound
ERG_HEX_FLAG	Indicates pollutant is a hexavalent chromium compound
PAMS_FLAG	Indicates pollutant is a PAMS compound
HEALTH_BENCHMARK_FLAG	Indicates if pollutant has a health benchmark value
UNIT_RISK_ESTIMATE	Unit Risk Estimate factor
REFERENCE_CONCENTRATION	Reference Concentration factor
NONCANCER_TARGET_SYSTEM_1	Target system affected by noncancer pollutant exposure
NONCANCER_TARGET_SYSTEM_2	Target system affected by noncancer pollutant exposure
NONCANCER_TARGET_SYSTEM_3	Target system affected by noncancer pollutant exposure
EPA_REGION_4_RISK_SCREENING_VALUE	EPA risk screening factor used as a screening approach
ATSDR_SHORT_TERM_VALUE	ATSDR short-term exposure risk factor
ATSDR_INTERMEDIATE_TERM_VALUE	ATSDR intermediate-term exposure risk factor
ATSDR_CHRONIC_VALUE	ATSDR chronic-term exposure risk factor
CAL_EPA_RELATIVE_EXPOSURE_LIMIT	California EPA Relative Exposure Limit factor
CAL_EPA_RELATIVE_EXPOSURE_LIMIT_DURATION	Sample duration for the CAL EPA REL
NAAQS_1_HOUR_VAL	Value for the 1-hour National Ambient Air Quality Standard (NAAQS)
NAAQS_3_HOUR_VAL	Value for the 3-hour NAAQS
NAAQS_8_HOUR_VAL	Value for the 8-hour NAAQS
NAAQS_DAILY_VAL	Value for the daily NAAQS
NAAQS_3_MONTH_ROLLING_VAL	Value for the 3-month rolling average NAAQS
NAAQS_QUARTERLY_VAL	Value for the quarterly average NAAQS
NAAQS_ANNUAL_VAL	Value for the annual average NAAQS
COMMENT	General comment

* = primary key field

5.5 Sampling Method Information

Table 5-5 presents data fields for a comprehensive list of sampling methodology codes listed in the AMA_SAMPLING_METHOD_CODE_DICTIONARY. The primary keys for this data table are the AQS_PARAMETER_CODE, AQS_METHODOLOGY_CODE, AQS_SAMPLE_DURATION_CODE, and the AQS_UNIT_CODE (denoted by a "*"). This data dictionary table includes the federal MDL in native units, as well as converted to standardized $\mu g/m^3$. A total of 3,994 records are in this data dictionary.

Table 5-5. Sampling Methodology Information Data Fields

Data Field	Data Description
*AQS_PARAMETER_CODE	AQS Parameter Identifier
PARAMETER_DESC	AQS Parameter Identifier Description
*AQS_METHODOLOGY_CODE	AQS Methodology Identifier
SAMPLE_COLLECTION_DESC	Sample Collection Description
SAMPLE_ANALYSIS_DESC	Sample Analysis Description
*AQS_SAMPLE_DURATION_CODE	Duration Identifier
DURATION_DESC	Duration Identifier Description
*AQS_UNIT_CODE	Unit of Measure Identifier
UNIT_DESC	Unit Description
FEDERAL_MDL_VALUE	Federal default method detection limit
FEDERAL_MDL_UNIT	Federal default method detection limit units
MDL_STD	Federal default method detection limit standardized to $\mu g/m^3$
SUMMARY_SCALE	AQS Field (unknown)
EQUIVALENT_METHOD_DESC	AQS Field (unknown)
REFERENCE_METHOD_ID	AQS Field (unknown)
COMMENT	General comment

* = primary key field

5.6   Date and Season Information

Table 5-6 presents data fields for every single day from 1973 to 2013 listed in the AMA_DATE_DICTIONARY. The primary key for this data table is the "DATE" (denoted by a "*"). This data dictionary table includes the corresponding day (Monday, Tuesday, etc.), day type (weekday or weekend), and calendar quarter in which the month belongs to (e.g., Quarter 1 = January, February, and March; Quarter 2 = April, May, and June). A total of 15,340 records are in this data dictionary.

