TECHNICAL MEMORANDUM

TO: Madeleine Strum, Dave Shelow, and Laurie Trinca
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: June 12, 2018

SUBJECT: Phase XII 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 XII Ambient Monitoring Archive (Archive). Under a prior Delivery Order, Eastern Research Group, Inc. (ERG) prepared Phase XI, 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 1990 to 2016.

ERG was tasked to develop Phase XII by updating the Archive to 2015, incorporate additional data not in the Archive, and provide general maintenance/cleanup of the Phase XI Archive. All work was performed under EPA Contract No. EP-D-14-030, Delivery Order 00-51 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.
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>
<thead>
<tr>
<th>Phase</th>
<th>Year Completed</th>
<th>Coverage Years</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>2001</td>
<td>1990-2000</td>
</tr>
<tr>
<td>II</td>
<td>2003</td>
<td>1990-2001</td>
</tr>
<tr>
<td>III</td>
<td>2004</td>
<td>1990-2002</td>
</tr>
<tr>
<td>IV</td>
<td>2005</td>
<td>1990-2003</td>
</tr>
<tr>
<td>V</td>
<td>2007</td>
<td>1973-2005</td>
</tr>
<tr>
<td>VI</td>
<td>2009</td>
<td>1973-2008</td>
</tr>
<tr>
<td>VII</td>
<td>2013 (Feb)</td>
<td>1973-2010</td>
</tr>
<tr>
<td>VIII</td>
<td>2013 (Oct)</td>
<td>1973-2012</td>
</tr>
<tr>
<td>IX</td>
<td>2015</td>
<td>1973-2013</td>
</tr>
<tr>
<td>X</td>
<td>2016</td>
<td>1973-2014</td>
</tr>
<tr>
<td>XI</td>
<td>2017</td>
<td>1990-2015</td>
</tr>
</tbody>
</table>

EPA previously developed the Phase XI Archive in February 2017, which contained 52 million HAP records from 1990 to 2014. The Phase XI Archive was the sixth 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 XII update, EPA requested that ERG:

- Retrieve 1990-2016 ambient HAP data from EPA’s Air Quality Subsystem (AQS);
- Retrieve non-HAP species, criteria pollutant, greenhouse gas (GHG), and meteorological data from 1990-2016;
- 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 XII Archive, there were eighteen primary data sources used. Information about each Data Source is presented in Sections 3.1 – 3.18.

3.1 Air Quality System (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 2016 were retrieved
in November 2017. Additionally, data from 1990-2015 were also retrieved to replace Phase XI database records since the Archive was last updated (February 2017). Subsequent data pulls were performed in February and April 2018 as EPA was alerted to new data being added into AQS. Over 52 million records from 2,290 sites and 365 parameters were incorporated into the Archive.

3.2 Allegheny County, PA

The Allegheny County Health Department (ACHD) in Pittsburgh, PA conducts addition metals and VOC sampling in the Pittsburgh area in which the data are not sent to AQS. As such, EPA coordinated with ACHD to obtain this data, as well as site metadata. More information on the ACHD can be found at: http://www.achd.net/air/index.php. A total of 5,251 records from 3 sites and 14 parameters were incorporated into the Archive.

3.3 Baldwin Hills Air Quality Study

Los Angeles County, in coordination with the South Coast Air Quality Management District conducted an air quality study in the Baldwin Hills area near oil and gas activities. More information on this study can be found at: http://planning.lacounty.gov/assets/upl/project/bh_air-quality-study.pdf. This data was sent to EPA for inclusion into the Archive, as it is not housed in AQS. A total of 7,146 records from one site\(^1\) and 15 parameters were incorporated into the Archive.

3.4 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 and ending in 2015. The study focused on establishing baseline air quality concentrations. A total of 1,734 records from six sites\(^2\) and one parameter were incorporated into the Archive.

---

\(^1\) A unique AMA_SITE_CODE identifier (06037BALD) was assigned based on the 2-digit state code, 3-digit county code, and the unique site code. The Baldwin Hills site is located in Los Angeles County, CA (FIPS = 06037) and the site identifier is “BALD”.

\(^2\) Unique AMA_SITE_CODE identifiers were assigned based on the 2-digit state code, 3-digit county code, and the unique site code. For example, the “24510PAM2” site is located in Baltimore City, MD (FIPS = 24510) and the site identifier is “PAM2”.

4
3.5  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\(^3\) for 49 parameters. More information can be found at: [http://fortworthtexas.gov/uploadedFiles/Gas_Wells/AirQualityStudy_final.pdf](http://fortworthtexas.gov/uploadedFiles/Gas_Wells/AirQualityStudy_final.pdf)

3.6  EPA Passive Sampling Tubes Study

EPA’s Office of Research and Development (ORD), in coordination with EPA Region 3 and the Department of Public Health in Philadelphia, conducted a multi-site, multi-pollutant air toxics study using passive sampling tubes. Over a 21-month period, 2-week duration samples were collected at 17 sites in South Philadelphia. More information on can be found at: [https://cfpub.epa.gov/si/si_public_file_download.cfm?p_download_id=527897](https://cfpub.epa.gov/si/si_public_file_download.cfm?p_download_id=527897). A total of 18,675 records from seventeen sites\(^4\) and nine parameters were incorporated into the Archive.

3.7  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 1991-2010 for the organic, PAH, and PCB

---

\(^3\) Unique AMA_SITE_CODE identifiers were assigned based on the 2-digit state code, 3-digit county code, and the unique site code. For example, the “48439LS02” site is located in Tarrant County, TX (FIPS = 48439) and the site identifier is “LS02”.

\(^4\) Unique AMA_SITE_CODE identifiers were assigned based on the 2-digit state code, 3-digit county code, and the unique site code. For example, the “42101PS04” site is located in Philadelphia County, PA (FIPS = 42101) and the site identifier is “PS04”.

compounds were retrieved from the IADN website (http://ec.gc.ca/data_donnees/STB-AQRD/Toxics/IADN/). A total of 162,836 records from eleven sites and 89 parameters were incorporated into the Archive.

3.8 Minnesota Air Toxics Data

The Minnesota Pollution Control Agency (MNPCA) oversees a large network of air toxics monitoring stations across the state. While the data were uploaded to AQS, EPA was alerted about data reporting issues that occurred when reporting to AQS, such as truncation of concentrations, missing method detection limits, and revised data. As such, MNPCA submitted their entire dataset from 2008-2016 to EPA for review and possible replacement of the data records that were in AQS for the Phase XII Archive. More information on the MNPCA air toxics program can be found at: https://www.pca.state.mn.us/air/air-monitoring-network-plan. A total of 135,850 records from 61 sites and 61 parameters were incorporated into the Archive.

3.9 Multiple Air Toxics Exposure Study (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, MATES-III, and MATES-IV 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. A total of 160,293 records from 17 sites and 59 parameters were incorporated into the Archive.

3.10 National Acid Deposition Program (NADP) Data

The National Acid Deposition Program (NADP) 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 through 2016 from the above networks were downloaded or sent to EPA via

---

5 Unique AMA_SITE_CODE identifiers were assigned based on the 2-digit state code, 3-digit county code, and the unique site code. For example, the “26019SDB1” site is located in Benzie County, MI (FIPS = 26019) and the site identifier is “SDB1”.
request, and processed from http://nadp.sws.uiuc.edu/data/. A total of 1,406,086 records from 165 sites\(^6\) and 4 parameters were incorporated into the Archive.

### 3.11 National Air Toxics Trends System (NATTS) Network Data Review

In Fall 2017, EPA prepared a final 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. The NATTS Network Assessment covers measurements from the 2003 through 2014 sampling years. More information on the NATTS Program can be found at: https://www3.epa.gov/ttn/amtic/natts.html. A total of 2,739 records from 3 sites and 71 parameters were incorporated into the Archive.

### 3.12 National Oceanic and Atmospheric Administration (NOAA) Global Monitoring System Data

Several air toxics (benzene, carbon tetrachloride, chloromethane, etc.) overlap with the trace gases collected by the National Oceanic and Atmospheric Administration’s (NOAA) Global Monitoring Division\(^7,8,9,10\) and a composite dataset was sent to EPA.\(^11\) These are 5-minute time-averaged samples collected weekly throughout the year at remote sites or other regional background sites not directly impacted by local sources. More information can be found on NOAA’s Global Monitoring Division (https://www.esrl.noaa.gov/gmd/), as well as other

---

\(^6\) Unique AMA_SITE_CODE identifiers were assigned based on the 2-digit state code, 3-digit county code, and the unique site code. For example, the “34023NJ30” site is located in Middlesex County, NJ (FIPS = 34023) and the site identifier is “NJ30”.

\(^7\) Methylene Chloride: https://www.esrl.noaa.gov/gmd/hats/gases/CH2Cl2.html

\(^8\) Tetrachloroethylene: https://www.esrl.noaa.gov/gmd/hats/gases/C2Cl4.html

\(^9\) Benzene: See information on the Tall Towers sites at https://www.esrl.noaa.gov/gmd/ccgg/insitu/.


\(^11\) Updated data provided by Stephen Montzka, A. NOAA/Earth System Research Laboratory/Global Monitoring Division
published references. A total of 158,780 records from seventeen sites and seven parameters were incorporated into the Archive.

3.13 Phase V/VII Database

The Phase V Database originally 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. The Phase VII Database consists of historical data that had been invalidated and are no longer in AQS. Nearly all of these records are for invalidated VOC data originally submitted by the Kentucky Department of Environmental Services. A total of 211,022 records from 147 sites and 165 parameters were incorporated into the Archive.

3.14 School Air Toxics, Phase 2

In 2009, as part of an 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

---


15 Unique AMA_SITE_CODE identifiers were assigned based on the 2-digit state code, 3-digit county code, and the unique site code. For example, the Mauna Loa site (Site ID = MLO), located in Hawaii County, HI (FIPS code = 15001) is assigned 15001NMLO.
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. Additionally, some records from the Alabama schools were not entered in AQS. 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/. A total of 800 records from 6 sites and 80 parameters were incorporated into the Archive.

3.15 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 from fourteen sites16 and 42 parameters were formatted for upload for the Archive. More information can be found at: http://www.sublettewyo.com/documentcenter/view/438.

3.16 Texas Commission on Environmental Quality

The State of Texas maintains a large archive of ambient HAP measurements on its Texas Air Monitoring Information System (TAMIS) website, which allows for ad-hoc queries (http://www17.tceq.texas.gov/tamis/index.cfm?fuseaction=home.welcome). Measurements from the TAMIS website were compared to those in AQS to identify missing data that could be included in the Archive. Priority was given to TAMIS data over AQS if overlaps were found. A total of 14,881,267 records from 122 sites and 79 parameters were incorporated into the Archive.

