TECHNICAL MEMORANDUM

TO: Madeleine Strum, Doris Chen, and Laurie Trinca
Delivery Order Managers
U.S. EPA Office of Air Quality Planning and Standards

FROM: Regi Oommen, Karla Faught, Jaime Hauser, Jennifer Sellers, and Steve Mendenhall, Eastern Research Group, Inc. (ERG)

DATE: January 24, 2020

SUBJECT: Compilation and Quality Assurance (QA) Summary Report for the Phase XIII Ambient Monitoring Archive for Hazardous Air Pollutants (HAPs)

1.0 INTRODUCTION

The purpose of this memorandum and accompanying appendices are to summarize improvements, modifications, and additional data incorporated into the development of EPA's Phase XIII Ambient Monitoring Archive (AMA, otherwise known as "the Archive"). Under a prior Delivery Order, Eastern Research Group, Inc. (ERG) prepared Phase XII of the AMA, which comprised of hazardous air pollutant (HAP) air toxics monitoring data collected from numerous federal, state, local, and tribal agencies from 1990 to 2016. Here, we use the word "Phase" as synonymous with version.

ERG was tasked to develop Phase XIII by updating the Archive to 2017, incorporate additional data not in the previous Archive, and provide general maintenance/cleanup of the Phase XII Archive. All work was performed under EPA Contract No. EP-D-14-030, Delivery Orders 00-51 and 00-62 entitled "Report Development – Data Characterization." This memorandum contains seven sections and six appendices, as presented in Table 1-1.

Table 1-1. Structure of the Memorandum

Section	Section Title
1	Introduction
2	Background Information
3	AMA Data Sources
4	QA Fixes and Data Changes
5	Database Structuring/Processing
6	Final Database
7	Final Output Data Files

Table 1-1. Structure of the Memorandum

Section	Section Title
Appendix A	Overlapping Records
Appendix B	Invalidated Records
Appendix C	Sampling Frequency Code Corrections
Appendix D	Questionable Values and Incorrectly Submitted ½ MDL Concentrations
Appendix E	Negative Concentrations and Incorrectly Assigned Qualifier Codes for "MD", "ND", and "SQ"
Appendix F	Program Ranking

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 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 timelines.

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
XII	2018	1990-2016

EPA developed the Phase XII Archive in June 2018, which contained nearly 70 million HAP records from 1990 to 2016. The Phase XII Archive was the seventh successful update built upon the re-engineered system that was developed for the Phase VI effort (Summer 2009). This re-engineering allowed ERG to simplify future updates. Data records were housed as their original sample durations from AQS, such as hourly measurements. Additionally, identification of possible non-detect data measurement records were substituted as one-half the method detection limit (MDL) value.

For the Phase XIII update, EPA requested that ERG:

- Retrieve 1990-2017 ambient HAP data from EPA's Air Quality Subsystem (AQS);
- Incorporate additional datasets, if available;
- Perform general housekeeping/cleanup of the new data retrieved from AQS;
- Standardize all descriptions (e.g., pollutant names, sampling methodology, etc.) and data fields;
- Assign and QA the AQS "Sampling Frequency Code" data based on sample dates;
- Assure each data record has a corresponding MDL;
- Identify sample values which were entered as ½ MDL (i.e., non-measured values);
- Identify sample values below MDL;
- Identify duplicative data reported in AQS from the reporting entity;
- Identify and maintain data records which have been invalidated;
- Perform range checks on reported data;
- Review and update data qualifier flags;
- Standardize all reported concentrations to local conditions using meteorological data using collocated or nearby monitoring sites, where applicable; and
- Prepare data files for posting to EPA's website.

3.0 AMA DATA SOURCES

For the Phase XIII Archive, there were twenty-one primary data sources used. Table 3-1 provides a summary of the final record counts of each data source used to populate Phase XIII Archive of HAPs. In total, there are over 75 million data records.