Table 5-6. Date and Season Information Data Fields

Data Field	Data Description
*DATE	Date of the sample (MM/DD/YYYY)
DATE_TXT	Date of the sample (MM/DD/YYYY) in text format
DAY_OF_WEEK	Sunday, Monday, Tuesday, Wednesday, Thursday, Friday, or Saturday
DAY_OF_WEEK_TYPE	Weekday or Weekend
YEAR	Calendar Year
MONTH	Month
DAY	Day
DATE_FORMATTED	Date of the sample (YYYYMMDD)
DAY_NUMBER	Numeric day count
QUARTER	Identifies the quarter within the calendar year

* = primary key field

5.7   Qualifier Code Information

Table 5-7 presents data fields for the data qualifier codes in the
AMA_QUALIFIER_CODE_DICTIONARY data table. The primary key for this data table is the
"AQS_QUALIFIER_CODE" (denoted by a "*"). This data dictionary table includes information
related to quality assurance issues, sampling problems, or information related to the
concentration records. A total of 203 records are in this data dictionary.

Table 5-7. Qualifier Information Data Fields

Data Field	Data Description
*AQS_QUALIFIER_CODE	Qualifier Identifier
QUALIFIER_DESC	Qualifier Description
QUALIFIER_TYPE	Type of Qualifier
QUALIFIER_TYPE_DESC	Type of Qualifier Description

* = primary key field

5.8   Sample Duration Information

Table 5-8 presents data fields for the sample duration codes in the
AMA_SAMPLE_DURATION_CODE_DICTIONARY. The primary key for this data table is
the "AQS_DURATION_CODE" (denoted by a "*"). This data dictionary table includes
information related to the length of the sample measurements quality assurance issues, sampling
problems, or information related to the concentration records. A total of 28 records are in this
data dictionary.

Table 5-8. Sample Duration Information Data Fields

Data Field	Data Description
*AQS_DURATION_CODE	Duration Identifier
DURATION_DESC	Duration Identifier Description
DURATION_INDICATOR	Duration Indicator Identifier
DURATION_LENGTH	Length of sampling
DURATION_UNIT	Unit of length for sampling

* = primary key field

5.9   Unit Code Information

Table 5-9 presents data fields for the unit codes in the
AMA_UNIT_CODE_DICTIONARY. The primary key for this data table is
"AQS_UNIT_CODE" (denoted by a "*"). A total of 17 records are in this data dictionary.

Table 5-9. Unit Information Data Fields

Data Field	Data Description
*AQS_UNIT_CODE	Unit of Measure Identifier
UNIT_DESCRIPTION	Unit Description
UNIT_ABBR	Abbreviation of Units
REPORTED	Flag to identify if unit code is to be reported in the Output table

* = primary key field

5.10   Collection Frequency Code Information

Table 5-10 presents data fields for the sampling collection frequency codes in the AMA_COLLECTION_FREQUENCY_CODES_DICTIONARY. The primary key for this data table is "Collection Frequency Code" (denoted by a "*"). A total of 18 records are in this data dictionary.

Table 5-10. Frequency Code Data Fields

Data Field	Data Description
*AQS_COLLECTION_FREQUENCY_CODE	Collection Frequency Code Identifier
COLLECTION_FREQUENCY_DESCRIPTION	Collection Frequency Code Description
DAILY_SAMPLE_NUMBER	Number of subdaily measurements (PAMS only)
DAILY_INTERVAL	Numeric equivalent of the collection frequency code

* = primary key field

5.11   Data Source Code Information

Table 5-11 presents data fields for the sampling collection frequency codes in the AMA_DATA_SOURCE_CODES_DICTIONARY. The primary key for this data table is "DATA_SOURCE" (denoted by a "*"). A total of 21 records are in this data dictionary.