3.17 West Virginia Special Studies

The West Virginia Division of Air Quality conducted multi-year metals measurements at two sites in WV targeting specific sources of interest. This data was sent to EPA for inclusion

---

16 Unique AMA_SITE_CODE identifiers were assigned based on the 2-digit state code, 3-digit county code, and the unique site code. For example, the “56035DANI” site is located in Sublette County, WY (FIPS = 56035) and the site identifier is “DANI”.
into the Archive, as it is not housed in AQS. A total of 2,065 records from 2 sites and 14 parameters were incorporated into the Archive.

3.18 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 2016 in Illinois, Indiana, and Michigan using XAct for targeting of specific sources. A total of 207,842 records from eight sites\(^\text{17}\) and 10 parameters were incorporated into the Archive.

3.19 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, all of the HAP measurements from the TAMIS website were uploaded in SQL Server and compared to the AQS data to identify missing data.

SQL Server also offers the ability to create primary key constraints on tables to ensure no duplication of records. In total, there are over 2.0 billion records in the Archive. Of that, over 53 million HAP records were in the blended master database.

4.0 QA FIXES

After a preliminary assessment of the Phase XII 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 \(\frac{1}{2}\) MDL. The concentrations for these

\(^\text{17}\) Unique AMA_SITE_CODE identifiers were assigned based on the 2-digit state code, 3-digit county code, and the unique site code. For example, the “18089XGAR” site is located in Lake County, IN (FIPS = 18089) and the site identifier is “XGAR”. 
records were changed to 0, and the SAMPLE_VALUE_FLAG field was populated with “ND”. The following approach was used to identify these records:

- Step 1: Identify all records in which the concentration is one-half MDL.
- 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.
- 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.
- 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 invalidated all VOC measurements analyzed by their laboratory since 1995 due to laboratory error (“AR” code). Finally, the following pollutants were invalidated due to special circumstances:
  - **Hexavalent Chromium:** All hexavalent chromium concentrations prior to 2005 were invalidated due to the sampling and analysis method not being officially approved by EPA.
  - **PAHs:** All polycyclic aromatic hydrocarbon (e.g., naphthalene, benzo(a)pyrene, anthracene, etc.) concentrations prior to 2007 were invalidated due to the sampling and analysis method not being officially approved by EPA.
  - **Acrolein:** All acrolein concentrations prior to 2005 were invalidated due to the sampling and analysis method not being officially approved by EPA.

- **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_QUALIFIER_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. Additionally, overlaps may occur between the xylenes as data could be reported as “total xylenes”, “m/p-xylene”, “m-xylene”, “o-xylene”, and/or “p-xylene”. Accordingly, “DUP” was assigned to the AQS_QUALIFIER_06 field to quickly identify these records as being invalidated.

• **AQS_QUALIFIER_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.

- **AQS_QUALIFIER_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:

  - CF-N: Calculation frequency codes which were null, and were populated by EPA.
  - CF-I-S: Calculation frequency codes incorrectly entered as “S”, which is
    Seasonal.
  - CF-I-P: Calculation frequency codes incorrectly entered as “P”, which is PAMS,
    every 6th day.
  - CF-I-J: Calculation frequency codes incorrectly entered as “J”, which is PAMS,
    every 3rd day.
  - CF-I-A: Calculation frequency codes incorrectly entered as “A”, which is PAMS,
    every day.
  - CF-I-14: Calculation frequency codes incorrectly entered as “14”, which is every
    14th day.
  - CF-I-12: Calculation frequency codes incorrectly entered as “12”, which is every
    7th day.
  - CF-I-11: Calculation frequency codes incorrectly entered as “11”, which is every
    30 days.
  - CF-I-10: Calculation frequency codes incorrectly entered as “10”, which is every
    24 days.
  - CF-I-9: Calculation frequency codes incorrectly entered as “9”, which is Random.
  - CF-I-8: Calculation frequency codes incorrectly entered as “8”, which is Stratified
    Random.
  - CF-I-7: Calculation frequency codes incorrectly entered a “7”, which is every
    12 days.
  - CF-I-6: Calculation frequency codes incorrectly entered a “6”, which is every
    6 days.
  - CF-I-5: Calculation frequency codes incorrectly entered a “5”, which is every
    5 days.
o CF-I-4: Calculation frequency codes incorrectly entered as “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.

o CF-I-0: Calculation frequency codes incorrectly entered a “0”, which is not a valid code.

• **AQS_QUALIFER_09**: This field is reserved for data records in which the sample value was suspected to be populated with $\frac{1}{2}$ 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.\textsuperscript{18}

• **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 XII 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 XI 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

<table>
<thead>
<tr>
<th>Data Field</th>
<th>Data Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>*AMA_SITE_CODE</td>
<td>Ambient Monitoring Archive (AMA) Site Code</td>
</tr>
<tr>
<td>*AQS_POC</td>
<td>Parameter Occurrence Code</td>
</tr>
<tr>
<td>*SAMPLE_DATE</td>
<td>Date Sample was taken</td>
</tr>
<tr>
<td>*SAMPLE_START_TIME</td>
<td>Time at which sample began</td>
</tr>
<tr>
<td>*AQS_PARAMETER_CODE</td>
<td>Air Quality Subsystem (AQS) Pollutant Code</td>
</tr>
<tr>
<td>HAP_FLAG</td>
<td>Flag to identify if HAP record</td>
</tr>
<tr>
<td>NON_HAP_FLAG</td>
<td>Flag to identify if Non-HAP record</td>
</tr>
<tr>
<td>CRITERIA_POLL_FLAG</td>
<td>Flag to identify if criteria pollutant record</td>
</tr>
<tr>
<td>GHG_POLL_FLAG</td>
<td>Flag to identify if GHG record</td>
</tr>
<tr>
<td>MET_DATA_FLAG</td>
<td>Flag to identify if meteorological data record</td>
</tr>
<tr>
<td>AQS_METHOD_CODE</td>
<td>Sampling Method Code</td>
</tr>
<tr>
<td>AQS_UNIT_CODE</td>
<td>Unit of Measure Code</td>
</tr>
<tr>
<td>AQS_MONITOR_PROTOCOL_ID</td>
<td>AQS Protocol ID for precision and accuracy records</td>
</tr>
<tr>
<td>AQS_QUALIFIER_01</td>
<td>Data Qualifier code field (reserved for reporting agency)</td>
</tr>
<tr>
<td>AQS_QUALIFIER_02</td>
<td>Data Qualifier code field (reserved for reporting agency)</td>
</tr>
<tr>
<td>AQS_QUALIFIER_03</td>
<td>Data Qualifier code field (reserved for reporting agency)</td>
</tr>
<tr>
<td>AQS_QUALIFIER_04</td>
<td>Data Qualifier code field (reserved for reporting agency)</td>
</tr>
<tr>
<td>AQS_QUALIFIER_05</td>
<td>Data Qualifier code field (reserved for reporting agency)</td>
</tr>
<tr>
<td>AQS_QUALIFIER_06</td>
<td>Data Qualifier code field (reserved for EPA QA – See Section 4.0)</td>
</tr>
<tr>
<td>AQS_QUALIFIER_07</td>
<td>Data Qualifier code field (reserved for EPA QA – See Section 4.0)</td>
</tr>
<tr>
<td>AQS_QUALIFIER_08</td>
<td>Data Qualifier code field (reserved for EPA QA – See Section 4.0)</td>
</tr>
<tr>
<td>AQS_QUALIFIER_09</td>
<td>Data Qualifier code field (reserved for EPA QA – See Section 4.0)</td>
</tr>
<tr>
<td>AQS_QUALIFIER_10</td>
<td>Data Qualifier code field (reserved for EPA QA – See Section 4.0)</td>
</tr>
<tr>
<td>ALTERNATE_MDL</td>
<td>Method detection limit (MDL), in native units, if entered by Entity</td>
</tr>
<tr>
<td>UNCERTAINTY</td>
<td>Estimate of uncertainty surrounding the data, if available</td>
</tr>
<tr>
<td>AQS_SAMPLING_FREQUENCY_CODE</td>
<td>Code identifying how often the measurements were collected</td>
</tr>
<tr>
<td>SAMPLE_VALUE_REPORTED</td>
<td>Reported concentration value (in native units, where possible)</td>
</tr>
<tr>
<td>SAMPLE_VALUE_ADJ</td>
<td>Adjusted concentration value</td>
</tr>
<tr>
<td>DATA_SOURCE</td>
<td>Identifies the data source for the data record</td>
</tr>
<tr>
<td>COMMENT</td>
<td>Reserved for comments</td>
</tr>
<tr>
<td>SAMPLE_VALUE_STD</td>
<td>Concentration value standardized to µg/m³</td>
</tr>
<tr>
<td>MDL_STD</td>
<td>MDL standardized to µg/m³</td>
</tr>
<tr>
<td>MDL_TYPE</td>
<td>Identifies the source of the standardized MDL</td>
</tr>
<tr>
<td>SAMPLE_VALUE_FLAG</td>
<td>Identifies if the concentration record is a non-detect (Flag = “ND”)</td>
</tr>
<tr>
<td>AQS_FEDERAL_MDL</td>
<td>Default Federal MDL value</td>
</tr>
<tr>
<td>AQS_FEDERAL_MDL_UNIT_CODE</td>
<td>Default Federal MDL value engineering Unit of Measure Code</td>
</tr>
</tbody>
</table>
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