Table 3-1. Data Source Information for HAP Records

Data Source	Data Years	# Sites	# Pollutants/ Parameters	HAP Data Record Count
Air Quality System (AQS) Database	1990-2017	2,333	365	54,853,944
Allegheny County, PA Health Department	2013-2017	3	14	6,931
Baldwin Hills Air Quality Study	2012-2013	1	16	7,455
Baltimore Inner Harbor Monitoring Study	2014-2015	6	1	1,734
City of Ft. Worth, TX Natural Gas Air Quality Study	2010	8	49	5,455
EPA Passive Sampling	2013-2015	17	9	18,675
EPA Region 3	2008-2017	2	14	2,779
Integrated Atmospheric Deposition Network	1999-2010	11	89	162,836
Minnesota Air Toxics	2008-2016	45	61	124,952
National Acid Deposition	1999-2017	166	4	2,125,990
NATTS Network Assessment	2003-2014	5	71	11,608
NOAA	1991-2017	18	7	684,392
Oregon DEQ	2012-2017	10	3	3,350
Pennsylvania Marcellus Shale Study	2012-2013	6	39	16,806
Phase V/VII Archive	1991-2010	145	165	202,153
School Air Toxics	2011-2012	6	80	800
South Coast AQMD	1999-2017	43	60	163,718
Sublette County, WY	2009-2010	14	42	37,398
TCEQ TAMIS	1992-2017	124	79	16,398,517
Utah State University-Vernal	2012-2017	4	9	13,802
XACT Monitoring Data	2011-2015	8	10	207,842
Totals	**1990-2017**	**2,629**	**378**	**75,051,137**

Information about each Data Source is presented in Sections 3.1 – 3.21. In the Archive, the field "DATA_SOURCE" identifies the data source and the date obtained as coming from one of the sources below (e.g., "AQS_20190405" equates to AQS data retrieved on April 29, 2019).

As part of its process to identify new sources of air toxics data, ERG reviewed state and local monitoring plans posted on EPA's website.[1] Additionally, ERG reviewed Community-Scale Air Toxics Ambient Monitoring (CSATAM) projects, and checked to determine if the monitoring data were uploaded to AQS.[2] Finally, ERG reviewed conference proceedings to

[1] State and Local Monitoring Plans are posted at: https://www.epa.gov/amtic/state-and-local-monitoring-plans
[2] More information on CSATAM projects are posted at: https://www3.epa.gov/ttn/amtic/local.html

identify data from air toxics projects that are not uploaded into AQS. In each of these situations, project sponsor/awardees would be contacted to obtain the data.

3.1 Air Quality System (AQS) Data

AQS is EPA's official repository of ambient monitoring data. Users of AQS can download data from pre-generated data files,[3] the AQS API,[4] or using standard/ad-hoc queries within the AQS data portal (which requires a User account and password).[5] 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 Stations (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. NATTS data are required to be submitted within 180 days at the end of the calendar year (120 days before 2018).[6]

AQS data for the 2017 data year were initially retrieved from the AQS data portal in December 2018 using the AMP501 ("Extract Raw Data") function. By using this function, the original data were obtained and not standardized. Additionally, data from 1990-2016 were also retrieved to replace Phase XII database records since the Archive was last updated (June 2018). Subsequent data pulls were performed in January, February, March, April, September, and October 2019 as EPA added new data into AQS. The pull dates are provided in the DATA_SOURCE field. Nearly 54.9 million HAP records from 2,333 sites and 365 parameters were incorporated into the Archive. Site and pollutant-specific MDLs were populated for approximately 28% of all the data records.

3.2 Allegheny County, PA

The Allegheny County Health Department (ACHD) in Pittsburgh, PA conducts metals and VOC sampling in the Pittsburgh area in which the data are not sent to AQS. As such, EPA

---

[3] Pre-generated Data Files from AQS are available at: https://aqs.epa.gov/aqsweb/airdata/download_files.html
[4] More information about the AQS API is found at: https://aqs.epa.gov/aqsweb/documents/data_api.html
[5] Can be accessed via the AQS Launch Web Application file at: https://www.epa.gov/aqsc
[6] As reported in Section 3.3.1.3.15in the Technical Assistance Document for the NATTS Program, Revision 3. (https://www3.epa.gov/ttnamti1/files/ambient/airtox/NATTS%20TAD%20Revision%203_FINAL%20October%202016.pdf)

coordinated with ACHD to obtain this data, as well as site metadata.[7] More information on the ACHD and their monitoring program can be found at: http://www.achd.net/air/index.php. A total of 6,931 records from 2013 through 2017 for three sites[8] and 14 parameters were incorporated into the Archive, which included new data for 2017 for Phase XIII. Site and pollutant-specific MDLs were provided for all the data records.