Table 5-11. Data Source Code Data Fields

Data Field	Data Description
*DATA_SOURCE	Data Source Code Identifier
DATA_SOURCE_DESCRIPTION	Data Source Code Description
NUM_RECORDS	Number of data records
MIN_YEAR	First year for the data source
MAX_YEAR	End year for the data source
NUM_PARAMETER_CODE	Number of HAPs for the data source
NUM_SITES	Number of monitoring sites for the data source
NUM_STATES	Number of states for the data source
NUM_COUNTIES	Number of counties for the data source

* = primary key field

6.0 FINAL DATABASE

The remainder of this memorandum focuses on HAPs. Table 6-1 provides a summary of the final record counts of each data source used to populate Phase X Archive of HAPs. In total, there are over 49 million data records.

Table 6-1. Data Source Information for HAP Records

Data Source	Data Years	# Sites	# Pollutants/ Parameters	HAP Data Record Count
AQS Data	1990-2014	2,212	359	45,338,081
IMPROVE Data	1990-2013	211	7	2,134,227
National Acid Deposition Program	2008-2013	25	3	836,436
Minnesota Pollution Control Agency	2008-2014	40	51	378,039
PHASE V Archive	1973-2006	419	167	211,535
EPA Region V	2012-2014	7	10	165,012
Integrated Atmospheric Deposition Network Data	1999-2008	11	88	141,938
MATES-II/III	1999-2007	17	52	133,010
Mercury Deposition Network	1999-2012	149	1	58,915
PHASE VIII Archive	1991-2010	19	16	56,304
Sublette County, Wyoming	2009-2010	14	42	37,398
Oregon DEQ XAct	2011	1	10	6,540
City of Ft. Worth, TX	2010	8	49	5,455
NATTS Network Assessment	2011-2014	3	21	5,310
Pinellas County DEM	2013-2014	2	61	4,753
South Carolina DHEC	2010-2013	1	53	2,347
Colorado DPHE	2010-2013	1	6	1,740
Maryland Department of the Environment	2014	6	1	1,048
EPA's School Air Toxics	2011-2012	6	80	800
Totals	**1973-2014**	**2,719**	**371**	**49,518,888**

Approximately 24% of the data records are non-detects, while less than 9% are null data records. It is important to note that null data records were not in EPA's Phase V database, and AQS raw data was reviewed for only the last 25 years (1990-2014); thus no conclusion can be made prior to 1990 about null data records. Finally, approximately 9% of the reported HAP records were below the MDL (BMDL). Table 6-2 provides a summary of these counts by year.

Table 6-2. HAP Summary Counts by Year

Year(s)	# HAP Records	# Non-Detect Records	% Non-Detect	# Null Data Records	% Null	# HAP Sample Values BMDL	% HAP Sample Values BMDL
<1990	56,648	92	0.2%	0	0.0%	2,086	3.7%
1990	149,049	66,515	44.6%	7,563	5.1%	11,130	7.5%
1991	184,638	83,964	45.5%	7,541	4.1%	14,537	7.9%

Table 6-2. HAP Summary Counts by Year (Continued)

