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 2-1. Summary of Prior HAP Archives

Phase	Year Completed	Coverage Years
I	2001	1990-2000
II	2003	1990-2001
III	2004	1990-2002
IV	2005	1990-2003
V	2007	1973-2005
VI	2009	1973-2008
VII	2013 (Feb)	1973-2010
VIII	2013 (Oct)	1973-2012
IX	2015	1973-2013
X	2016	1973-2014
XI	2017	1990-2015

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".

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

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. 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[5] 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/.

[10] Carbon tetrachloride: NOAA GMD ftp site data downloaded 3/30/2017 from: ftp://ftp.cmdl.noaa.gov/hats/solvents/CCl4/flasks/Otto/monthly/

[11] Updated data provided by Stephen Montzka, A. NOAA/Earth System Research Laboratory/Global Monitoring Division

published references.[12,13,14] A total of 158,780 records from seventeen sites[15] 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

---

[12] Hossaini, R., M.P. Chipperfield, A. Saiz-Lopez, J.J. Harrison, R. von Glasow, R. Sommariva, E. Atlas, M. Navarro, S.A. Montzka, W. Feng, S. Dhomse, C. Harth, J. Mühle, C. Lunder, S. O-Doherty, D. Young, S. Reimann, M.K. Vollmer, P. Krummel, and P.F. Bernath, Growth in stratospheric chlorine from short-lived chemicals not controlled by the Montreal Protocol., Geophys. Res. Lett., 42, 11, 4573-4580, 2015.

[13] Michael C. McCarthy, Hilary R. Hafner & Stephen A. Montzka (2006) Background Concentrations of 18 Air Toxics for North America, Journal of the Air & Waste Management Association, 56:1, 3-11, DOI: 10.1080/10473289.2006.10464436.

[14] L. Hu (CIRES and NOAA), S. A. Montzka (NOAA), B. R. Miller (CIRES and NOAA), A. E. Andrews (NOAA), J. B. Miller (NOAA) S. J. Lehman (INSTAAR, CU-Boulder), C. Sweeney (CIRES and NOAA), S. Miller (Stanford University), K. Thoning (NOAA), C. Siso (CIRES and NOAA), E. Atlas (University of Miami), D. Blake (University of California Irvine), J. A. de Gouw (CIRES and NOAA), J. B. Gilman (CIRES and NOAA), G. Dutton (NOAA), J. W. Elkins (NOAA), B. D. Hall (NOAA), H. Chen (University of Groningen, the Netherlands), M. L. Fischer (Lawrence Berkeley National Laboratory), M. Mountain (Atmospheric and Environmental Research), T. Nehrkorn (Atmospheric and Environmental Research), S. C. Biraud (Lawrence Berkeley National Laboratory), F. Moore (CIRES and NOAA) and P. P. Tans (NOAA), Continued emissions of carbon tetrachloride from the United States nearly two decades after its phaseout for dispersive uses, *Proceedings of the National Academy of Sciences*, doi:10.1073/pnas.1522284113, 2016.

[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 sites[16] 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[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 ½ MDL. The concentrations for these

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

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

- AQS_QUALIFER_06: This field is reserved for data records which were identified as duplicates and were invalidated. For example, duplicates were identified if a concentration record was reported as both a local condition and a standard condition. While the parameter codes may be different, they are the same. As such, the local condition record was retained and the standard condition was invalidated. 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_QUALIFER_07: This field is reserved for data records in which the sample value was invalidated as a result of the NATTS Network Assessment or through discussions with the Data Owners (e.g., the state agency). Accordingly, "INV" was

assigned to the AQS_QUALIFIER_07 field to quickly identify these records as being invalidated.