3.3 Baldwin Hills Air Quality Study

Los Angeles County, in coordination with the South Coast Air Quality Management District (SCAQMD) conducted an air quality study in the Baldwin Hills area near oil and gas activities in 2012 and 2013. More information on this study can be found at: http://planning.lacounty.gov/assets/upl/project/bh_air-quality-study.pdf. These data were sent to ERG from the SCAQMD contractor for inclusion into the Archive, as it is not housed in AQS.[9] A total of 7,455 records from one site[10] and 16 parameters were incorporated into the Archive. Pollutant-specific MDLs were provided for all of the metals data. However, the pollutant MDLs obtained from the Proton-Transfer-Reaction Time-of-Flight Mass Spectometry (PTR-TOFMS) were obtained from the manufacturer.[11]

3.4 Baltimore Inner Harbor Monitoring Study

The Maryland Department of the Environment and U.S. EPA Region 3 oversaw a special hexavalent chromium monitoring study at six sites in the Baltimore Inner Harbor from 2014 to 2015. The study focused on establishing baseline air quality concentrations.[12] These data were sent to ERG from the SCAQMD contractor for inclusion into the Archive, as it is not housed in

---

7 Monitoring results provided by ACHD directly to EPA via e-mail from Mr. Darrell Stern, ACHD on 5/7/2019.
8 The three sites are: Avalon (420030002); Lawrenceville (420030008); and Liberty (420030064).
9 Email from Mr. Mike McCarthy, Sonoma Technology to Mr. Regi Oommen, ERG on 4/25/2016.
10 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".
11 Per the manufacturer, the level of detection (LOD) for the pollutants of interest (2,4-dinitrotoluene, benzene, naphthalene, 1,3-butadiene, acrolein, and total xylenes) is less than 1 pptv (https://www.ionicon.com/information/technology/ionicon-ptr-tofms-series-performance).
12 https://mde.maryland.gov/programs/LAND/RecyclingandOperationsprogram/Documents/Publications/HPD_DDP%20Air%20Monitoring%20Plan_2013-11-12.pdf

AQS.[13] A total of 1,734 records from six sites[14] and one parameter were incorporated into the Archive. Pollutant-specific MDLs were provided for all records.

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. During this two-month study, 5,455 concentrations were generated at eight monitoring sites[15] for 49 parameters. More information can be found at: http://fortworthtexas.gov/uploadedFiles/Gas_Wells/AirQualityStudy_final.pdf. ERG, as the contract lab, received permission from DEM to include the data into the Archive. Pollutant-specific MDLs were provided for all records.

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 from 2013 through 2015, two-week duration samples were collected in South Philadelphia. More information can be found at: https://cfpub.epa.gov/si/si_public_file_download.cfm?p_download_id=527897. These data were sent to ERG from the City of Philadelphia for inclusion into the Archive, as it is not housed in AQS.[16] A total of 18,675 records from seventeen sites[17] and nine parameters were incorporated into the Archive. Pollutant-specific MDLs were provided for all records.

3.7 EPA Region 3

The West Virginia Division of Air Quality conducted multi-year (2008 through 2017) metals measurements at two sites in WV targeting specific sources of interest. Filter samples

---

[13] Email from Ms. Jaime Hauser, ERG to Mr. Regi Oommen, ERG on 12/19/2016.

[14] 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".

[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 "48439LS02" site is located in Tarrant County, TX (FIPS = 48439) and the site identifier is "LS02".

[16] Email from Ms. Hallie Weiss, City of Philadelphia to Mr. Regi Oommen, ERG on 12/12/2017.

[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 "42101PS04" site is located in Philadelphia County, PA (FIPS = 42101) and the site identifier is "PS04".

were sent for analysis to the EPA Region 3 lab, who also coordinated these data to be sent to EPA for inclusion into the Archive, as it is not housed in AQS.[18] A total of 2,779 records from two sites and 14 parameters were incorporated into the Archive, which included 2017 data. Pollutant-specific MDLs were provided for all records.

3.8 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. U.S. data from 1991-2010 for the organic, polycyclic aromatic hydrocarbon (PAH), and polychlorinated biphenyls (PCB) compounds were retrieved from the IADN website.[19] Recent data (2011-present) only covers sites in Canada. A total of 162,836 records from eleven sites[20] and 89 parameters were incorporated into the Archive. Pollutant-specific MDLs were provided for all records.