Year(s)	# HAP Records	# Non-Detect Records	% Non-Detect	# Null Data Records	% Null	# HAP Sample Values BMDL	% HAP Sample Values BMDL
1992	219,913	94,298	42.9%	13,702	6.2%	16,883	7.7%
1993	295,058	105,967	35.9%	18,612	6.3%	26,051	8.8%
1994	497,311	144,208	29.0%	31,410	6.3%	31,299	6.3%
1995	862,296	220,039	25.5%	60,055	7.0%	39,277	4.6%
1996	1,037,934	258,775	24.9%	65,551	6.3%	61,829	6.0%
1997	1,260,542	284,664	22.6%	88,290	7.0%	62,128	4.9%
1998	1,454,114	308,653	21.2%	130,441	9.0%	73,448	5.1%
1999	1,568,915	342,962	21.9%	184,809	11.8%	81,847	5.2%
2000	1,703,948	414,659	24.3%	136,515	8.0%	116,095	6.8%
2001	2,047,073	490,053	23.9%	207,405	10.1%	158,528	7.7%
2002	2,119,510	544,939	25.7%	176,960	8.3%	195,340	9.2%
2003	2,169,229	550,615	25.4%	178,434	8.2%	189,187	8.7%
2004	2,538,779	623,001	24.5%	224,575	8.8%	217,541	8.6%
2005	2,847,615	701,181	24.6%	254,166	8.9%	261,301	9.2%
2006	2,873,234	702,389	24.4%	247,908	8.6%	241,228	8.4%
2007	3,093,773	699,415	22.6%	276,302	8.9%	234,110	7.6%
2008	3,013,861	685,625	22.7%	296,831	9.8%	225,629	7.5%
2009	3,259,689	753,902	23.1%	296,410	9.1%	276,396	8.5%
2010	3,353,470	785,030	23.4%	314,228	9.4%	321,387	9.6%
2011	3,508,212	869,482	24.8%	328,048	9.4%	373,718	10.7%
2012	3,272,918	736,585	22.5%	322,148	9.8%	420,385	12.8%
2013	3,113,340	743,673	23.9%	278,977	9.0%	430,962	13.8%
2014	3,017,819	707,718	23.5%	232,100	7.7%	365,928	12.1%
Totals	**49,518,888**	**11,918,404**	**24.1%**	**4,378,981**	**8.8%**	**4,448,250**	**9.0%**

Of the 11,918,404 HAP non-detects in the master database, less than 4% (473,993 records) were suspected as being non-detects in which a concentration equal to ½ MDL were either intentionally or mistakenly substituted. Table 6-3 provides an overview of these records by state, as well as whether the MDL that was used was a default federal MDL or one entered by the user.

Table 6-3. Non-Detect Records Populated with ½ MDL by State

State	Total # of ND	Total # Surrogate	# Fed MDL Surrogate	# Entity-Provided MDL Surrogates	Time Period of Surrogates
Alabama	78,307	3	2	1	1977-1996
Alaska	55,566	0	0	0	NA
Arizona	174,476	0	0	0	NA
Arkansas	24,308	0	0	0	NA
California	886,107	389,289	254,758	134,531	1985-2014
Colorado	172,485	31	31	0	2002
Connecticut	272,793	0	0	0	NA
Delaware	81,698	193	32	161	2000-2012

Table 6-3. Non-Detect Records Populated with ½ MDL by State (Continued)

State	Total # of ND	Total # Surrogate	# Fed MDL Surrogate	# Entity-Provided MDL Surrogates	Time Period of Surrogates
District of Columbia	129,697	100	77	23	1997-2008
Florida	155,997	14,641	122	14,519	1989-2006
Georgia	589,938	0	0	0	NA
Hawaii	42,772	0	0	0	NA
Idaho	50,568	10,621	0	10,621	2002-2008
Illinois	473,643	18	17	1	2005
Indiana	344,827	31	31	0	1990
Iowa	75,531	0	0	0	NA
Kansas	130,738	1	1	0	1990
Kentucky	89,979	0	0	0	NA
Louisiana	172,534	6	1	5	1977-1997
Maine	743,153	0	0	0	NA
Maryland	167,471	392	392	0	1997-2000
Massachusetts	360,192	0	0	0	NA
Michigan	329,145	57	57	0	1992-1994
Minnesota	397,725	5	5	0	1999
Mississippi	78,855	0	0	0	NA
Missouri	139,273	0	0	0	NA
Montana	108,155	2	2	0	1991
Nebraska	20,970	0	0	0	NA
Nevada	48,229	1	0	1	1979
New Hampshire	237,752	110	110	0	2002-2004
New Jersey	227,404	15	15	0	1993-2005
New Mexico	57,845	0	0	0	NA
New York	235,357	9,842	9,842	0	1990-1999
North Carolina	104,463	1,140	1,112	28	2002-2008
North Dakota	37,490	2	2	0	2000
Ohio	166,664	3	0	3	2004-2005
Oklahoma	85,368	0	0	0	NA
Oregon	148,920	39,327	2,117	37,210	1999-2007
Pennsylvania	482,746	1,086	850	236	2000-2013
Rhode Island	185,594	0	0	0	NA
South Carolina	200,811	16	16	0	1993-1994
South Dakota	62,476	0	0	0	NA
Tennessee	61,889	138	138	0	1990-1998
Texas	2,515,620	2,051	2,051	0	1994-2009
Utah	93,466	0	0	0	NA
Vermont	109,424	9	8	1	1995-2002
Virginia	120,466	244	113	131	2000-2012
Washington	142,500	4,619	4	4,615	1995-2006
West Virginia	27,544	0	0	0	NA
Wisconsin	110,925	0	0	0	NA
Wyoming	80,167	0	0	0	NA
Puerto Rico	13,634	0	0	0	NA
Virgin Islands	14,747	0	0	0	NA
Total	**11,918,404**	**473,993**	**271,906**	**202,087**	**1977-2014**