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

- CF-I-4: Calculation frequency codes incorrectly entered as "4", which is every 4 days.
- CF-I-3: Calculation frequency codes incorrectly entered a "3", which is every 3 days.
- CF-I-2: Calculation frequency codes incorrectly entered a "2", which is every 2 days.
- CF-I-1: Calculation frequency codes incorrectly entered a "1", which is every day.
- 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 ½ MDL or in which the pollutant code equals 43505, which is "Acrolein – Unverified". Accordingly, "SM" and "QV" were assigned, respectively, to the AQS_QUALIFIER_09 field to quickly identify these records. For the "QV" data records, results of a recent short-term laboratory study have raised questions about the consistency and reliability of monitoring results of acrolein. Because of the uncertain accuracy of acrolein measurements, OAQPS has changed the name of the existing acrolein parameter code in AQS (43505) to "Acrolein - Unverified" to indicate the current level of uncertainty that exists with the data already reported to AQS. Correspondingly, a new parameter code (43509) has been created in AQS for "Acrolein - Verified." Whether or not all or a subset of existing data remain in the unverified parameter code, or are re-categorized as verified and moved / reported to this new parameter code, is a choice over which each owning agency has complete discretion. Until such time as agencies evaluate their acrolein monitoring procedures and the quality of reported data, EPA recommends that already-reported data remain in the unverified method code.[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.

[18] Found at: "Data Quality Evaluation Guidelines for Ambient Air Acrolein Measurements. OAQPS. December 17, 2010. Internet address: https://www3.epa.gov/ttn/amtic/files/ambient/airtox/20101217acroleindataqualityeval.pdf

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

Data Field	Data Description
*AMA_SITE_CODE	Ambient Monitoring Archive (AMA) Site Code
*AQS_POC	Parameter Occurrence Code
*SAMPLE_DATE	Date Sample was taken
*SAMPLE_START_TIME	Time at which sample began
*AQS_PARAMETER_CODE	Air Quality Subsystem (AQS) Pollutant Code
HAP_FLAG	Flag to identify if HAP record
NON_HAP_FLAG	Flag to identify if Non-HAP record
CRITERIA_POLL_FLAG	Flag to identify if criteria pollutant record
GHG_POLL_FLAG	Flag to identify if GHG record
MET_DATA_FLAG	Flag to identify if meteorological data record
AQS_METHOD_CODE	Sampling Method Code
AQS_UNIT_CODE	Unit of Measure Code
AQS_SAMPLE_DURATION_CODE	Sample Duration Code
AQS_NULL_DATA_CODE	Data Qualifier code for null sample values
AQS_MONITOR_PROTOCOL_ID	AQS Protocol ID for precision and accuracy records
AQS_QUALIFIER_01	Data Qualifier code field (reserved for reporting agency)
AQS_QUALIFIER_02	Data Qualifier code field (reserved for reporting agency)
AQS_QUALIFIER_03	Data Qualifier code field (reserved for reporting agency)
AQS_QUALIFIER_04	Data Qualifier code field (reserved for reporting agency)
AQS_QUALIFIER_05	Data Qualifier code field (reserved for reporting agency)
AQS_QUALIFIER_06	Data Qualifier code field (reserved for EPA QA – See Section 4.0)
AQS_QUALIFIER_07	Data Qualifier code field (reserved for EPA QA – See Section 4.0)
AQS_QUALIFIER_08	Data Qualifier code field (reserved for EPA QA – See Section 4.0)
AQS_QUALIFIER_09	Data Qualifier code field (reserved for EPA QA – See Section 4.0)
AQS_QUALIFIER_10	Data Qualifier code field (reserved for EPA QA – See Section 4.0)
ALTERNATE_MDL	Method detection limit (MDL), in native units, if entered by Entity
UNCERTAINTY	Estimate of uncertainty surrounding the data, if available
AQS_SAMPLING_FREQUENCY_CODE	Code identifying how often the measurements were collected
SAMPLE_VALUE_REPORTED	Reported concentration value (in native units, where possible)
SAMPLE_VALUE_ADJ	Adjusted concentration value
DATA_SOURCE	Identifies the data source for the data record
COMMENT	Reserved for comments
SAMPLE_VALUE_STD	Concentration value standardized to $\mu g/m^3$
MDL_STD	MDL standardized to $\mu g/m^3$
MDL_TYPE	Identifies the source of the standardized MDL
SAMPLE_VALUE_FLAG	Identifies if the concentration record is a non-detect (Flag = "ND")
AQS_FEDERAL_MDL	Default Federal MDL value
AQS_FEDERAL_MDL_UNIT_CODE	Default Federal MDL value engineering Unit of Measure Code