3.9 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 removed that data from AQS and provided their entire dataset from 2008-2016 to ERG for inclusion in Phase XIII Archive.[21] 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 124,952 records from 45

[18] Email from Mr. Howard Schmidt, EPA Region 3 to Mr. Regi Oommen, ERG on 2/27/2018.
[19] http://donnees.ec.gc.ca/data/air/monitor/monitoring-of-combined-atmospheric-gases-and-particles/?lang=en
[20] 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".
[21] Email from Ms. Kellie Gavin, MNPCA to Mr. Regi Oommen, ERG on 3/5/2018.

sites and 61 parameters were incorporated into the Archive. Pollutant-specific MDLs were provided for all records.

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 from 1999 through 2017 from the MDN and AMNet networks were downloaded from http://nadp.slh.wisc.edu/data/. A total of 2,125,990 records from 166 sites[22] and 4 parameters were incorporated into the Archive. Pollutant-specific MDLs were provided for all records.

3.11 National Air Toxics Trends Station (NATTS) Network Data Review

In Fall 2017, ERG, under contract to EPA, prepared a final report on data reporting for the NATTS Network. As per the requirements of the NATTS Network, participating sites are to report data to AQS. During this data review, a number of concentrations reported to AQS were identified as incorrect (and never corrected in AQS). Additionally, certain datasets were identified as missing from AQS, and were obtained from the NATTS Operators. The corrected and missing data were not submitted to AQS, and were obtained by EPA 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 11,608 records from 5 sites and 71 parameters from 2003 through 2014 were incorporated into the Archive. Pollutant-specific MDLs were provided for all records.

[22] 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".

3.12 National Oceanic and Atmospheric Administration (NOAA)

Select air toxics data are collected at the National Oceanic and Atmospheric Administration's (NOAA) monitoring sites, often in remote locations. Three measurement programs from NOAA sites were incorporated into the Archive:

- Global Monitoring Division (GMD): Air toxics data collected under the Global Monitoring Data program include several air toxics: methylene chloride; tetrachloroethylene; benzene; and carbon tetrachloride.[23,24,25,26] A composite dataset was sent to EPA.[27] 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 published references.[28,29,30] Measurements through 2017 were sent to EPA for all pollutants, except carbon tetrachloride, which extended through 2014.

- Chromatograph for Atmospheric Trace Species (CATS): Long-term in-situ hourly measurements for halocarbons, including carbon tetrachloride (the only HAP) since 1998 through 2017 at three U.S. sites (Mauna Loa, HI; Niwot Ridge, CO; and Pt. Barrow, AK). The CATS Gas Chromatographs are custom built instruments with four separate channels. Each channel is comprised of a pair of separation columns, flow controllers, an air selection valve, and an electron capture detector. More information can be found at: https://www.esrl.noaa.gov/gmd/hats/insitu/cats/.

---

[23] Methylene Chloride: https://www.esrl.noaa.gov/gmd/hats/gases/CH2Cl2.html

[24] Tetrachloroethylene: https://www.esrl.noaa.gov/gmd/hats/gases/C2Cl4.html

[25] Benzene data was provided by Stephen Montzka, A. NOAA/Earth System Research Laboratory/Global Monitoring Division to Madeleine Strum. April 16, 2018 via e-mail.

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

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

[28] 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.

[29] 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.

[30] 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.

- OTTO Flask Data: The data reported are from samples collected approximately once per week in matching, concurrent, flask pairs and later analyzed on a gas chromatograph with electron capture detection (GC-ECD) located in Boulder, Colorado, USA. This system uses two standard reference gases for calibration and has been in operation since very early in 1995. Weekly, 5-minute carbon tetrachloride measurement data from seven sites from 2015-2017 were retrieved at (ftp://ftp.cmdl.noaa.gov/hats/solvents/CCl4/flasks/Otto/pairs/). More information can be found at: http://www.esrl.noaa.gov/gmd/hats/flask/flasks.html.

A total of 684,392 records from 1991 through 2017 for eighteen sites[31] and seven parameters were incorporated into the Archive. Pollutant-specific MDLs were provided for all records.

3.13 Oregon Department of Environmental Quality Carbonyls

In Summer 2019, EPA was alerted by the Oregon Department of Environmental Quality (ODEQ) of incorrectly submitted carbonyl compound (acetaldehyde, formaldehyde, and propionaldehyde) concentrations residing in AQS from 2012 through 2017. As such, a total of 3,350 revised concentrations from 10 sites and three parameters were sent by ODEQ and incorporated into the Archive.[32] Pollutant-specific MDLs were provided for all records. These data will eventually be updated in AQS by ODEQ.