In the Phase IX database, data has been stored with native sample durations, as presented in Table 6-4.

Table 6-4. Phase X HAP Database Sample Duration Counts by Year

| Year(s) | Sub-Daily Records | | | | | | | | | Daily Records | Weekly/ Monthly/ Variable Records |
	Sub-Hourly	1-hour	2-hour	3-hour	4-hour	5-hour	6-hour	8-hour	12-hour		
<1990	0	0	0	12,381	0	0	0	0	0	44,267	0
1990	0	0	0	756	0	0	0	0	400	147,813	80
1991	0	0	0	493	0	0	0	0	0	184,135	10
1992	0	0	0	1,302	0	0	0	0	0	218,611	0
1993	0	38,579	0	21,401	0	0	872	0	0	234,206	0
1994	0	154,837	0	59,000	0	0	0	0	0	283,474	0
1995	0	459,749	0	84,192	2,088	0	133	0	0	316,134	0
1996	0	565,991	0	120,502	6,876	0	0	0	0	344,565	0
1997	0	770,316	0	119,846	3,843	0	0	0	0	366,537	0
1998	0	933,031	0	153,593	2,799	0	0	0	0	364,691	0
1999	610	982,071	0	154,112	0	0	0	2,130	0	428,273	1,719
2000	201	1,049,784	0	137,269	1,797	0	0	1,578	0	511,363	1,956
2001	1	1,177,792	0	135,038	5,879	0	0	0	6,092	719,614	2,657
2002	274	1,086,801	0	134,088	10,664	0	0	0	4,290	880,238	3,155
2003	278	1,082,626	0	116,193	9,641	0	0	0	2,262	954,581	3,648
2004	90,079	1,249,572	0	100,965	17,659	0	0	1,648	1,108	1,073,672	4,076
2005	101,521	1,409,594	0	104,265	14,526	0	0	7,458	0	1,205,970	4,281
2006	98,740	1,545,459	0	113,262	5,073	0	0	2,264	0	1,103,763	4,673
2007	311,763	1,595,975	0	125,786	0	0	2,020	0	0	1,053,199	5,030
2008	368,060	1,468,540	4,053	111,048	18	6	2,015	0	1,975	1,052,543	5,603
2009	366,978	1,590,963	86,802	112,028	1,077	384	3	0	1,089	1,094,677	5,688
2010	393,223	1,707,614	89,757	114,234	1,149	330	0	0	1,134	1,040,448	5,581
2011	277,016	2,047,798	86,472	103,425	687	225	26	0	0	987,190	5,373
2012	246,188	1,902,780	45,579	89,613	930	69	218	0	0	982,066	5,475
2013	171,364	1,863,473	44,916	87,756	264	93	0	0	0	945,474	0
2014	286,558	1,734,847	0	88,016	0	0	0	0	0	908,398	0
Totals	2,712,854	26,418,192	357,579	2,400,564	84,970	1,107	5,287	15,078	18,350	17,445,902	59,005

7.0   FINAL OUTPUT DATA FILES

The raw ambient monitoring data are housed in the "Ambient Monitoring Archive" data table. For the public release files, the key data fields in the Phase IX raw table are presented in Table 7-1. Primary key fields are denoted by a "*".