<table>
<thead>
<tr>
<th>Data Field</th>
<th>Data Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>*AMA_SITE_CODE</td>
<td>Site Identifier made up of STATE_FIPS, COUNTY_FIPS, and LOCAL_SITE_ID</td>
</tr>
<tr>
<td>STATE_FIPS</td>
<td>State Code</td>
</tr>
<tr>
<td>COUNTY_FIPS</td>
<td>County Code</td>
</tr>
<tr>
<td>STATE_COUNTY_FIPS</td>
<td>Combination of the State and County FIPS</td>
</tr>
<tr>
<td>COUNTY_NAME</td>
<td>County Name</td>
</tr>
<tr>
<td>LOCAL_SITE_ID</td>
<td>Local Site Identifier</td>
</tr>
<tr>
<td>SITE_NAME</td>
<td>Name of Site, if available</td>
</tr>
<tr>
<td>CENSUS_TRACT_ID_2000</td>
<td>U.S. Census Tract Identifier for Year 2000</td>
</tr>
<tr>
<td>CENSUS_TRACT_ID_2010</td>
<td>U.S. Census Tract Identifier for Year 2010</td>
</tr>
<tr>
<td>CENSUS_TRACT_POPULATION_2000</td>
<td>U.S. Census Tract population for Year 2000</td>
</tr>
<tr>
<td>CENSUS_TRACT_POPULATION_2010</td>
<td>U.S. Census Tract population for Year 2010</td>
</tr>
<tr>
<td>CENSUS_BLOCK_ID_12</td>
<td>U.S. Census Block Identifier for Year 2010</td>
</tr>
<tr>
<td>ADDRESS</td>
<td>Monitoring Site Address</td>
</tr>
<tr>
<td>CITY</td>
<td>Monitoring Site City</td>
</tr>
<tr>
<td>STATE_ABBR</td>
<td>Monitoring Site State Abbreviation</td>
</tr>
<tr>
<td>ZIP_CODE</td>
<td>Monitoring Site Zip Code</td>
</tr>
<tr>
<td>EPA_REGION</td>
<td>EPA Region Designation</td>
</tr>
<tr>
<td>SUPPORT_AGENCY_CODE</td>
<td>Code for the Support Agency</td>
</tr>
<tr>
<td>SUPPORT_AGENCY</td>
<td>Support Agency Name</td>
</tr>
<tr>
<td>NATTS_SITE_FLAG</td>
<td>Identifies the site as a NATTS Site</td>
</tr>
<tr>
<td>UATMP_SITE_FLAG</td>
<td>Identifies the site as a UATMP Site</td>
</tr>
<tr>
<td>PAMS_SITE_FLAG</td>
<td>Identifies the site as a PAMS Site</td>
</tr>
<tr>
<td>IMPROVE_SITE_FLAG</td>
<td>Identifies the site as an IMPROVE Site</td>
</tr>
<tr>
<td>CASTNET_SITE_FLAG</td>
<td>Identifies the site as an CASTNET Site</td>
</tr>
<tr>
<td>PM_SUPERSITES_SITE_FLAG</td>
<td>Identifies the site as an PM Supersites Site</td>
</tr>
<tr>
<td>PILOT_SITE_FLAG</td>
<td>Identifies the site as an EPA Pilot site</td>
</tr>
<tr>
<td>POST_KATRINA_SITE_FLAG</td>
<td>Identifies the site as a Post-Katrina UATMP site</td>
</tr>
<tr>
<td>CSATAMP_SITE_CYCLE_FLAG</td>
<td>Identifies the site as a Community-Scale Air Toxics Monitoring site</td>
</tr>
<tr>
<td>CANDIDATE_NCORE_SITE_FLAG</td>
<td>Identifies the site as a potential NCORE monitoring site</td>
</tr>
<tr>
<td>SCHOOL_AIR_TOXICS_SITE_FLAG</td>
<td>Identifies the site as a School Air Toxics monitoring site</td>
</tr>
<tr>
<td>BP_OIL_SPILL_SITE_FLAG</td>
<td>Identifies the site as a BP Oil Spill monitoring site</td>
</tr>
<tr>
<td>LEAD_NAAQS_SITE_FLAG</td>
<td>Identifies the site as a Lead NAAQS monitoring site</td>
</tr>
<tr>
<td>MONITOR_LATITUDE</td>
<td>Vertical coordinates of the monitoring site</td>
</tr>
<tr>
<td>Data Field</td>
<td>Data Description</td>
</tr>
<tr>
<td>----------------------------</td>
<td>----------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>MONITOR_LONGITUDE</td>
<td>Horizontal coordinates of the monitoring site</td>
</tr>
<tr>
<td>DATUM</td>
<td>Coordinate data system</td>
</tr>
<tr>
<td>UTM_NORTHING</td>
<td>Universal Transverse Mercator Projection Y-coordinate value</td>
</tr>
<tr>
<td>UTM_EASTING</td>
<td>Universal Transverse Mercator Projection X-coordinate value</td>
</tr>
<tr>
<td>UTM_ZONE</td>
<td>Zone for the UTM coordinates</td>
</tr>
<tr>
<td>ELEVATION</td>
<td>Elevation of the monitoring site, in meters</td>
</tr>
<tr>
<td>LOCATION_TYPE</td>
<td>Type of location</td>
</tr>
<tr>
<td>LAND_USE</td>
<td>Use of land</td>
</tr>
<tr>
<td>DATE_SITE_ESTABLISHED</td>
<td>Data in which the site was operational</td>
</tr>
<tr>
<td>DATE_SITE_CLOSED</td>
<td>Data in which the site ceased operations</td>
</tr>
<tr>
<td>CBSA_NAME</td>
<td>Core-Based Statistical Area name</td>
</tr>
<tr>
<td>CBSA_TYPE</td>
<td>CBSA type (metropolitan or micropolitan)</td>
</tr>
<tr>
<td>URBAN_AREA_NAME</td>
<td>Shortened MSA name</td>
</tr>
<tr>
<td>MONITOR_TRAFFIC_COUNT</td>
<td>Traffic passing by the monitoring site</td>
</tr>
<tr>
<td>TRAFFIC_COUNT_YEAR</td>
<td>Year of traffic count</td>
</tr>
<tr>
<td>RFG_MANDATED_AREA_FLAG</td>
<td>Indicates the site is in a reformulated gasoline Mandated regulated area</td>
</tr>
<tr>
<td>RFG_OPT_IN_AREA_FLAG</td>
<td>Indicates the site is in a reformulated gasoline Opt-In regulated area</td>
</tr>
<tr>
<td>RFG_OPT_OUT_AREA_FLAG</td>
<td>Indicates the site is in a reformulated gasoline Opt-Out regulated area</td>
</tr>
<tr>
<td>WINTER_OXYGENATED_AREA_FLAG</td>
<td>Indicates the site is in a Winter Oxygenated regulation area</td>
</tr>
<tr>
<td>CLOSEST_NWS_STATION</td>
<td>Closest National Weather Service (NWS) station</td>
</tr>
<tr>
<td>CLOSEST_NWS_STATION_WBAN</td>
<td>Closest National Weather Service (NWS) station identifier</td>
</tr>
<tr>
<td>CLOSEST_NWS_STATION_DISTANCE_MILES</td>
<td>Distance between the monitoring site and the closest NWS station</td>
</tr>
<tr>
<td>CLOSEST_NWS_STATION_BEARING_FROM_EAST</td>
<td>Bearing angle from the east of the monitoring site and the closest NWS station</td>
</tr>
<tr>
<td>SECOND_CLOSEST_NWS_STATION</td>
<td>Second closest National Weather Service (NWS) station</td>
</tr>
<tr>
<td>SECOND_CLOSEST_NWS_STATION_WBAN</td>
<td>Second closest National Weather Service (NWS) station identifier</td>
</tr>
<tr>
<td>SECOND_CLOSEST_NWS_STATION_DISTANCE_MILES</td>
<td>Distance between the monitoring site and the second closest NWS station</td>
</tr>
<tr>
<td>SECOND_CLOSEST_NWS_STATION_BEARING_FROM_EAST</td>
<td>Bearing angle from the east of the monitoring site and the second closest NWS station</td>
</tr>
<tr>
<td>COMMENT</td>
<td>General comment</td>
</tr>
</tbody>
</table>

* = primary key field
A number of useful metadata are provided related to site location, monitoring programs, demographic/population activities, and regulatory applicability. A total of 2,565 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 407,649 records are in this data dictionary.

#### Table 5-3. Monitor Information Data Fields

<table>
<thead>
<tr>
<th>Data Field</th>
<th>Data Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>*AMA_SITE_CODE</td>
<td>Site Identifier made up of STATE_FIPS, COUNTY_FIPS, and LOCAL_SITE_ID</td>
</tr>
<tr>
<td>*AQS_POC</td>
<td>Parameter Occurrence Code</td>
</tr>
<tr>
<td>*AQS_PARAMETER_CODE</td>
<td>AQS Pollutant Identifier</td>
</tr>
<tr>
<td>*SAMPLE_YEAR</td>
<td>Year of Sampling</td>
</tr>
<tr>
<td>MIN_DATE</td>
<td>Start date of measurements for the Sample Year</td>
</tr>
<tr>
<td>MAX_DATE</td>
<td>End date of measurements for the Sample Year</td>
</tr>
<tr>
<td>MONITOR_CODE</td>
<td>Site Identifier made up of AMA_SITE_CODE, AQS_POC, and AQS_PARAMETER_CODE</td>
</tr>
<tr>
<td>PROGRAM</td>
<td>Program associated with each monitor, if available</td>
</tr>
<tr>
<td>MONITOR_OBJECTIVE</td>
<td>Sampling Objective of the Monitor</td>
</tr>
<tr>
<td>MONITOR_TYPE</td>
<td>Type of Monitor</td>
</tr>
<tr>
<td>MONITOR_DESIGNATION</td>
<td>Indicates whether the monitor is the primary, secondary, or not determined</td>
</tr>
<tr>
<td>COUNT_RECORD</td>
<td>Number of AMA HAP Records</td>
</tr>
<tr>
<td>COUNT_CONCENTRATION</td>
<td>Number of AMA HAP Concentrations</td>
</tr>
<tr>
<td>ERG_COMMENT</td>
<td>Comment field</td>
</tr>
<tr>
<td>SAMPLING_FREQUENCY_DESCRIPTION</td>
<td>Description of the collection frequency</td>
</tr>
<tr>
<td>SAMPLING_DURATION_DESCRIPTION</td>
<td>Description of the sample duration</td>
</tr>
<tr>
<td>PRIORITY_TRENDS</td>
<td>Ranking of monitor datasets</td>
</tr>
<tr>
<td>AQS_METHOD_CODE</td>
<td>AQS Method Code(s) per monitor</td>
</tr>
</tbody>
</table>

* = primary key field

The PRIORITY_TRENDS data field prioritizes each monitor based on program requirements, sampling and analytical methods, temporal coverage, and method quality.
objectives (e.g., completeness or sensitivity), and can be helpful in data analysis trends. For example, benzene data collected under the NATTS Program are required to meet more stringent method quality objectives, as compared to benzene data collected under the PAMS Program. Thus, the benzene from the NATTS Program will have a higher priority ranking than the benzene from the PAMS Program.

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 1,486 records are in the master data dictionary, of which 586 are HAPs.