3.14 Pennsylvania Department of Environmental Protection

To evaluate impacts from oil and gas wells in the Marcellus Shale area of Pennsylvania, HAP measurements from 2012 through 2013 were collected. The sampling results provided basic information about the types of pollutants emitted to the atmosphere during selected phases of gas extraction operations in the Marcellus Shale formation. The project placed emphasis on characterizing concentrations of criteria pollutants and HAPs near permanent facilities related to the Marcellus Shale gas industry in Washington County, PA. More information is available at: https://www.dep.pa.gov/Business/Energy/OilandGasPrograms/OilandGasMgmt/Oil-and-Gas-Related-Topics/Pages/Air.aspx.

[31] 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.

[32] Email from Mr. Chris Moore, ODEQ to Mr. Regi Oommen, ERG on 9/27/2019.

A total of 16,806 records for six sites[33] and 39 parameters were incorporated into the Archive. Pollutant-specific MDLs were provided for all records.

3.15 Phase V/VII Database

The Phase V Database originally consisted of over nine million daily (e.g., no hourly or minute data) 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, retained for posterity, are for invalidated VOC data originally submitted by the Kentucky Department of Environmental Services. A total of 202,153 records from 1991 through 2010 for 145 sites and 165 parameters were incorporated into the Archive. Pollutant-specific MDLs were provided for the majority of the records.

3.16 School Air Toxics Ambient Monitoring Program

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 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 newspaper series on air toxics at schools, and information from state and local air pollution agencies. Phase 1 Sampling took place in 2009-2010 in 59 schools across the U.S., while Phase 2 Sampling followed up at 22 schools in 2010-2011. Nearly all of the data resides in AQS, with the exception of:

- Special VOC measurements taken at two schools during the Phase 2 Sampling: Enterprise High School in Enterprise, MS and Temple Elementary is Diboll, TX.
- Some records from the four Alabama schools.

[33] 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 Henderson site (Site ID = HEND), located in Washington County, PA (FIPS code = 42125) is assigned 42125HEND.

These data were retrieved by EPA and formatted for 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. Pollutant-specific MDLs were provided for all records.

3.17 South Coast Air Quality Management District

The South Coast Air Quality Management District (SCAQMD) sponsored air quality data characterization studies called the Multiple Air Toxics Exposure Study (MATES). MATES-II (1999), MATES-III (2004-2007), and MATES-IV (2012-2013) data were obtained from SCAQMD. More information can be found at: https://www.aqmd.gov/home/air-quality/air-quality-studies/health-studies/.

SCAQMD also launched two hexavalent chromium studies:

- Western Riverside County: In 2008, after a five-month intensive investigation of toxic air pollution in the western areas of Riverside and San Bernardino Counties, SCAQMD identified cement production in the area as a source of elevated levels of the cancer-causing chemical hexavalent chromium. As a result, SCAQMD took aggressive action to reduce the hexavalent chromium emissions from the two cement plants in the area in order to lower the cancer risk levels in the communities around the facilities. Sampling took place at seventeen sites from 2008-2011.[34] More information can be found at: http://www.aqmd.gov/docs/default-source/air-quality/special-monitoring-and-emissions-studies/hexavalent-chromium-study/hexavalent-chromium-air-monitoring-data.pdf?sfvrsn=2.

- Compton Community Air Toxics Initiative (CATI). SCAQMD has been measuring levels of ambient air hexavalent chromium near several industrial facilities in the Compton area since June 2017. This air monitoring effort at ten monitoring sites[35] was aimed to identify and prioritize high-risk facilities with the potential to emit hexavalent chromium, then use the latest air monitoring technology to confirm specific sources of high emissions. More information can be found at: http://www.aqmd.gov/docs/default-source/air-toxics-initiative/compton/updated-air-monitoring-plan.pdf?sfvrsn=14.

[34] 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, Site TXI-1 (TX01), located in Riverside County, CA (FIPS code = 06065) is assigned 06065TX01.

[35] 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, Site #1C (CS01), located in Los Angeles County, CA (FIPS code = 06037) is assigned 06037CS01.

A total of 163,718 records from 43 sites and 60 parameters were incorporated into the Archive. Pollutant-specific MDLs were provided for all records.

3.18 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 to February 2010. More information on the sampling design and analysis of the measurements can be found at: http://www.sublettewyo.com/documentcenter/view/438. Nearly 37,400 HAP concentrations from fourteen sites[36] and 42 parameters were formatted for upload for the Archive.[37] Pollutant-specific MDLs were provided for all records.