Table 5-1. Ambient Monitoring Archive Data Input Fields

Data Field	Data Description
BELOW_MDL_FLAG	Identifies if the non-zero sample value is less than the MDL (Flag = Y")
TRANSACTION_DATE	Date in which the data record entered the Archive
CONV_FLAG	Flag to identify concentration records which need to be converted (Flag = 1) to local conditions using local temperature and pressure
TEMP_STD	Ambient temperature standardized to degrees Celsius for concentration records which need converted to local conditions
TEMP_STD_UNITS	Units for the standardized temperature
TEMP_SOURCE	Source of data for the ambient temperature
PRESS_STD	Ambient pressure standardized to mmHg for concentration records which need converted to local conditions
PRESS_STD_UNITS	Units for the standardized pressure
PRESS_SOURCE	Source of data for the barometric pressure
SAMPLE_VALUE_STD_LC	Concentration value standardized to $\mu g/m^3$, local conditions
LC_TYPE	Flag to indicate if the SAMPLE_VALUE_STD_LC is local "L" or standard "S"
WIND_SPEED_MPH	Wind speed in miles per hour (if available) for the concentration record
WIND_DIRECTION_DEGREES	Wind direction in degrees (if available) for the concentration record
WIND_SPEED_SOURCE	Source for the wind speed value
WIND_DIRECTION_SOURCE	Source for the wind direction value

* = primary key field

Sample values populated with a 0 indicate a non-detect, and a corresponding "ND" flag is populated in the SAMPLING_VALUE_FLAG field. Similarly, sample values with no data (or null) indicate that the sample or the pollutant concentration was invalidated by the responsible agency or EPA for any number of reasons.

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

5.2 Site Information

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

Table 5-2. Site Information Data Fields

Data Field	Data Description
*AMA_SITE_CODE	Site Identifier made up of STATE_FIPS, COUNTY_FIPS, and LOCAL_SITE_ID
STATE_FIPS	State Code
COUNTY_FIPS	County Code
STATE_COUNTY_FIPS	Combination of the State and County FIPS
COUNTY_NAME	County Name
LOCAL_SITE_ID	Local Site Identifier
SITE_NAME	Name of Site, if available
CENSUS_TRACT_ID_2000	U.S. Census Tract Identifier for Year 2000
CENSUS_TRACT_ID_2010	U.S. Census Tract Identifier for Year 2010
CENSUS_TRACT_POPULATION_2000	U.S. Census Tract population for Year 2000
CENSUS_TRACT_POPULATION_2010	U.S. Census Tract population for Year 2010
CENSUS_BLOCK_ID_12	U.S. Census Block Identifer for Year 2010
ADDRESS	Monitoring Site Address
CITY	Monitoring Site City
STATE_ABBR	Monitoring Site State Abbreviation
ZIP_CODE	Monitoring Site Zip Code
EPA_REGION	EPA Region Designation
SUPPORT_AGENCY_CODE	Code for the Support Agency
SUPPORT_AGENCY	Support Agency Name
NATTS_SITE_FLAG	Identifies the site as a NATTS Site
UATMP_SITE_FLAG	Identifies the site as a UATMP Site
PAMS_SITE_FLAG	Identifies the site as a PAMS Site
IMPROVE_SITE_FLAG	Identifies the site as an IMPROVE Site
CASTNET_SITE_FLAG	Identifies the site as an CASTNET Site
PM_SUPERSITES_SITE_FLAG	Identifies the site as an PM Supersites Site
PILOT_SITE_FLAG	Identifies the site as an EPA Pilot site
POST_KATRINA_SITE_FLAG	Identifies the site as a Post-Katrina UATMP site
CSATAMP_SITE_CYCLE_FLAG	Identifies the site as a Community-Scale Air Toxics Monitoring site
CANDIDATE_NCORE_SITE_FLAG	Identifies the site as a potential NCORE monitoring site
SCHOOL_AIR_TOXICS_SITE_FLAG	Identifies the site as a School Air Toxics monitoring site
BP_OIL_SPILL_SITE_FLAG	Identifies the site as a BP Oil Spill monitoring site
LEAD_NAAQS_SITE_FLAG	Identifies the site as a Lead NAAQS monitoring site
MONITOR_LATITUDE	Vertical coordinates of the monitoring site