<table>
<thead>
<tr>
<th>Data Field</th>
<th>Data Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>REPORTED</td>
<td>Flag to identify if parameter code is to be reported in the Output file</td>
</tr>
<tr>
<td>*AQS_PARAMETER_CODE</td>
<td>AQS Pollutant Identifier</td>
</tr>
<tr>
<td>AQS_PARAMETER_NAME</td>
<td>Pollutant or Parameter Name</td>
</tr>
<tr>
<td>POLLUTANT_CASNUM</td>
<td>Pollutant CAS Number, if available</td>
</tr>
<tr>
<td>NEI_POLLUTANT_ID</td>
<td>National Emissions Inventory Pollutant Code</td>
</tr>
<tr>
<td>POLLUTANT_TYPE</td>
<td>Pollutant Grouping Type</td>
</tr>
<tr>
<td>REPORTING_PARAMETER_NAME</td>
<td>Reported Parameter Name</td>
</tr>
<tr>
<td>REPORTING_CATEGORY_NAME</td>
<td>Reported Pollutant Grouping Name</td>
</tr>
<tr>
<td>NUM_CARBON</td>
<td>Number of carbons</td>
</tr>
<tr>
<td>MOLECULAR_WEIGHT</td>
<td>Molecular weight of pollutant</td>
</tr>
<tr>
<td>NATTS_MQO_CORE_HAP</td>
<td>Designated as a priority EPA hazardous air pollutant (HAP)</td>
</tr>
<tr>
<td>URBAN_33_POLL_FLAG</td>
<td>Designated as an urban-33 pollutant</td>
</tr>
<tr>
<td>HAP_FLAG</td>
<td>Indicates pollutant is a HAP</td>
</tr>
<tr>
<td>CAP_FLAG</td>
<td>Indicates pollutant is a criteria air pollutant</td>
</tr>
<tr>
<td>GHG_FLAG</td>
<td>Indicates pollutant is a greenhouse gas air pollutant</td>
</tr>
<tr>
<td>NON_HAP_FLAG</td>
<td>Indicates pollutant is a non-HAP</td>
</tr>
<tr>
<td>MET_DATA_FLAG</td>
<td>Indicates parameter is meteorological data</td>
</tr>
<tr>
<td>TO15_FLAG</td>
<td>Indicates pollutant is a TO-15 compound</td>
</tr>
<tr>
<td>TO11A_FLAG</td>
<td>Indicates pollutant is a TO-11A compound</td>
</tr>
<tr>
<td>IO3_5_FLAG</td>
<td>Indicates pollutant is an IO3.5 compound</td>
</tr>
<tr>
<td>TO13_FLAG</td>
<td>Indicates pollutant is a TO-13A compound</td>
</tr>
<tr>
<td>8270C_FLAG</td>
<td>Indicates pollutant is a 8270 compound</td>
</tr>
<tr>
<td>SNMOC_FLAG</td>
<td>Indicates pollutant is a SNMOC compound</td>
</tr>
</tbody>
</table>
Table 5-4. Pollutant Information Data Fields

<table>
<thead>
<tr>
<th>Data Field</th>
<th>Data Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>ERG_HEX_FLAG</td>
<td>Indicates pollutant is a hexavalent chromium compound</td>
</tr>
<tr>
<td>PAMS_FLAG</td>
<td>Indicates pollutant is a PAMS compound</td>
</tr>
<tr>
<td>HEALTH_BENCHMARK_FLAG</td>
<td>Indicates if pollutant has a health benchmark value</td>
</tr>
<tr>
<td>UNIT_RISK_ESTIMATE</td>
<td>Unit Risk Estimate factor</td>
</tr>
<tr>
<td>REFERENCE_CONCENTRATION</td>
<td>Reference Concentration factor</td>
</tr>
<tr>
<td>NONCANCER_TARGET_SYSTEM_1</td>
<td>Target system affected by noncancer pollutant exposure</td>
</tr>
<tr>
<td>NONCANCER_TARGET_SYSTEM_2</td>
<td>Target system affected by noncancer pollutant exposure</td>
</tr>
<tr>
<td>NONCANCER_TARGET_SYSTEM_3</td>
<td>Target system affected by noncancer pollutant exposure</td>
</tr>
<tr>
<td>EPA_REGION_4_RISK Screening VALUE</td>
<td>EPA risk screening factor used as a screening approach</td>
</tr>
<tr>
<td>ATSDR_SHORT_TERM_VALUE</td>
<td>ATSDR short-term exposure risk factor</td>
</tr>
<tr>
<td>ATSDR_INTERMEDIATE_TERM VALUE</td>
<td>ATSDR intermediate-term exposure risk factor</td>
</tr>
<tr>
<td>ATSDR_CHRONIC_VALUE</td>
<td>ATSDR chronic-term exposure risk factor</td>
</tr>
<tr>
<td>CAL_EPA_RELATIVE_EXPOSURE_LIMIT</td>
<td>California EPA Relative Exposure Limit factor</td>
</tr>
<tr>
<td>CAL_EPA_RELATIVE_EXPOSURE_LIMIT DURATION</td>
<td>Sample duration for the CAL EPA REL</td>
</tr>
<tr>
<td>NAAQS_1_HOURVAL</td>
<td>Value for the 1-hour National Ambient Air Quality Standard (NAAQS)</td>
</tr>
<tr>
<td>NAAQS_3_HOURVal</td>
<td>Value for the 3-hour NAAQS</td>
</tr>
<tr>
<td>NAAQS_8_HOURVal</td>
<td>Value for the 8-hour NAAQS</td>
</tr>
<tr>
<td>NAAQS_DAILYVAL</td>
<td>Value for the daily NAAQS</td>
</tr>
<tr>
<td>NAAQS_3_MONTH.Rolling.VAL</td>
<td>Value for the 3-month rolling average NAAQS</td>
</tr>
<tr>
<td>NAAQS_QUARTERLY.VAL</td>
<td>Value for the quarterly average NAAQS</td>
</tr>
<tr>
<td>NAAQS_ANNUAL.VAL</td>
<td>Value for the annual average NAAQS</td>
</tr>
<tr>
<td>COMMENT</td>
<td>General comment</td>
</tr>
</tbody>
</table>

* = 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 μg/m³. A total of 4,221 records are in this data dictionary.

Table 5-5. Sampling Methodology Information Data Fields

<table>
<thead>
<tr>
<th>Data Field</th>
<th>Data Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>*AQS_PARAMETER_CODE</td>
<td>AQS Parameter Identifier</td>
</tr>
<tr>
<td>PARAMETER_DESC</td>
<td>AQS Parameter Identifier Description</td>
</tr>
<tr>
<td>*AQS_METHODOLOGY_CODE</td>
<td>AQS Methodology Identifier</td>
</tr>
</tbody>
</table>
Table 5-5. Sampling Methodology Information Data Fields

<table>
<thead>
<tr>
<th>Data Field</th>
<th>Data Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>SAMPLE_COLLECTION_DESC</td>
<td>Sample Collection Description</td>
</tr>
<tr>
<td>SAMPLE_ANALYSIS_DESC</td>
<td>Sample Analysis Description</td>
</tr>
<tr>
<td>AQS_SAMPLE_DURATION_CODE</td>
<td>Duration Identifier</td>
</tr>
<tr>
<td>DURATION_DESC</td>
<td>Duration Identifier Description</td>
</tr>
<tr>
<td>AQS_UNIT_CODE</td>
<td>Unit of Measure Identifier</td>
</tr>
<tr>
<td>UNIT_DESC</td>
<td>Unit Description</td>
</tr>
<tr>
<td>FEDERAL_MDL_VALUE</td>
<td>Federal default method detection limit</td>
</tr>
<tr>
<td>FEDERAL_MDL_UNIT</td>
<td>Federal default method detection limit units</td>
</tr>
<tr>
<td>MDL_STD</td>
<td>Federal default method detection limit standardized to $\mu g/m^3$</td>
</tr>
<tr>
<td>SUMMARY_SCALE</td>
<td>AQS Field (unknown)</td>
</tr>
<tr>
<td>EQUIVALENT_METHOD_DESC</td>
<td>AQS Field (unknown)</td>
</tr>
<tr>
<td>REFERENCE_METHOD_ID</td>
<td>AQS Field (unknown)</td>
</tr>
<tr>
<td>COMMENT</td>
<td>General comment</td>
</tr>
</tbody>
</table>

* = primary key field

5.6 Date and Season Information

Table 5-6 presents data fields for every single day from 1990 to 2016 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 9,862 records are in this data dictionary.

Table 5-6. Date and Season Information Data Fields

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

* = 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 217 records are in this data dictionary.

<table>
<thead>
<tr>
<th>Data Field</th>
<th>Data Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>*AQS_QUALIFIER_CODE</td>
<td>Qualifier Identifier</td>
</tr>
<tr>
<td>QUALIFIER_DESC</td>
<td>Qualifier Description</td>
</tr>
<tr>
<td>QUALIFIER_TYPE</td>
<td>Type of Qualifier</td>
</tr>
<tr>
<td>QUALIFIER_TYPE_DESC</td>
<td>Type of Qualifier Description</td>
</tr>
</tbody>
</table>

* = 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>
<thead>
<tr>
<th>Data Field</th>
<th>Data Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>*AQS_DURATION_CODE</td>
<td>Duration Identifier</td>
</tr>
<tr>
<td>DURATION_DESC</td>
<td>Duration Identifier Description</td>
</tr>
<tr>
<td>DURATION_INDICATOR</td>
<td>Duration Indicator Identifier</td>
</tr>
<tr>
<td>DURATION_LENGTH</td>
<td>Length of sampling</td>
</tr>
<tr>
<td>DURATION_UNIT</td>
<td>Unit of length for sampling</td>
</tr>
</tbody>
</table>

* = 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 20 records are in this data dictionary.
Table 5-9. Unit Information Data Fields

<table>
<thead>
<tr>
<th>Data Field</th>
<th>Data Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>*AQS UNIT_CODE</td>
<td>Unit of Measure Identifier</td>
</tr>
<tr>
<td>UNIT_DESCRIPTION</td>
<td>Unit Description</td>
</tr>
<tr>
<td>UNIT_ABBR</td>
<td>Abbreviation of Units</td>
</tr>
<tr>
<td>REPORTED</td>
<td>Flag to identify if unit code is to be reported in the Output table</td>
</tr>
</tbody>
</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 24 records are in this data dictionary.

Table 5-10. Frequency Code Data Fields

<table>
<thead>
<tr>
<th>Data Field</th>
<th>Data Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>*AQS COLLECTION_FREQUENCY_CODE</td>
<td>Collection Frequency Code Identifier</td>
</tr>
<tr>
<td>COLLECTION_FREQUENCY_DESCRIPTION</td>
<td>Collection Frequency Code Description</td>
</tr>
<tr>
<td>DAILY_SAMPLE_NUMBER</td>
<td>Number of subdaily measurements (PAMS only)</td>
</tr>
<tr>
<td>DAILY_INTERVAL</td>
<td>Numeric equivalent of the collection frequency code</td>
</tr>
</tbody>
</table>

* = 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 42 records are in this data dictionary.

Table 5-11. Data Source Code Data Fields

<table>
<thead>
<tr>
<th>Data Field</th>
<th>Data Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>*DATA_SOURCE</td>
<td>Data Source Code Identifier</td>
</tr>
<tr>
<td>DATA_SOURCE_DESCRIPTION</td>
<td>Data Source Code Description</td>
</tr>
<tr>
<td>NUM_RECORDS</td>
<td>Number of data records</td>
</tr>
<tr>
<td>MIN_YEAR</td>
<td>First year for the data source</td>
</tr>
<tr>
<td>MAX_YEAR</td>
<td>End year for the data source</td>
</tr>
<tr>
<td>NUM_PARAMETER_CODE</td>
<td>Number of HAPs for the data source</td>
</tr>
<tr>
<td>NUM_SITES</td>
<td>Number of monitoring sites for the data source</td>
</tr>
<tr>
<td>NUM_STATES</td>
<td>Number of states for the data source</td>
</tr>
<tr>
<td>NUM_COUNTIES</td>
<td>Number of counties for the data source</td>
</tr>
</tbody>
</table>

* = 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 XII Archive of HAPs. In total, there are nearly 70 million data records.