3.19 Texas Commission on Environmental Quality (TCEQ)

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 for non-identical overlaps. A total of 16,398,517 records from 1992 through 2017 for 124 sites and 79 parameters were incorporated into the Archive. The pollutant-method specific MDLs were pulled from the TAMIS website and included in the concentration record.

3.20 Utah State University

Utah State University (USU) in Vernal, UT collects HAP measurements during wintertime in and around oil and gas wells in the northeastern Utah. USU is carrying out a comprehensive research program to understand and provide solutions for the Basin's air quality problems. This is a cooperative effort with Uintah and Duchesne Counties, local industry, the Utah Division of Air Quality, the Ute Indian Tribe, the TriCounty Health Department, research

[36] 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".

[37] Email from Ms. Cara Keslar, Wyoming DEP to Mr. Regi Oommen, ERG on 7/13/2014.

teams at other Utah universities and universities around U.S., and several federal agencies (BLM, EPA, DOE). Over 13,800 HAP concentrations from four sites[38] and 9 parameters from 2012 to 2017 were formatted for upload for the Archive.[39] Sample durations included a mixture of 1-hr, 3-hr, and 24-hr measurements. Pollutant-specific MDLs were provided for approximately 15% of all records. More information on the sampling program can be found at: http://binghamresearch.usu.edu/files/15_0331%20annual%20report%20ute%20tribe.pdf.

3.21 XAct Monitoring Data

U.S. EPA purchased XAct Monitoring Measurement Systems as a result of the 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 the XAct system in a 2011 study of particulate matter (PM) metals. Measurements data were sent by ODEQ to EPA and were processed for the Archive.[40] After this study, EPA Region 5 conducted several monitoring campaigns, ranging from two- to six-months from 2012 to 2016 in Illinois, Indiana, and Michigan using XAct for targeting specific sources. A total of 207,842 records from eight sites[41] and 10 parameters were incorporated into the Archive. Pollutant-specific MDLs were provided for all records.

4.0 QA FIXES AND DATA CHANGES

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

- **Pollutant Name Update**. In the Archive pollutant dictionary, all pollutants analyzed via the TO-13A method were changed from "(Tsp) STP" to "(total tsp and vapor)". For example, parameter code 17141 was changed from "naphthalene (Tsp) STP" to "naphthalene (total tsp and vapor)".

[38] 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 "49047HRPL" site is located in Uintah County, UT (FIPS = 49047) and the site identifier is "HRPL".
[39] Email from Mr. Seth Lyman, USU to Mr. Regi Oommen, ERG on 4/19/2019.
[40] Email from Ms. Aida Biberic, ODEQ to Mr. Dave Shelow, EPA on 6/24/2013.
[41] 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".

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

    - 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, mark as being an incorrectly substituted record for non-detects.

- **Negative Concentrations**. Over 217,000 concentrations were reported 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.

- **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" (which is "Miscellaneous Void") in the AQS_NULL_DATA_CODE data field and as "INV" (which is "Invalidated") 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). All hexavalent chromium concentrations prior to 2005, all polycyclic aromatic hydrocarbon (e.g., naphthalene, benzo(a)pyrene, anthracene, etc.) concentrations prior to 2007, and 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. Both conditions were retained in the Archive, while standard conditions were invalidated.

- **Revised Concentrations**. Through the NATTS Network Assessment and UATMP, small sets of blanks data were mistakenly entered into AQS, and were nulled out accordingly. Additionally, outlier concentrations were identified, and in some cases, revised data were sent to EPA.

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

- **Inconsistency of Coding**. ERG evaluated AQS coding of the following Qualifier Codes for inconsistencies:

    - MD: This Qualifier Code is used to designate reported concentrations between the 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 Qualifier Code 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 Qualifier Code 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 Qualifier Code 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 Qualifier Code flag was removed.

    - SQ: This Qualifier Code is used to designate reported concentrations compared to the Sample Quantitation Limit (PQL), which is 3.18 times the MDL.[42] Concentration records were deemed "inconsistent" if they were assigned "SQ", but the reported values were greater than 3.18 times the MDL. As such, the Qualifier Code flag was removed.