Table 5-2. Site Information Data Fields

Data Field	Data Description
MONITOR_LONGITUDE	Horizontal coordinates of the monitoring site
DATUM	Coordinate data system
UTM_NORTHING	Universal Transverse Mercator Projection Y-coordinate value
UTM_EASTING	Universal Transverse Mercator Projection X-coordinate value
UTM_ZONE	Zone for the UTM coordinates
ELEVATION	Elevation of the monitoring site, in meters
LOCATION_TYPE	Type of location
LAND_USE	Use of land
DATE_SITE_ESTABLISHED	Data in which the site was operational
DATE_SITE_CLOSED	Data in which the site ceased operations
CBSA_NAME	Core-Based Statistical Area name
CBSA_TYPE	CBSA type (metropolitan or micropolitan)
URBAN_AREA_NAME	Shortened MSA name
MONITOR_TRAFFIC_COUNT	Traffic passing by the monitoring site
TRAFFIC_COUNT_YEAR	Year of traffic count
RFG_MANDATED_AREA_FLAG	Indicates the site is in a reformulated gasoline Mandated regulated area
RFG_OPT_IN_AREA_FLAG	Indicates the site is in a reformulated gasoline Opt-In regulated area
RFG_OPT_OUT_AREA_FLAG	Indicates the site is in a reformulated gasoline Opt-Out regulated area
WINTER_OXYGENATED_AREA_FLAG	Indicates the site is in a Winter Oxygenated regulation area
CLOSEST_NWS_STATION	Closest National Weather Service (NWS) station
CLOSEST_NWS_STATION_WBAN	Closest National Weather Service (NWS) station identifier
CLOSEST_NWS_STATION_DISTANCE_MILES	Distance between the monitoring site and the closest NWS station
CLOSEST_NWS_STATION_BEARING_FROM_EAST	Bearing angle from the east of the monitoring site and the closest NWS station
SECOND_CLOSEST_NWS_STATION	Second closest National Weather Service (NWS) station
SECOND_CLOSEST_NWS_STATION_WBAN	Second closest National Weather Service (NWS) station identifier
SECOND_CLOSEST_NWS_STATION_DISTANCE_MILES	Distance between the monitoring site and the second closest NWS station
SECOND_CLOSEST_NWS_STATION_BEARING_FROM_EAST	Bearing angle from the east of the monitoring site and the second closest NWS station
COMMENT	General comment

* = primary key field

A number of useful metadata are provided related to site location, 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

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

* = 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 5-4. Pollutant Information Data Fields

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

Table 5-4. Pollutant Information Data Fields

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

* = primary key field

5.5 Sampling Method Information

Table 5-5 presents data fields for a comprehensive list of sampling methodology codes listed in the AMA_SAMPLING_METHOD_CODE_DICTIONARY. The primary keys for this data table are the AQS_PARAMETER_CODE, AQS_METHODOLOGY_CODE, AQS_SAMPLE_DURATION_CODE, and the AQS_UNIT_CODE (denoted by a "*"). This data dictionary table includes the federal MDL in native units, as well as converted to standardized µg/m^3. A total of 4,221 records are in this data dictionary.

Table 5-5. Sampling Methodology Information Data Fields

Data Field	Data Description
*AQS_PARAMETER_CODE	AQS Parameter Identifier
PARAMETER_DESC	AQS Parameter Identifier Description
*AQS_METHODOLOGY_CODE	AQS Methodology Identifier

Table 5-5. Sampling Methodology Information Data Fields

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

* = primary key field

5.6 Date and Season Information

Table 5-6 presents data fields for every single day from 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

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

* = primary key field

5.7 Qualifier Code Information

Table 5-7 presents data fields for the data qualifier codes in the AMA_QUALIFIER_CODE_DICTIONARY data table. The primary key for this data table is the "AQS_QUALIFIER_CODE" (denoted by a "*"). This data dictionary table includes information related to quality assurance issues, sampling problems, or information related to the concentration records. A total of 217 records are in this data dictionary.