**Table 6-1. Data Source Information for HAP Records**

<table>
<thead>
<tr>
<th>Data Source</th>
<th>Data Years</th>
<th># Sites</th>
<th># Pollutants/Parameters</th>
<th>HAP Data Record Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>AQS Data</td>
<td>1990-2016</td>
<td>2,290</td>
<td>365</td>
<td>52,532,573</td>
</tr>
<tr>
<td>Texas Commission on Environmental Quality</td>
<td>1992-2016</td>
<td>122</td>
<td>79</td>
<td>14,881,267</td>
</tr>
<tr>
<td>National Acid Deposition Program</td>
<td>2008-2016</td>
<td>165</td>
<td>4</td>
<td>1,406,086</td>
</tr>
<tr>
<td>PHASE V and Phase VII Archives</td>
<td>1991-2010</td>
<td>147</td>
<td>165</td>
<td>211,022</td>
</tr>
<tr>
<td>XACT Monitoring</td>
<td>2011-2015</td>
<td>8</td>
<td>10</td>
<td>207,842</td>
</tr>
<tr>
<td>Integrated Atmospheric Deposition Network</td>
<td>1999-2010</td>
<td>11</td>
<td>89</td>
<td>162,836</td>
</tr>
<tr>
<td>Multiple Air Toxics Exposure Studies (II, III, and IV)</td>
<td>1999-2013</td>
<td>17</td>
<td>59</td>
<td>160,293</td>
</tr>
<tr>
<td>NOAA Global Monitoring Division</td>
<td>1991-2016</td>
<td>17</td>
<td>7</td>
<td>158,780</td>
</tr>
<tr>
<td>Minnesota Pollution Control Agency</td>
<td>2008-2016</td>
<td>46</td>
<td>61</td>
<td>135,850</td>
</tr>
<tr>
<td>Sublette County, Wyoming</td>
<td>2009-2010</td>
<td>14</td>
<td>42</td>
<td>37,398</td>
</tr>
<tr>
<td>EPA Passive Sampling Study</td>
<td>2013-2015</td>
<td>17</td>
<td>9</td>
<td>18,675</td>
</tr>
<tr>
<td>Baldwin Hills Air Quality Study</td>
<td>2012-2013</td>
<td>1</td>
<td>15</td>
<td>7,146</td>
</tr>
<tr>
<td>City of Ft. Worth, TX</td>
<td>2010</td>
<td>8</td>
<td>49</td>
<td>5,455</td>
</tr>
<tr>
<td>Allegheny County, PA</td>
<td>2013-2016</td>
<td>3</td>
<td>14</td>
<td>5,251</td>
</tr>
<tr>
<td>NATTS Network Assessment</td>
<td>2010-2014</td>
<td>3</td>
<td>71</td>
<td>2,739</td>
</tr>
<tr>
<td>EPA Region III for WV</td>
<td>2008-2016</td>
<td>2</td>
<td>14</td>
<td>2,065</td>
</tr>
<tr>
<td>Maryland Department of the Environment</td>
<td>2014-2015</td>
<td>6</td>
<td>1</td>
<td>1,734</td>
</tr>
<tr>
<td>EPA's School Air Toxics</td>
<td>2011-2012</td>
<td>6</td>
<td>80</td>
<td>800</td>
</tr>
<tr>
<td><strong>Totals</strong></td>
<td><strong>1990-2016</strong></td>
<td><strong>2,565</strong></td>
<td><strong>378</strong></td>
<td><strong>69,937,812</strong></td>
</tr>
</tbody>
</table>

Less 22% of the data records are non-detects, while approximately 15% are null data records. It is important to note that null data records were not in EPA’s Phase V database, thus no conclusion can be made about null data records. Finally, less than 10% 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**

<table>
<thead>
<tr>
<th>Year</th>
<th># HAP Records</th>
<th># Non-Detect Records</th>
<th>% Non-Detect</th>
<th># Null Data Records</th>
<th>% Null</th>
<th># HAP Sample Values BMDL</th>
<th>% HAP Sample Values BMDL</th>
</tr>
</thead>
<tbody>
<tr>
<td>1990</td>
<td>142,126</td>
<td>62,942</td>
<td>44.3%</td>
<td>6,204</td>
<td>4.4%</td>
<td>14,034</td>
<td>9.9%</td>
</tr>
<tr>
<td>1991</td>
<td>175,666</td>
<td>78,892</td>
<td>44.9%</td>
<td>5,749</td>
<td>3.3%</td>
<td>18,783</td>
<td>10.7%</td>
</tr>
<tr>
<td>1992</td>
<td>207,973</td>
<td>88,642</td>
<td>42.6%</td>
<td>11,238</td>
<td>5.4%</td>
<td>25,415</td>
<td>12.2%</td>
</tr>
<tr>
<td>1993</td>
<td>286,397</td>
<td>101,352</td>
<td>35.4%</td>
<td>19,990</td>
<td>7.0%</td>
<td>32,891</td>
<td>11.5%</td>
</tr>
</tbody>
</table>
Table 6-2. HAP Summary Counts by Year

<table>
<thead>
<tr>
<th>Year</th>
<th># HAP Records</th>
<th># Non-Detect Records</th>
<th>% Non-Detect</th>
<th># Null Data Records</th>
<th>% Null</th>
<th># HAP Sample Values BMDL</th>
<th>% HAP Sample Values BMDL</th>
</tr>
</thead>
<tbody>
<tr>
<td>1994</td>
<td>488,682</td>
<td>140,009</td>
<td>28.7%</td>
<td>32,678</td>
<td>6.7%</td>
<td>42,348</td>
<td>8.7%</td>
</tr>
<tr>
<td>1995</td>
<td>880,780</td>
<td>214,170</td>
<td>24.3%</td>
<td>91,015</td>
<td>10.3%</td>
<td>54,672</td>
<td>6.2%</td>
</tr>
<tr>
<td>1996</td>
<td>1,129,426</td>
<td>252,913</td>
<td>22.4%</td>
<td>158,842</td>
<td>14.1%</td>
<td>82,487</td>
<td>7.3%</td>
</tr>
<tr>
<td>1997</td>
<td>1,326,261</td>
<td>278,134</td>
<td>21.0%</td>
<td>168,074</td>
<td>12.7%</td>
<td>84,222</td>
<td>6.4%</td>
</tr>
<tr>
<td>1998</td>
<td>1,536,342</td>
<td>302,017</td>
<td>19.7%</td>
<td>226,214</td>
<td>14.7%</td>
<td>95,976</td>
<td>6.2%</td>
</tr>
<tr>
<td>1999</td>
<td>1,673,443</td>
<td>337,239</td>
<td>20.2%</td>
<td>305,316</td>
<td>18.2%</td>
<td>101,463</td>
<td>6.1%</td>
</tr>
<tr>
<td>2000</td>
<td>1,823,136</td>
<td>407,314</td>
<td>22.3%</td>
<td>271,915</td>
<td>14.9%</td>
<td>139,521</td>
<td>7.7%</td>
</tr>
<tr>
<td>2001</td>
<td>2,159,662</td>
<td>477,518</td>
<td>22.1%</td>
<td>347,964</td>
<td>16.1%</td>
<td>223,818</td>
<td>9.9%</td>
</tr>
<tr>
<td>2002</td>
<td>2,252,877</td>
<td>531,850</td>
<td>23.6%</td>
<td>342,224</td>
<td>15.2%</td>
<td>273,759</td>
<td>10.1%</td>
</tr>
<tr>
<td>2003</td>
<td>2,479,804</td>
<td>542,650</td>
<td>21.9%</td>
<td>397,255</td>
<td>16.0%</td>
<td>215,562</td>
<td>8.7%</td>
</tr>
<tr>
<td>2004</td>
<td>2,978,195</td>
<td>640,708</td>
<td>21.5%</td>
<td>491,418</td>
<td>16.5%</td>
<td>252,356</td>
<td>8.5%</td>
</tr>
<tr>
<td>2005</td>
<td>3,437,633</td>
<td>722,814</td>
<td>21.0%</td>
<td>584,850</td>
<td>17.0%</td>
<td>308,949</td>
<td>9.0%</td>
</tr>
<tr>
<td>2006</td>
<td>3,472,362</td>
<td>744,640</td>
<td>21.4%</td>
<td>539,275</td>
<td>15.5%</td>
<td>301,930</td>
<td>8.7%</td>
</tr>
<tr>
<td>2007</td>
<td>3,655,404</td>
<td>762,686</td>
<td>20.9%</td>
<td>489,670</td>
<td>13.4%</td>
<td>299,994</td>
<td>8.2%</td>
</tr>
<tr>
<td>2008</td>
<td>3,628,561</td>
<td>782,957</td>
<td>21.6%</td>
<td>554,306</td>
<td>15.3%</td>
<td>273,759</td>
<td>7.5%</td>
</tr>
<tr>
<td>2009</td>
<td>3,841,448</td>
<td>857,908</td>
<td>22.3%</td>
<td>498,573</td>
<td>13.0%</td>
<td>349,900</td>
<td>9.1%</td>
</tr>
<tr>
<td>2010</td>
<td>3,976,020</td>
<td>879,438</td>
<td>22.1%</td>
<td>558,994</td>
<td>14.1%</td>
<td>401,585</td>
<td>10.1%</td>
</tr>
<tr>
<td>2011</td>
<td>4,130,186</td>
<td>960,458</td>
<td>23.3%</td>
<td>604,882</td>
<td>14.6%</td>
<td>408,779</td>
<td>9.9%</td>
</tr>
<tr>
<td>2012</td>
<td>4,405,802</td>
<td>922,899</td>
<td>20.9%</td>
<td>600,430</td>
<td>13.6%</td>
<td>465,965</td>
<td>10.6%</td>
</tr>
<tr>
<td>2013</td>
<td>4,679,865</td>
<td>1,004,152</td>
<td>21.5%</td>
<td>661,457</td>
<td>14.1%</td>
<td>472,802</td>
<td>10.1%</td>
</tr>
<tr>
<td>2014</td>
<td>5,105,617</td>
<td>1,068,311</td>
<td>20.9%</td>
<td>691,285</td>
<td>13.5%</td>
<td>497,084</td>
<td>9.7%</td>
</tr>
<tr>
<td>2015</td>
<td>4,981,803</td>
<td>1,022,432</td>
<td>20.5%</td>
<td>676,126</td>
<td>13.6%</td>
<td>488,378</td>
<td>9.8%</td>
</tr>
<tr>
<td>2016</td>
<td>5,082,341</td>
<td>1,051,992</td>
<td>20.7%</td>
<td>667,705</td>
<td>13.1%</td>
<td>552,832</td>
<td>10.9%</td>
</tr>
<tr>
<td>Totals</td>
<td>69,937,812</td>
<td>15,237,037</td>
<td>21.8%</td>
<td>10,003,649</td>
<td>14.3%</td>
<td>6,384,685</td>
<td>9.1%</td>
</tr>
</tbody>
</table>