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

- AQS_QUALIFER_06: This field is reserved for data records which were identified as duplicates or overlaps and were invalidated. 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 pollutant, but with concentrations reported for different temperature and pressure conditions. 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" (parameter code 45102), "*m/p*-xylene" (parameter code 45109), "*m*-xylene" (parameter code 45205), "*o*-xylene" (parameter code 45204), and/or "*p*-xylene" (parameter code 45206). Accordingly, "OVR" was assigned to the

[42] U.S. EPA, 2016. Technical Assistance Document for the National Air Toxics Trends Stations Programs, Rev. 3. Found at: https://www3.epa.gov/ttnamti1/files/ambient/airtox/NATTS%20TAD%20Revision%203_FINAL%20October%202016.pdf

AQS_QUALIFIER_06 field to identify these invalidated records. Table 4-1 summarizes the fate of multiple reporting for the xylene records, where "X' indicates there is a valid concentration. Appendix A presents the records that were invalidated.

Table 4-1. Xylene Overlap Scenarios

Overlap Scenarios					Fate
xylene(s) (45102)	***m,p*-xylene (45109)**	***o*-xylene (45204)**	***m*-xylene (45205)**	***p*-xylene (45206)**	
X	X	X	X	X	Invalidate 45102 and 45109
X	X	X	X		Invalidate 45102 and 45205
X	X	X			Invalidate 45102
X	X		X		Invalidate 45109 and 45205
X	X			X	Invalidate 45109 and 45206
X	X				Invalidate 45109
X		X			Invalidate 45204
X			X		Invalidate 45205
X				X	Invalidate 45206
	X	X	X	X	Invalidate 45109
	X	X	X		Invalidate 45205
	X	X		X	Invalidate 45206
	X	X			No overlap
	X	X			No overlap
	X		X	X	Invalidate 45109
	X		X		Invalidate 45205
	X			X	Invalidate 45206
		X	X	X	No overlap
		X	X		No overlap
		X		X	No overlap
			X	X	No overlap

- 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 these invalidated records. Appendix B presents the records that were invalidated.

- AQS_QUALIFER_08: This field is reserved for data records in which the Collection Frequency Code was not populated in the concentration and/or monitor data, or if the value entered was suspected to be incorrect. Accordingly, "CF" was assigned to the AQS_QUALIFIER_08 field to identify these records. Appendix C presents the records that were changed.

- 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" ("surrogate method used") and "QV" ("questionable value") were assigned, respectively, to the AQS_QUALIFIER_09 field to 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.[43] Lastly, "PC" ("potential calculation error") is assigned in this field. Appendix D presents the records that were identified.

- 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 identify these records. Additionally, records in which the Data Qualifier was inconsistent in its coding of "MS", "MD", "ND", "PQ", and "SQ" were noted in this field. Appendix E presents the records that were identified.

[43] 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

5.0 DATABASE STRUCTURE/PROCESSING

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 and overlapped 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 75 million HAP records were in the blended master database.

After merging the data, ERG calculates the "SAMPLE_VALUE_REPORTED" to a standardized concentration in $\mu g/m^3$, using the following procedures:

- For AQS_UNIT_CODE = 001 (micrograms per cubic meter, standard conditions), no change
- For AQS_UNIT_CODE = 008 (parts per billion by volume, ppbv):

  $$\text{concentration, } \mu g/m^3 = (\text{concentration, ppbv})*(\text{molecular weight})*(24.45)$$

- For AQS_UNIT_CODE = 007 (parts per million by volume, ppmv):

  $$\text{concentration, } \mu g/m^3 = (1000)*(\text{concentration, ppbv})*(\text{molecular weight})*(24.45)$$

- For AQS_UNIT_CODE = 121 (parts per trillion by volume, pptv):

  $$\text{concentration, } \mu g/m^3 = \frac{(\text{concentration, ppbv})*(\text{molecular weight})*(24.45)}{(1000)}$$

- For AQS_UNIT_CODE = 078 (parts per billion by carbon, ppbC):

  $$\text{concentration, } \mu g/m^3 = \frac{(\text{concentration, ppbv})*(\text{molecular weight})*(24.45)}{(\text{number of carbons})}$$

- For AQS_UNIT_CODE = 101 (parts per million by carbon, ppmC):

  $$\text{concentration, } \mu g/m^3 = \frac{(1000)*(\text{concentration, ppbv})*(\text{molecular weight})*(24.45)}{(\text{number of carbons})}$$

- For AQS_UNIT_CODE = 003 (nanograms per cubic meter, standard conditions):

$$\text{concentration, } \mu g/m^3 = \frac{(\text{concentration, } ng/m^3)}{(1000)}$$