Table 5-7. Qualifier Information Data Fields

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

* = primary key field

5.8 Sample Duration Information

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

Table 5-8. Sample Duration Information Data Fields

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

* = primary key field

5.9 Unit Code Information

Table 5-9 presents data fields for the unit codes in the AMA_UNIT_CODE_DICTIONARY. The primary key for this data table is "AQS_UNIT_CODE" (denoted by a "*"). A total of 20 records are in this data dictionary.

Table 5-9. Unit Information Data Fields

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

* = primary key field

5.10 Collection Frequency Code Information

Table 5-10 presents data fields for the sampling collection frequency codes in the AMA_COLLECTION_FREQUENCY_CODES_DICTIONARY. The primary key for this data table is "Collection Frequency Code" (denoted by a "*"). A total of 24 records are in this data dictionary.

Table 5-10. Frequency Code Data Fields

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

* = primary key field

5.11 Data Source Code Information

Table 5-11 presents data fields for the sampling collection frequency codes in the AMA_DATA_SOURCE_CODES_DICTIONARY. The primary key for this data table is "DATA_SOURCE" (denoted by a "*"). A total of 42 records are in this data dictionary.

Table 5-11. Data Source Code Data Fields

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

* = primary key field

6.0 FINAL DATABASE

The remainder of this memorandum focuses on HAPs. Table 6-1 provides a summary of the final record counts of each data source used to populate Phase XII Archive of HAPs. In total, there are nearly 70 million data records.

Table 6-1. Data Source Information for HAP Records

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

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

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

Table 6-2. HAP Summary Counts by Year

Year	# HAP Records	# Non-Detect Records	% Non-Detect	# Null Data Records	% Null	# HAP Sample Values BMDL	% HAP Sample Values BMDL
1994	488,682	140,009	28.7%	32,678	6.7%	42,348	8.7%
1995	880,780	214,170	24.3%	91,015	10.3%	54,672	6.2%
1996	1,129,426	252,913	22.4%	158,842	14.1%	82,487	7.3%
1997	1,326,261	278,134	21.0%	168,074	12.7%	84,222	6.4%
1998	1,536,342	302,017	19.7%	226,214	14.7%	95,976	6.2%
1999	1,673,443	337,239	20.2%	305,316	18.2%	101,463	6.1%
2000	1,823,136	407,314	22.3%	271,915	14.9%	139,521	7.7%
2001	2,159,662	477,518	22.1%	347,964	16.1%	179,180	8.3%
2002	2,252,877	531,850	23.6%	342,224	15.2%	223,818	9.9%
2003	2,479,804	542,650	21.9%	397,255	16.0%	215,562	8.7%
2004	2,978,195	640,708	21.5%	491,418	16.5%	252,356	8.5%
2005	3,437,633	722,814	21.0%	584,850	17.0%	308,949	9.0%
2006	3,472,362	744,640	21.4%	539,275	15.5%	301,930	8.7%
2007	3,655,404	762,686	20.9%	489,670	13.4%	299,994	8.2%
2008	3,628,561	782,957	21.6%	554,306	15.3%	273,759	7.5%
2009	3,841,448	857,908	22.3%	498,573	13.0%	349,900	9.1%
2010	3,976,020	879,438	22.1%	558,994	14.1%	401,585	10.1%
2011	4,130,186	960,458	23.3%	604,882	14.6%	408,779	9.9%
2012	4,405,802	922,899	20.9%	600,430	13.6%	465,965	10.6%
2013	4,679,865	1,004,152	21.5%	661,457	14.1%	472,802	10.1%
2014	5,105,617	1,068,311	20.9%	691,285	13.5%	497,084	9.7%
2015	4,981,803	1,022,432	20.5%	676,126	13.6%	488,378	9.8%
2016	5,082,341	1,051,992	20.7%	667,705	13.1%	552,832	10.9%
Totals	**69,937,812**	**15,237,037**	**21.8%**	**10,003,649**	**14.3%**	**6,384,685**	**9.1%**