Table 7-1. Ambient Monitoring Archive Output Fields

Data Field	Data Description
STATE_ABBR	Two-letter abbreviation for the state with the monitoring site
*AMA_SITE_CODE	Ambient Monitoring Archive (AMA) Site Code
*AQS_POC	Parameter Occurrence Code (POC)
PROGRAM	Identifies Monitoring Program, if available

Table 7-1. Ambient Monitoring Archive Output Fields (Continued)

Data Field	Data Description
YEAR	Year of sampling date
QUARTER	Calendar quarter of the sampling date
*SAMPLE_DATE	Date Sample was taken
*SAMPLE_START_TIME	Time at which sample began
*AQS_PARAMETER_CODE	Air Quality System (AQS) Pollutant Code
AQS_PARAMETER_NAME	AQS pollutant name
DATA_SOURCE	Identifies the source of the data record
DURATION_DESC	Translated AQS Sample Duration description
SAMPLE_VALUE_REPORTED	Reported sample value from the data source
AQS_UNIT_CODE	Unit of Measure Code for the native sample value
UNIT_DESC	Translated AQS Unit of Measure description
SAMPLING_FREQUENCY_CODE	Collection Frequency code (1=Daily; 2=EveryOtherDay; 3=Every3Days; 4=Every4Days; 5=Every5Days; 6=Every6Days; 7=Every12Days; 8=StratifiedRandom; 9=Random; 10=Every24Days; 11=Every30Days; 12=Every7Days; 14=Every14Day; A=PAMS Daily; J=PAMS 3rdDay; P=PAMS 6thDay; S=Seasonal)
COMMENT	Reserved for comments
SAMPLE_VALUE_STD_FINAL_UG_M3	Concentration value standardized to µg/m^3, local conditions
SAMPLE_VALUE_STD_FINAL_TYPE	Final Concentration type for analysis (L = Local Conditions, S = Standard Conditions)
AQS_PARAMETER_CODE_FINAL	Final AQS Pollutant Code for analysis
AQS_PARAMETER_NAME_FINAL	Final AQS Pollutant Name for analysis
MDL_STD_UG_M3	MDL standardized to µg/m^3
MDL_TYPE	Identifies the source of the standardized MDL
AQS_NULL_DATA_CODE	Data Qualifier code for null sample values
AQS_QUALIFIER_01	Data Qualifier code field 1
AQS_QUALIFIER_02	Data Qualifier code field 2
AQS_QUALIFIER_03	Data Qualifier code field 3
AQS_QUALIFIER_04	Data Qualifier code field 4
AQS_QUALIFIER_05	Data Qualifier code field 5
AQS_QUALIFIER_06	Data Qualifier code field 6
AQS_QUALIFIER_07	Data Qualifier code field 7
AQS_QUALIFIER_08	Data Qualifier code field 8
AQS_QUALIFIER_09	Data Qualifier code field 9
AQS_QUALIFIER_10	Data Qualifier code field 10
AQS_METHOD_CODE	Sampling and Analysis Method Code
SAMPLE_COLLECTION_DESC	Translated AQS Sampling Collection description
SAMPLE_ANALYSIS_DESC	Translated AQS Analysis Method description
SAMPLE_VALUE_FLAG	Identifies if the concentration record is a non-detect
BELOW_MDL_FLAG	Identifies if the non-zero sample value is less than the MDL
CENSUS_TRACT_ID	U.S. Census tract identifier in which the monitoring site is located
CENSUS_TRACT_POPULATION_2010	2010 Population within the census tract
MONITOR_LATITUDE	Y-Coordinate Value in decimal degrees
MONITOR_LONGITUDE	X-Coordinate Value in decimal degrees