Of the 15,237,037 HAP non-detects in the master database, approximately 3% (460,858 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

<table>
<thead>
<tr>
<th>State</th>
<th>Total # of ND</th>
<th>Total # Surrogate</th>
<th># Fed MDL Surrogate</th>
<th># Entity-Provided MDL Surrogates</th>
<th>Time Period of Surrogates</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alabama</td>
<td>77,783</td>
<td>2</td>
<td>2</td>
<td>0</td>
<td>1993-1996</td>
</tr>
<tr>
<td>Alaska</td>
<td>59,551</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>NA</td>
</tr>
<tr>
<td>Arizona</td>
<td>183,618</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>NA</td>
</tr>
<tr>
<td>Arkansas</td>
<td>26,242</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>NA</td>
</tr>
<tr>
<td>California</td>
<td>958,594</td>
<td>375,508</td>
<td>253,660</td>
<td>121,848</td>
<td>1990-2016</td>
</tr>
<tr>
<td>Colorado</td>
<td>153,225</td>
<td>31</td>
<td>31</td>
<td>0</td>
<td>2002</td>
</tr>
<tr>
<td>Connecticut</td>
<td>277,696</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>NA</td>
</tr>
<tr>
<td>Delaware</td>
<td>83,632</td>
<td>193</td>
<td>32</td>
<td>161</td>
<td>2000-2012</td>
</tr>
</tbody>
</table>
## Table 6-3. Non-Detect Records Populated with ½ MDL by State

<table>
<thead>
<tr>
<th>State</th>
<th>Total # of ND</th>
<th>Total # Surrogate</th>
<th># Fed MDL Surrogate</th>
<th># Entity-Provided MDL Surrogates</th>
<th>Time Period of Surrogates</th>
</tr>
</thead>
<tbody>
<tr>
<td>District of Columbia</td>
<td>145,471</td>
<td>100</td>
<td>77</td>
<td>23</td>
<td>1997-2008</td>
</tr>
<tr>
<td>Florida</td>
<td>174,464</td>
<td>14,640</td>
<td>122</td>
<td>14,518</td>
<td>1990-2006</td>
</tr>
<tr>
<td>Georgia</td>
<td>585,466</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>NA</td>
</tr>
<tr>
<td>Hawaii</td>
<td>34,175</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>NA</td>
</tr>
<tr>
<td>Idaho</td>
<td>53,358</td>
<td>10,621</td>
<td>0</td>
<td>10,621</td>
<td>2002-2008</td>
</tr>
<tr>
<td>Illinois</td>
<td>497,105</td>
<td>18</td>
<td>17</td>
<td>1</td>
<td>2005</td>
</tr>
<tr>
<td>Indiana</td>
<td>383,639</td>
<td>31</td>
<td>31</td>
<td>0</td>
<td>1990</td>
</tr>
<tr>
<td>Iowa</td>
<td>81,230</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>NA</td>
</tr>
<tr>
<td>Kansas</td>
<td>132,946</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1990</td>
</tr>
<tr>
<td>Kentucky</td>
<td>96,377</td>
<td>655</td>
<td>0</td>
<td>655</td>
<td>2006-2007</td>
</tr>
<tr>
<td>Louisiana</td>
<td>168,016</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1999</td>
</tr>
<tr>
<td>Maine</td>
<td>784,860</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>NA</td>
</tr>
<tr>
<td>Maryland</td>
<td>179,960</td>
<td>392</td>
<td>392</td>
<td>0</td>
<td>1997-2000</td>
</tr>
<tr>
<td>Massachusetts</td>
<td>394,960</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>NA</td>
</tr>
<tr>
<td>Michigan</td>
<td>358,935</td>
<td>57</td>
<td>57</td>
<td>0</td>
<td>1992-1994</td>
</tr>
<tr>
<td>Minnesota</td>
<td>460,603</td>
<td>5</td>
<td>5</td>
<td>0</td>
<td>1999</td>
</tr>
<tr>
<td>Mississippi</td>
<td>81,310</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>NA</td>
</tr>
<tr>
<td>Missouri</td>
<td>196,601</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>NA</td>
</tr>
<tr>
<td>Montana</td>
<td>107,596</td>
<td>2</td>
<td>2</td>
<td>0</td>
<td>1991</td>
</tr>
<tr>
<td>Nebraska</td>
<td>22,807</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>NA</td>
</tr>
<tr>
<td>Nevada</td>
<td>43,160</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>NA</td>
</tr>
<tr>
<td>New Hampshire</td>
<td>288,886</td>
<td>110</td>
<td>110</td>
<td>0</td>
<td>2002-2004</td>
</tr>
<tr>
<td>New Jersey</td>
<td>275,765</td>
<td>15</td>
<td>15</td>
<td>0</td>
<td>1993-2005</td>
</tr>
<tr>
<td>New Mexico</td>
<td>63,401</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>NA</td>
</tr>
<tr>
<td>New York</td>
<td>261,198</td>
<td>9,842</td>
<td>9,842</td>
<td>0</td>
<td>1990-1999</td>
</tr>
<tr>
<td>North Carolina</td>
<td>140,085</td>
<td>1,140</td>
<td>1,112</td>
<td>28</td>
<td>2002-2008</td>
</tr>
<tr>
<td>North Dakota</td>
<td>39,532</td>
<td>2</td>
<td>2</td>
<td>0</td>
<td>2000</td>
</tr>
<tr>
<td>Ohio</td>
<td>203,118</td>
<td>3</td>
<td>0</td>
<td>3</td>
<td>2004-2005</td>
</tr>
<tr>
<td>Oklahoma</td>
<td>101,865</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>NA</td>
</tr>
<tr>
<td>Pennsylvania</td>
<td>516,189</td>
<td>1,086</td>
<td>850</td>
<td>236</td>
<td>2000-2013</td>
</tr>
<tr>
<td>Rhode Island</td>
<td>195,525</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>NA</td>
</tr>
<tr>
<td>South Carolina</td>
<td>212,059</td>
<td>16</td>
<td>16</td>
<td>0</td>
<td>1993-1994</td>
</tr>
<tr>
<td>South Dakota</td>
<td>64,702</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>NA</td>
</tr>
<tr>
<td>Tennessee</td>
<td>63,978</td>
<td>138</td>
<td>138</td>
<td>0</td>
<td>1990-1998</td>
</tr>
<tr>
<td>Texas</td>
<td>5,105,429</td>
<td>2,051</td>
<td>2,051</td>
<td>0</td>
<td>1994-2009</td>
</tr>
<tr>
<td>Utah</td>
<td>92,884</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>NA</td>
</tr>
<tr>
<td>Vermont</td>
<td>120,468</td>
<td>9</td>
<td>8</td>
<td>1</td>
<td>1995-2002</td>
</tr>
<tr>
<td>Virginia</td>
<td>126,305</td>
<td>244</td>
<td>113</td>
<td>131</td>
<td>2000-2012</td>
</tr>
<tr>
<td>Washington</td>
<td>137,149</td>
<td>4,619</td>
<td>4</td>
<td>4,615</td>
<td>1995-2006</td>
</tr>
<tr>
<td>West Virginia</td>
<td>30,336</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>NA</td>
</tr>
<tr>
<td>Wisconsin</td>
<td>125,913</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>NA</td>
</tr>
<tr>
<td>Wyoming</td>
<td>85,183</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>NA</td>
</tr>
<tr>
<td>Puerto Rico</td>
<td>13,981</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>NA</td>
</tr>
<tr>
<td>Virgin Islands</td>
<td>9,326</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>NA</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>15,237,037</strong></td>
<td><strong>460,858</strong></td>
<td><strong>270,807</strong></td>
<td><strong>190,051</strong></td>
<td><strong>1990-2016</strong></td>
</tr>
</tbody>
</table>
In the Phase XII database, data has been stored with native sample durations, as presented in Table 6-4.