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

State	Total # of ND	Total # Surrogate	# Fed MDL Surrogate	# Entity-Provided MDL Surrogates	Time Period of Surrogates
Alabama	77,783	2	2	0	1993-1996
Alaska	59,551	0	0	0	NA
Arizona	183,618	0	0	0	NA
Arkansas	26,242	0	0	0	NA
California	958,594	375,508	253,660	121,848	1990-2016
Colorado	153,225	31	31	0	2002
Connecticut	277,696	0	0	0	NA
Delaware	83,632	193	32	161	2000-2012

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

State	Total # of ND	Total # Surrogate	# Fed MDL Surrogate	# Entity-Provided MDL Surrogates	Time Period of Surrogates
District of Columbia	145,471	100	77	23	1997-2008
Florida	174,464	14,640	122	14,518	1990-2006
Georgia	585,466	0	0	0	NA
Hawaii	34,175	0	0	0	NA
Idaho	53,358	10,621	0	10,621	2002-2008
Illinois	497,105	18	17	1	2005
Indiana	383,639	31	31	0	1990
Iowa	81,230	0	0	0	NA
Kansas	132,946	1	1	0	1990
Kentucky	96,377	655	0	655	2006-2007
Louisiana	168,016	1	1	0	1999
Maine	784,860	0	0	0	NA
Maryland	179,960	392	392	0	1997-2000
Massachusetts	394,960	0	0	0	NA
Michigan	358,935	57	57	0	1992-1994
Minnesota	460,603	5	5	0	1999
Mississippi	81,310	0	0	0	NA
Missouri	196,601	0	0	0	NA
Montana	107,596	2	2	0	1991
Nebraska	22,807	0	0	0	NA
Nevada	43,160	0	0	0	NA
New Hampshire	288,886	110	110	0	2002-2004
New Jersey	275,765	15	15	0	1993-2005
New Mexico	63,401	0	0	0	NA
New York	261,198	9,842	9,842	0	1990-1999
North Carolina	140,085	1,140	1,112	28	2002-2008
North Dakota	39,532	2	2	0	2000
Ohio	203,118	3	0	3	2004-2005
Oklahoma	101,865	0	0	0	NA
Oregon	160,380	39,326	2,116	37,210	1999-2007
Pennsylvania	516,189	1,086	850	236	2000-2013
Rhode Island	195,525	0	0	0	NA
South Carolina	212,059	16	16	0	1993-1994
South Dakota	64,702	0	0	0	NA
Tennessee	63,978	138	138	0	1990-1998
Texas	5,105,429	2,051	2,051	0	1994-2009
Utah	92,884	0	0	0	NA
Vermont	120,468	9	8	1	1995-2002
Virginia	126,305	244	113	131	2000-2012
Washington	137,149	4,619	4	4,615	1995-2006
West Virginia	30,336	0	0	0	NA
Wisconsin	125,913	0	0	0	NA
Wyoming	85,183	0	0	0	NA
Puerto Rico	13,981	0	0	0	NA
Virgin Islands	9,326	0	0	0	NA
Total	**15,237,037**	**460,858**	**270,807**	**190,051**	**1990-2016**