* = primary key field

In the public release files, EPA is not outputting "Acrolein – unverified" (parameter code = 43505) due to the unreliability of the measurements. Similarly, the following parameter codes

are not included in the Ambient Monitoring Archive output files, as they are combined pollutants which cannot be disaggregated for air quality use:

- 45110: Styrene and O-Xylene
- 45111: M(and P)-Xylene and Bromoform
- 45112: O-Xylene and 1,1,2,2-Tetrachloroethane
- 45115: Benzene and 1,2-Dichloroethane

Additionally, AMA records which have deposition units, such as nanogram per liter, are not outputted in the public release files. Further, AMA records prior to 1990 are not being outputted. Finally, AMA records in which there is no latitude or longitude coordinate pair are not in the public release files. Table 7-2 presents a summary of the final counts in the Output files by state. Over 98% of the Output records are in local conditions. Local condition records are initially identified as:

- Concentration records in which the reported unit codes are local conditions, such as: 105, 108
- All null or zero concentration records, regardless of reported unit
- All VOC concentration records if the sampling and analytical method codes indicated canister sampling
- All carbonyl concentrations if the data were collected by samplers under EPA's UATMP. These monitors are defaulted to collect local conditions.

For the remaining concentration records, EPA obtained, where possible, the local ambient temperature and pressure data to match the same temporal time frame of the concentration record. For example, hourly temperature and pressure were obtained for hourly measurements and daily temperature and pressure were obtained for daily measurements. The hierarchy for selecting temperature and pressure data was the following:

- Average daily temperature (AQS parameter code = 68105) and average daily pressure from AQS (AQS parameter code = 68108).
- The hourly temperature (AQS parameter code = 62101) and barometric pressure (AQS parameter code = 64101) observations from AQS to gap-fill for missing days.

- Hourly air temperature and station pressure observations from the closest National Weather Service (NWS) stations were used as a surrogate.

The calculation to convert from standard conditions (SC) to local conditions (LC) is:

$$\text{concentration, LC} = \frac{(\text{concentration, SC})*(298\ K)*(\text{local pressure in millimeters of mercury})}{(\text{local temperature in degrees Kelvin})*(760\ \text{millimeters of mercury})}$$

Table 7-2. Summary of Output Record Counts By State

State	Total # Output Records	Total # Local Condition Records	Total # Standard Condition Records	% Local Condition Records
Alabama	262,260	262,016	244	99.91%
Alaska	108,647	108,647	0	100.00%
Arizona	495,836	495,330	506	99.90%
Arkansas	57,024	56,798	226	99.60%
California	3,555,976	3,378,352	177,624	95.00%
Colorado	451,413	440,386	11,027	97.56%
Connecticut	1,198,443	1,194,623	3,820	99.68%
Delaware	258,607	254,309	4,298	98.34%
District of Columbia	499,964	498,638	1,326	99.73%
Florida	644,037	640,844	3,193	99.50%
Georgia	1,763,103	1,762,283	820	99.95%
Hawaii	143,383	115,284	28,099	80.40%
Idaho	99,451	99,104	347	99.65%
Illinois	1,150,498	1,109,272	41,226	96.42%
Indiana	1,901,783	1,887,932	13,851	99.27%
Iowa	157,988	157,642	346	99.78%
Kansas	210,392	177,231	33,161	84.24%
Kentucky	345,684	345,615	69	99.98%
Louisiana	709,275	702,897	6,378	99.10%
Maine	1,637,747	1,633,471	4,276	99.74%
Maryland	991,955	891,381	100,574	89.86%
Massachusetts	1,644,809	1,640,961	3,848	99.77%
Michigan	1,131,809	1,118,312	13,497	98.81%
Minnesota	1,046,257	1,034,736	11,521	98.90%
Mississippi	233,766	233,302	464	99.80%
Missouri	646,279	633,016	13,263	97.95%
Montana	239,270	229,535	9,735	95.93%
Nebraska	56,959	52,397	4,562	91.99%
Nevada	101,535	101,529	6	99.99%
New Hampshire	671,159	651,095	20,064	97.01%
New Jersey	1,115,285	1,111,981	3,304	99.70%
New Mexico	144,259	143,547	712	99.51%
New York	1,459,742	1,404,049	55,693	96.18%
North Carolina	374,988	373,022	1,966	99.48%