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

<table>
<thead>
<tr>
<th>Year(s)</th>
<th>Sub-Hourly</th>
<th>1-hour</th>
<th>2-hour</th>
<th>3-hour</th>
<th>4-hour</th>
<th>5-hour</th>
<th>6-hour</th>
<th>8-hour</th>
<th>12-hour</th>
<th>Daily Records</th>
<th>Monthly/Variable Records</th>
</tr>
</thead>
<tbody>
<tr>
<td>1990</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>756</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>400</td>
<td>140,890</td>
<td>80</td>
</tr>
<tr>
<td>1991</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>493</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
<td>175,161</td>
<td>10</td>
</tr>
<tr>
<td>1992</td>
<td>21</td>
<td>0</td>
<td>0</td>
<td>1,302</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
<td>206,650</td>
<td>0</td>
</tr>
<tr>
<td>1993</td>
<td>39</td>
<td>38,579</td>
<td>0</td>
<td>21,401</td>
<td>0</td>
<td>0</td>
<td>872</td>
<td>0</td>
<td>0</td>
<td>225,506</td>
<td>0</td>
</tr>
<tr>
<td>1994</td>
<td>98</td>
<td>154,837</td>
<td>0</td>
<td>59,000</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
<td>274,747</td>
<td>0</td>
</tr>
<tr>
<td>1995</td>
<td>275</td>
<td>482,527</td>
<td>0</td>
<td>84,192</td>
<td>2,088</td>
<td>0</td>
<td>133</td>
<td>0</td>
<td>0</td>
<td>311,565</td>
<td>0</td>
</tr>
<tr>
<td>1996</td>
<td>425</td>
<td>659,926</td>
<td>0</td>
<td>120,852</td>
<td>6,876</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
<td>341,347</td>
<td>0</td>
</tr>
<tr>
<td>1997</td>
<td>567</td>
<td>837,469</td>
<td>0</td>
<td>120,476</td>
<td>3,843</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>363,906</td>
<td>0</td>
</tr>
<tr>
<td>1998</td>
<td>704</td>
<td>1,013,689</td>
<td>0</td>
<td>154,287</td>
<td>2,799</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
<td>364,863</td>
<td>0</td>
</tr>
<tr>
<td>1999</td>
<td>1,509</td>
<td>1,088,437</td>
<td>0</td>
<td>154,112</td>
<td>0</td>
<td>0</td>
<td>2,130</td>
<td>0</td>
<td>0</td>
<td>425,536</td>
<td>1,719</td>
</tr>
<tr>
<td>2000</td>
<td>1,250</td>
<td>1,171,823</td>
<td>0</td>
<td>137,269</td>
<td>1,797</td>
<td>0</td>
<td>1,578</td>
<td>0</td>
<td></td>
<td>507,463</td>
<td>1,956</td>
</tr>
<tr>
<td>2001</td>
<td>1,068</td>
<td>1,303,216</td>
<td>0</td>
<td>135,038</td>
<td>5,879</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>6,092</td>
<td>705,712</td>
<td>2,657</td>
</tr>
<tr>
<td>2002</td>
<td>1,608</td>
<td>1,235,225</td>
<td>0</td>
<td>134,088</td>
<td>10,664</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>4,290</td>
<td>863,847</td>
<td>3,155</td>
</tr>
<tr>
<td>2003</td>
<td>1,911</td>
<td>1,403,818</td>
<td>0</td>
<td>116,193</td>
<td>9,641</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2,262</td>
<td>942,331</td>
<td>3,648</td>
</tr>
<tr>
<td>2004</td>
<td>91,618</td>
<td>1,698,222</td>
<td>0</td>
<td>100,965</td>
<td>17,659</td>
<td>0</td>
<td>0</td>
<td>1,648</td>
<td>1,108</td>
<td>1,062,899</td>
<td>4,076</td>
</tr>
<tr>
<td>2005</td>
<td>103,315</td>
<td>2,013,926</td>
<td>0</td>
<td>104,265</td>
<td>14,526</td>
<td>0</td>
<td>0</td>
<td>7,458</td>
<td>1,189</td>
<td>1,189,862</td>
<td>4,281</td>
</tr>
<tr>
<td>2006</td>
<td>102,091</td>
<td>2,162,792</td>
<td>0</td>
<td>113,262</td>
<td>5,073</td>
<td>0</td>
<td>0</td>
<td>2,264</td>
<td>0</td>
<td>1,082,207</td>
<td>4,673</td>
</tr>
<tr>
<td>2007</td>
<td>320,891</td>
<td>2,167,974</td>
<td>0</td>
<td>125,786</td>
<td>0</td>
<td>0</td>
<td>2,020</td>
<td>0</td>
<td>0</td>
<td>1,033,703</td>
<td>5,030</td>
</tr>
<tr>
<td>2008</td>
<td>384,251</td>
<td>2,095,297</td>
<td>4,053</td>
<td>111,048</td>
<td>18</td>
<td>6</td>
<td>2,015</td>
<td>0</td>
<td>1,975</td>
<td>1,024,295</td>
<td>5,603</td>
</tr>
<tr>
<td>2009</td>
<td>385,821</td>
<td>2,171,025</td>
<td>86,802</td>
<td>113,564</td>
<td>1,077</td>
<td>384</td>
<td>3</td>
<td>0</td>
<td>1,089</td>
<td>1,075,995</td>
<td>5,688</td>
</tr>
<tr>
<td>2010</td>
<td>416,969</td>
<td>2,323,679</td>
<td>89,757</td>
<td>117,660</td>
<td>1,149</td>
<td>330</td>
<td>0</td>
<td>0</td>
<td>1,134</td>
<td>1,019,761</td>
<td>5,581</td>
</tr>
<tr>
<td>2011</td>
<td>298,184</td>
<td>2,669,063</td>
<td>86,560</td>
<td>110,691</td>
<td>687</td>
<td>225</td>
<td>26</td>
<td>0</td>
<td>0</td>
<td>959,377</td>
<td>5,373</td>
</tr>
<tr>
<td>2012</td>
<td>265,150</td>
<td>3,010,018</td>
<td>45,897</td>
<td>105,165</td>
<td>930</td>
<td>69</td>
<td>218</td>
<td>0</td>
<td>0</td>
<td>972,880</td>
<td>5,475</td>
</tr>
<tr>
<td>2013</td>
<td>188,544</td>
<td>3,368,041</td>
<td>44,916</td>
<td>99,790</td>
<td>264</td>
<td>93</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>973,906</td>
<td>4,311</td>
</tr>
<tr>
<td>2014</td>
<td>304,797</td>
<td>3,665,461</td>
<td>65,490</td>
<td>100,772</td>
<td>345</td>
<td>27</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>955,792</td>
<td>12,933</td>
</tr>
<tr>
<td>2015</td>
<td>326,623</td>
<td>3,555,559</td>
<td>71,964</td>
<td>79,996</td>
<td>6</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>946,221</td>
<td>1,431</td>
</tr>
<tr>
<td>2016</td>
<td>419,238</td>
<td>3,694,904</td>
<td>38,187</td>
<td>40,997</td>
<td>0</td>
<td>0</td>
<td>456</td>
<td>352</td>
<td>0</td>
<td>888,207</td>
<td>0</td>
</tr>
</tbody>
</table>

**Totals** 3,616,969 43,985,507 533,626 2,563,420 85,321 1,137 5,287 15,534 18,702 19,034,629 77,680

### 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 XII raw table are presented in Table 7-1. Primary key fields are denoted by a “*”.

### Table 7-1. Ambient Monitoring Archive Output Fields

<table>
<thead>
<tr>
<th>Data Field</th>
<th>Data Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>STATE_ABBR</td>
<td>Two-letter abbreviation for the state with the monitoring site</td>
</tr>
<tr>
<td>*AMA_SITE_CODE</td>
<td>Ambient Monitoring Archive (AMA) Site Code</td>
</tr>
<tr>
<td>*AQS_POC</td>
<td>Parameter Occurrence Code (POC)</td>
</tr>
<tr>
<td>Data Field</td>
<td>Data Description</td>
</tr>
<tr>
<td>--------------------------------</td>
<td>----------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>PROGRAM</td>
<td>Identifies Monitoring Program, if available</td>
</tr>
<tr>
<td>YEAR</td>
<td>Year of sampling date</td>
</tr>
<tr>
<td>QUARTER</td>
<td>Calendar quarter of the sampling date</td>
</tr>
<tr>
<td>*SAMPLE_DATE</td>
<td>Date Sample was taken</td>
</tr>
<tr>
<td>*SAMPLE_START_TIME</td>
<td>Time at which sample began</td>
</tr>
<tr>
<td>*AQS_PARAMETER_CODE</td>
<td>Air Quality System (AQS) Pollutant Code</td>
</tr>
<tr>
<td>AQS_PARAMETER_NAME</td>
<td>AQS pollutant name</td>
</tr>
<tr>
<td>DATA_SOURCE</td>
<td>Identifies the source of the data record</td>
</tr>
<tr>
<td>DURATION_DESC</td>
<td>Translated AQS Sample Duration description</td>
</tr>
<tr>
<td>SAMPLE_VALUE_REPORTED</td>
<td>Reported sample value from the data source</td>
</tr>
<tr>
<td>AQS_UNIT_CODE</td>
<td>Unit of Measure Code for the native sample value</td>
</tr>
<tr>
<td>UNIT_DESC</td>
<td>Translated AQS Unit of Measure description</td>
</tr>
<tr>
<td>SAMPLING_FREQUENCY_CODE</td>
<td>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; 90=Every90Days; A=PAMS Daily; H or J=PAMS 3rdDay; P=PAMS 6thDay; S=Seasonal; Y=TwicePerWeek; Z=Every9Days)</td>
</tr>
<tr>
<td>COMMENT</td>
<td>Reserved for comments</td>
</tr>
<tr>
<td>SAMPLE_VALUE_STD_FINAL_UG_M3</td>
<td>Concentration value standardized to µg/m³, local conditions</td>
</tr>
<tr>
<td>SAMPLE_VALUE_STD_FINAL_TYPE</td>
<td>Final Concentration type for analysis (L = Local Conditions, S = Standard Conditions)</td>
</tr>
<tr>
<td>AQS_PARAMETER_CODE_FINAL</td>
<td>Final AQS Pollutant Code for analysis</td>
</tr>
<tr>
<td>AQS_PARAMETER_NAME_FINAL</td>
<td>Final AQS Pollutant Name for analysis</td>
</tr>
<tr>
<td>ALTERNATE_MDL_REPORTED</td>
<td>Reported MDL in native units</td>
</tr>
<tr>
<td>MDL_STD_UG_M3</td>
<td>MDL standardized to µg/m³</td>
</tr>
<tr>
<td>MDL_TYPE</td>
<td>Identifies the source of the standardized MDL</td>
</tr>
<tr>
<td>AQS_NULL_DATA_CODE</td>
<td>Data Qualifier code for null sample values</td>
</tr>
<tr>
<td>AQS_QUALIFIER_01</td>
<td>Data Qualifier code field 1</td>
</tr>
<tr>
<td>AQS_QUALIFIER_02</td>
<td>Data Qualifier code field 2</td>
</tr>
<tr>
<td>AQS_QUALIFIER_03</td>
<td>Data Qualifier code field 3</td>
</tr>
<tr>
<td>AQS_QUALIFIER_04</td>
<td>Data Qualifier code field 4</td>
</tr>
<tr>
<td>AQS_QUALIFIER_05</td>
<td>Data Qualifier code field 5</td>
</tr>
<tr>
<td>AQS_QUALIFIER_06</td>
<td>Data Qualifier code field 6</td>
</tr>
<tr>
<td>AQS_QUALIFIER_07</td>
<td>Data Qualifier code field 7</td>
</tr>
<tr>
<td>AQS_QUALIFIER_08</td>
<td>Data Qualifier code field 8</td>
</tr>
<tr>
<td>AQS_QUALIFIER_09</td>
<td>Data Qualifier code field 9</td>
</tr>
<tr>
<td>AQS_QUALIFIER_10</td>
<td>Data Qualifier code field 10</td>
</tr>
<tr>
<td>AQS_METHOD_CODE</td>
<td>Sampling and Analysis Method Code</td>
</tr>
<tr>
<td>SAMPLE_COLLECTION_DESC</td>
<td>Translated AQS Sampling Collection description</td>
</tr>
<tr>
<td>SAMPLE_ANALYSIS_DESC</td>
<td>Translated AQS Analysis Method description</td>
</tr>
<tr>
<td>BELOW_MDL_FLAG</td>
<td>Identifies if the concentration record is a non-detect</td>
</tr>
<tr>
<td>CENSUS_TRACT_ID</td>
<td>U.S. Census tract identifier in which the monitoring site is located</td>
</tr>
<tr>
<td>CENSUS_TRACT_POPULATION_2010</td>
<td>2010 Population within the census tract</td>
</tr>
<tr>
<td>MONITOR_LATITUDE</td>
<td>Y-Coordinate Value in decimal degrees</td>
</tr>
<tr>
<td>MONITOR_LONGITUDE</td>
<td>X-Coordinate Value in decimal degrees</td>
</tr>
</tbody>
</table>

* = 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. Nearly 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:

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

| 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                         | 285,086           | 281,238          | 3,848            | 98.65%           |
| Alaska                          | 191,539           | 180,227          | 11,312           | 94.09%           |
| Arizona                         | 528,021           | 527,398          | 623              | 99.88%           |
| Arkansas                        | 63,640            | 63,503           | 137              | 99.78%           |
| California                      | 4,082,104         | 3,873,142        | 208,962          | 94.88%           |
| Colorado                        | 444,394           | 414,584          | 29,810           | 93.29%           |
| Connecticut                     | 1,238,858         | 1,235,038        | 3,820            | 99.69%           |
| Delaware                        | 265,327           | 261,098          | 4,229            | 98.41%           |
| District of Columbia            | 554,902           | 553,577          | 1,325            | 99.76%           |
| Florida                         | 728,354           | 725,110          | 3,244            | 99.55%           |
| Georgia                         | 1,868,276         | 1,867,359        | 917              | 99.95%           |
| Hawaii                          | 171,536           | 142,559          | 28,977           | 83.11%           |
| Idaho                           | 106,446           | 105,855          | 591              | 99.44%           |
| Illinois                        | 1,225,345         | 1,163,440        | 61,905           | 94.95%           |
| Indiana                         | 2,282,409         | 2,266,024        | 16,385           | 99.28%           |
| Iowa                            | 187,216           | 171,428          | 15,788           | 91.57%           |
| Kansas                          | 218,193           | 185,032          | 33,161           | 84.80%           |
| Kentucky                        | 357,765           | 357,739          | 26               | 99.99%           |
| Louisiana                       | 755,405           | 750,607          | 4,798            | 99.36%           |
| Maine                           | 1,745,434         | 1,733,236        | 12,198           | 99.30%           |
| Maryland                        | 1,100,376         | 997,166          | 103,210          | 90.62%           |
| Massachusetts                   | 1,852,240         | 1,844,366        | 7,874            | 99.57%           |
| Michigan                        | 1,205,872         | 1,176,547        | 29,325           | 97.57%           |
| Minnesota                       | 1,143,832         | 1,115,210        | 28,622           | 97.50%           |
| Mississippi                     | 268,013           | 267,536          | 477              | 99.82%           |
| Missouri                        | 841,367           | 807,152          | 34,215           | 95.93%           |
| Montana                         | 233,867           | 224,132          | 9,735            | 95.84%           |
| Nebraska                        | 62,508            | 57,946           | 4,562            | 92.70%           |
| Nevada                          | 88,411            | 88,411           | 0                | 100.00%          |
| New Hampshire                   | 779,339           | 771,592          | 7,747            | 99.01%           |
| New Jersey                      | 1,286,390         | 1,281,616        | 4,774            | 99.63%           |
| New Mexico                      | 161,697           | 160,985          | 712              | 99.56%           |
Table 7-2. Summary of Output Record Counts by State

<table>
<thead>
<tr>
<th>State</th>
<th>Total # Output Records</th>
<th>Total # Local Condition Records</th>
<th>Total # Standard Condition Records</th>
<th>% Local Condition Records</th>
</tr>
</thead>
<tbody>
<tr>
<td>New York</td>
<td>1,679,957</td>
<td>1,618,151</td>
<td>61,806</td>
<td>96.32%</td>
</tr>
<tr>
<td>North Carolina</td>
<td>426,512</td>
<td>424,619</td>
<td>1,893</td>
<td>99.56%</td>
</tr>
<tr>
<td>North Dakota</td>
<td>79,396</td>
<td>79,396</td>
<td>0</td>
<td>100.00%</td>
</tr>
<tr>
<td>Ohio</td>
<td>505,876</td>
<td>489,858</td>
<td>16,018</td>
<td>96.83%</td>
</tr>
<tr>
<td>Oklahoma</td>
<td>300,585</td>
<td>299,848</td>
<td>737</td>
<td>99.75%</td>
</tr>
<tr>
<td>Oregon</td>
<td>370,482</td>
<td>364,944</td>
<td>5,538</td>
<td>98.51%</td>
</tr>
<tr>
<td>Pennsylvania</td>
<td>1,367,344</td>
<td>1,324,513</td>
<td>42,831</td>
<td>96.87%</td>
</tr>
<tr>
<td>Rhode Island</td>
<td>752,509</td>
<td>752,495</td>
<td>14</td>
<td>100.00%</td>
</tr>
<tr>
<td>South Carolina</td>
<td>534,128</td>
<td>416,684</td>
<td>117,444</td>
<td>78.01%</td>
</tr>
<tr>
<td>South Dakota</td>
<td>130,588</td>
<td>130,588</td>
<td>0</td>
<td>100.00%</td>
</tr>
<tr>
<td>Tennessee</td>
<td>176,440</td>
<td>166,128</td>
<td>10,312</td>
<td>94.16%</td>
</tr>
<tr>
<td>Texas</td>
<td>31,287,934</td>
<td>31,249,545</td>
<td>38,389</td>
<td>99.88%</td>
</tr>
<tr>
<td>Utah</td>
<td>300,924</td>
<td>299,916</td>
<td>1,008</td>
<td>99.67%</td>
</tr>
<tr>
<td>Vermont</td>
<td>756,968</td>
<td>326,740</td>
<td>430,228</td>
<td>43.16%</td>
</tr>
<tr>
<td>Virginia</td>
<td>587,661</td>
<td>585,701</td>
<td>1,960</td>
<td>99.67%</td>
</tr>
<tr>
<td>Washington</td>
<td>393,366</td>
<td>391,825</td>
<td>1,541</td>
<td>99.61%</td>
</tr>
<tr>
<td>West Virginia</td>
<td>138,353</td>
<td>130,964</td>
<td>7,389</td>
<td>94.66%</td>
</tr>
<tr>
<td>Wisconsin</td>
<td>3,386,068</td>
<td>3,351,668</td>
<td>34,400</td>
<td>98.98%</td>
</tr>
<tr>
<td>Wyoming</td>
<td>157,484</td>
<td>155,667</td>
<td>1,817</td>
<td>98.85%</td>
</tr>
<tr>
<td>Puerto Rico</td>
<td>31,465</td>
<td>31,284</td>
<td>176</td>
<td>99.44%</td>
</tr>
<tr>
<td>Virgin Islands</td>
<td>27,324</td>
<td>27,324</td>
<td>0</td>
<td>100.00%</td>
</tr>
<tr>
<td>TOTALS</td>
<td>69,715,526</td>
<td>68,268,716</td>
<td>1,446,810</td>
<td>97.92%</td>
</tr>
</tbody>
</table>

Table 7-3 presents a summary of the final counts in the Output files by year. From 2001 to 2016, approximately 98.61% of the data records are in local conditions.

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

<table>
<thead>
<tr>
<th>Year</th>
<th>Total # Output Records</th>
<th>Total # Local Condition Records</th>
<th>Total # Standard Condition Records</th>
<th>% Local Concentration Records</th>
</tr>
</thead>
<tbody>
<tr>
<td>1990</td>
<td>140,823</td>
<td>92,269</td>
<td>48,554</td>
<td>65.52%</td>
</tr>
<tr>
<td>1991</td>
<td>174,334</td>
<td>119,635</td>
<td>54,699</td>
<td>68.62%</td>
</tr>
<tr>
<td>1992</td>
<td>206,429</td>
<td>145,118</td>
<td>61,311</td>
<td>70.30%</td>
</tr>
<tr>
<td>1993</td>
<td>285,701</td>
<td>202,939</td>
<td>82,762</td>
<td>71.03%</td>
</tr>
<tr>
<td>1994</td>
<td>486,616</td>
<td>392,505</td>
<td>94,111</td>
<td>80.66%</td>
</tr>
<tr>
<td>1995</td>
<td>879,509</td>
<td>784,917</td>
<td>94,592</td>
<td>89.24%</td>
</tr>
<tr>
<td>1996</td>
<td>1,127,509</td>
<td>1,040,288</td>
<td>87,221</td>
<td>92.26%</td>
</tr>
<tr>
<td>1997</td>
<td>1,321,453</td>
<td>1,279,980</td>
<td>41,473</td>
<td>96.86%</td>
</tr>
<tr>
<td>1998</td>
<td>1,532,026</td>
<td>1,508,233</td>
<td>23,793</td>
<td>98.45%</td>
</tr>
<tr>
<td>1999</td>
<td>1,666,764</td>
<td>1,654,805</td>
<td>11,959</td>
<td>99.28%</td>
</tr>
<tr>
<td>2000</td>
<td>1,817,826</td>
<td>1,807,442</td>
<td>10,384</td>
<td>99.43%</td>
</tr>
<tr>
<td>2001</td>
<td>2,154,472</td>
<td>2,143,607</td>
<td>10,865</td>
<td>99.50%</td>
</tr>
<tr>
<td>2002</td>
<td>2,246,094</td>
<td>2,226,360</td>
<td>19,734</td>
<td>99.12%</td>
</tr>
<tr>
<td>2003</td>
<td>2,472,307</td>
<td>2,451,895</td>
<td>20,412</td>
<td>99.17%</td>
</tr>
</tbody>
</table>
### Table 7-3. Summary of Output Record Counts by Year

<table>
<thead>
<tr>
<th>Year</th>
<th>Total # Output Records</th>
<th>Total # Local Condition Records</th>
<th>Total # Standard Condition Records</th>
<th>% Local Concentration Records</th>
</tr>
</thead>
<tbody>
<tr>
<td>2004</td>
<td>2,968,920</td>
<td>2,953,994</td>
<td>14,926</td>
<td>99.50%</td>
</tr>
<tr>
<td>2005</td>
<td>3,413,593</td>
<td>3,402,593</td>
<td>11,000</td>
<td>99.68%</td>
</tr>
<tr>
<td>2006</td>
<td>3,452,662</td>
<td>3,433,286</td>
<td>19,376</td>
<td>99.44%</td>
</tr>
<tr>
<td>2007</td>
<td>3,641,458</td>
<td>3,526,833</td>
<td>114,625</td>
<td>96.85%</td>
</tr>
<tr>
<td>2008</td>
<td>3,614,774</td>
<td>3,415,129</td>
<td>199,645</td>
<td>94.48%</td>
</tr>
<tr>
<td>2009</td>
<td>3,825,761</td>
<td>3,656,522</td>
<td>169,239</td>
<td>95.58%</td>
</tr>
<tr>
<td>2010</td>
<td>3,961,150</td>
<td>3,889,435</td>
<td>71,715</td>
<td>98.19%</td>
</tr>
<tr>
<td>2011</td>
<td>4,115,760</td>
<td>4,076,073</td>
<td>39,687</td>
<td>99.04%</td>
</tr>
<tr>
<td>2012</td>
<td>4,391,461</td>
<td>4,356,369</td>
<td>35,092</td>
<td>99.20%</td>
</tr>
<tr>
<td>2013</td>
<td>4,671,052</td>
<td>4,640,023</td>
<td>31,029</td>
<td>99.34%</td>
</tr>
<tr>
<td>2014</td>
<td>5,096,994</td>
<td>5,074,829</td>
<td>22,165</td>
<td>99.57%</td>
</tr>
<tr>
<td>2015</td>
<td>4,972,669</td>
<td>4,968,238</td>
<td>4,431</td>
<td>99.91%</td>
</tr>
<tr>
<td>2016</td>
<td>5,077,409</td>
<td>5,025,399</td>
<td>52,010</td>
<td>98.98%</td>
</tr>
<tr>
<td>Total</td>
<td>69,715,526</td>
<td>68,268,716</td>
<td>1,446,810</td>
<td>97.92%</td>
</tr>
</tbody>
</table>