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

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

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

Table 7-1. Ambient Monitoring Archive Output Fields

Data Field	Data Description
PROGRAM	Identifies Monitoring Program, if available
YEAR	Year of sampling date
QUARTER	Calendar quarter of the sampling date
*SAMPLE_DATE	Date Sample was taken
*SAMPLE_START_TIME	Time at which sample began
*AQS_PARAMETER_CODE	Air Quality System (AQS) Pollutant Code
AQS_PARAMETER_NAME	AQS pollutant name
DATA_SOURCE	Identifies the source of the data record
DURATION_DESC	Translated AQS Sample Duration description
SAMPLE_VALUE_REPORTED	Reported sample value from the data source
AQS_UNIT_CODE	Unit of Measure Code for the native sample value
UNIT_DESC	Translated AQS Unit of Measure description
SAMPLING_FREQUENCY_CODE	Collection Frequency code (1=Daily; 2=EveryOtherDay; 3=Every3Days; 4=Every4Days; 5=Every5Days; 6=Every6Days; 7=Every12Days; 8=StratifiedRandom; 9=Random; 10=Every24Days; 11=Every30Days; 12=Every7Days; 14=Every14Day; 90=Every90Days; A=PAMS Daily; H or J=PAMS 3rdDay; P=PAMS 6thDay; S=Seasonal; Y=TwicePerWeek; Z=Every9Days)
COMMENT	Reserved for comments
SAMPLE_VALUE_STD_FINAL_UG_M3	Concentration value standardized to µg/m^3, local conditions
SAMPLE_VALUE_STD_FINAL_TYPE	Final Concentration type for analysis (L = Local Conditions, S = Standard Conditions)
AQS_PARAMETER_CODE_FINAL	Final AQS Pollutant Code for analysis
AQS_PARAMETER_NAME_FINAL	Final AQS Pollutant Name for analysis
ALTERNATE_MDL_REPORTED	Reported MDL in native units
MDL_STD_UG_M3	MDL standardized to µg/m^3
MDL_TYPE	Identifies the source of the standardized MDL
AQS_NULL_DATA_CODE	Data Qualifier code for null sample values
AQS_QUALIFIER_01	Data Qualifier code field 1
AQS_QUALIFIER_02	Data Qualifier code field 2
AQS_QUALIFIER_03	Data Qualifier code field 3
AQS_QUALIFIER_04	Data Qualifier code field 4
AQS_QUALIFIER_05	Data Qualifier code field 5
AQS_QUALIFIER_06	Data Qualifier code field 6
AQS_QUALIFIER_07	Data Qualifier code field 7
AQS_QUALIFIER_08	Data Qualifier code field 8
AQS_QUALIFIER_09	Data Qualifier code field 9
AQS_QUALIFIER_10	Data Qualifier code field 10
AQS_METHOD_CODE	Sampling and Analysis Method Code
SAMPLE_COLLECTION_DESC	Translated AQS Sampling Collection description
SAMPLE_ANALYSIS_DESC	Translated AQS Analysis Method description
SAMPLE_VALUE_FLAG	Identifies if the concentration record is a non-detect
BELOW_MDL_FLAG	Identifies if the non-zero sample value is less than the MDL
CENSUS_TRACT_ID	U.S. Census tract identifier in which the monitoring site is located
CENSUS_TRACT_POPULATION_2010	2010 Population within the census tract
MONITOR_LATITUDE	Y-Coordinate Value in decimal degrees
MONITOR_LONGITUDE	X-Coordinate Value in decimal degrees

* = primary key field

In the public release files, EPA is not outputting "Acrolein – unverified" (parameter code = 43505) due to the unreliability of the measurements. Similarly, the following parameter codes are not included in the Ambient Monitoring Archive output files, as they are combined pollutants which cannot be disaggregated for air quality use:

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

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

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

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

State	Total # Output Records	Total # Local Condition Records	Total # Standard Condition Records	% Local Condition Records
New York	1,679,957	1,618,151	61,806	96.32%
North Carolina	426,512	424,619	,893	99.56%
North Dakota	79,396	79,396	0	100.00%
Ohio	505,876	489,858	16,018	96.83%
Oklahoma	300,585	299,848	737	99.75%
Oregon	370,482	364,944	5,538	98.51%
Pennsylvania	1,367,344	1,324,513	42,831	96.87%
Rhode Island	752,509	752,495	14	100.00%
South Carolina	534,128	416,684	117,444	78.01%
South Dakota	130,588	130,588	0	100.00%
Tennessee	176,440	166,128	10,312	94.16%
Texas	31,287,934	31,249,545	38,389	99.88%
Utah	300,924	299,916	1,008	99.67%
Vermont	756,968	326,740	430,228	43.16%
Virginia	587,661	585,701	1,960	99.67%
Washington	393,366	391,825	1,541	99.61%
West Virginia	138,353	130,964	7,389	94.66%
Wisconsin	3,386,068	3,351,668	34,400	98.98%
Wyoming	157,484	155,667	1,817	98.85%
Puerto Rico	31,465	31,289	176	99.44%
Virgin Islands	27,324	27,324	0	100.00%
TOTALS	**69,715,526**	**68,268,716**	**1,446,810**	**97.92%**

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

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

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

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