Table 7-2. Summary of Output Record Counts By State (Continued)

State	Total # Output Records	Total # Local Condition Records	Total # Standard Condition Records	% Local Condition Records
North Dakota	72,952	72,942	10	99.99%
Ohio	413,086	411,131	1,955	99.53%
Oklahoma	244,760	243,984	776	99.68%
Oregon	316,448	316,330	118	99.96%
Pennsylvania	1,265,961	1,223,810	42,151	96.67%
Rhode Island	702,174	702,160	14	100.00%
South Carolina	493,670	469,035	24,635	95.01%
South Dakota	129,122	129,122	0	100.00%
Tennessee	169,020	158,708	10,312	93.90%
Texas	15,116,701	15,087,290	29,411	99.81%
Utah	276,923	275,923	1,000	99.64%
Vermont	718,367	717,601	766	99.89%
Virginia	579,018	577,058	1,960	99.66%
Washington	399,463	398,074	1,389	99.65%
West Virginia	122,345	116,937	5,408	95.58%
Wisconsin	2,571,407	2,563,611	7,796	99.70%
Wyoming	144,505	144,505	0	100.00%
Puerto Rico	29,920	29,744	176	99.41%
Virgin Islands	43,965	43,965	0	100.00%
TOTALS	**49,319,390**	**48,621,467**	**697,923**	**98.58%**

Table 7-3 presents a summary of the final counts in the Output files by year. From 2000 to 2014, approximately 99.8% of the data records are in local conditions.

Table 7-3. Summary of Output Record Counts By Year

Year	Total # Output Records	Total # Local Condition Records	Total # Standard Condition Records	% Local Concentration Records
<1990	56,648	33,942	22,706	59.92%
1990	147,746	98,790	48,956	66.86%
1991	183,306	129,581	53,725	70.69%
1992	218,369	159,063	59,306	72.84%
1993	294,362	212,870	81,492	72.32%
1994	495,245	401,858	93,387	81.14%
1995	861,025	765,993	95,032	88.96%
1996	1,036,017	948,360	87,657	91.54%
1997	1,255,734	1,215,146	40,588	96.77%
1998	1,449,798	1,426,881	22,917	98.42%
1999	1,562,236	1,549,416	12,820	99.18%
2000	1,698,638	1,690,441	8,197	99.52%
2001	2,041,883	2,036,681	5,202	99.75%
2002	2,112,727	2,108,684	4,043	99.81%

Table 7-3. Summary of Output Record Counts By Year
(Continued)

Year	Total # Output Records	Total # Local Condition Records	Total # Standard Condition Records	% Local Concentration Records
2003	2,161,732	2,159,627	2,105	99.90%
2004	2,529,703	2,527,519	2,184	99.91%
2005	2,824,879	2,822,538	2,341	99.92%
2006	2,855,520	2,853,304	2,216	99.92%
2007	3,080,261	3,079,351	910	99.97%
2008	3,000,098	2,999,886	212	99.99%
2009	3,244,555	3,237,073	7,482	99.77%
2010	3,339,481	3,332,172	7,309	99.78%
2011	3,494,792	3,476,485	18,307	99.48%
2012	3,259,153	3,249,197	9,956	99.69%
2013	3,105,532	3,097,358	8,174	99.74%
2014	3,009,950	3,009,251	699	99.98%
Total	**49,319,390**	**48,621,467**	**697,923**	**98.58%**