- For AQS_UNIT_CODE = 004 (nanograms per cubic meter, 0° Celsius):

$$\text{concentration, } \mu g/m^3 = \frac{(\text{concentration, } ng/m^3)*(273\ K)}{(1000)*(298\ K)}$$

- For AQS_UNIT_CODE = 074 (picograms per cubic meter, standard conditions):

$$\text{concentration, } \mu g/m^3 = \frac{(\text{concentration, } ng/m^3)}{(1000000)}$$

- For AQS_UNIT_CODE = 174 (picograms per cubic meter, 0° Celsius):

$$\text{concentration, } \mu g/m^3 = \frac{(\text{concentration, } ng/m^3)*(273\ K)}{(1000000)*(298\ K)}$$

- For AQS_UNIT_CODE = 002 (micrograms per cubic meter, 0° Celsius):

$$\text{concentration, } \mu g/m^3 = \frac{(\text{concentration, } ng/m^3)*(273\ K)}{(298\ K)}$$

- For AQS_UNIT_CODE = 105 (micrograms per cubic meter, local conditions):

$$\text{concentration, } \mu g/m^3 = \frac{(\text{concentration, } ng/m^3)*(\text{Local temperature in K})*(760\ mm\ Hg)}{(298\ K)*(\text{Local pressure in mm Hg})}$$

- For AQS_UNIT_CODE = 108 (nanograms per cubic meter, local conditions):

$$\text{concentration, } \mu g/m^3 = \frac{(\text{concentration, } ng/m^3)*(\text{Local temperature in K})*(760\ mm\ Hg)}{(1000)*(298\ K)*(\text{Local pressure in mm Hg})}$$

Phase XIII 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 ("foreign keys" are columns in a relational database table that provides a link between data in two tables.) Relational tables ensure data integrity and provide more scalability. To translate the data in the Ambient Monitoring Archive, ERG developed nine data dictionary tables. These dictionaries are critical in properly describing and standardizing the raw data, and

are recommended for providing additional context to the concentration records. AQS data dictionaries were initially retrieved from EPA's AQS website, which provided the metadata information for the AQS-submitted data. Data elements that were not in the AQS data dictionaries were subsequently added. The ten data dictionaries are presented in Sections 5.1 through 5.10 below.

5.1 Site Information

Table 5-1 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 5-1. 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[1]	State Code
COUNTY_FIPS[1]	County Code
STATE_COUNTY_FIPS	Combination of the State and County FIPS
COUNTY_NAME	County Name
LOCAL_SITE_ID[1]	Local Site Identifier
AQS_SITE_NAME[1]	Site Name in AQS
AMA_SITE_NAME	Additional/alternative name of site, if available
CENSUS_TRACT_ID_2000	U.S. Census Tract Identifier for Year 2000
CENSUS_TRACT_ID_2010[1]	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[1]	U.S. Census Block Identifier for Year 2010
ADDRESS[1]	Monitoring Site Address
CITY[1]	Monitoring Site City
STATE_ABBR	Monitoring Site State Abbreviation
ZIP_CODE[1]	Monitoring Site Zip Code
EPA_REGION	EPA Region Designation
SUPPORT_AGENCY_CODE[1]	Code for the Support Agency
SUPPORT_AGENCY[1]	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

Table 5-1. Site Information Data Fields

Data Field	Data Description
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[1]	Vertical coordinates of the monitoring site
MONITOR_LONGITUDE[1]	Horizontal coordinates of the monitoring site
DATUM[1]	Coordinate data system
UTM_NORTHING[1]	Universal Transverse Mercator Projection Y-coordinate value
UTM_EASTING[1]	Universal Transverse Mercator Projection X-coordinate value
UTM_ZONE[1]	Zone for the UTM coordinates
ELEVATION[1]	Elevation of the monitoring site, in meters
LOCATION_TYPE[1]	Type of location, which is typically populated in AQS
LAND_USE[1]	Use of land
DATE_SITE_ESTABLISHED[1]	Data in which the site was operational
DATE_SITE_CLOSED[1]	Data in which the site ceased operations
CBSA_NAME	Core-Based Statistical Area name
CBSA_TYPE	CBSA type (metropolitan or micropolitan)
URBAN_AREA_NAME	Alternate MSA name
MONITOR_TRAFFIC_COUNT[2]	Traffic passing by the monitoring site
TRAFFIC_COUNT_YEAR[2]	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

Table 5-1. Site Information Data Fields

Data Field	Data Description
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
[1] Data field in the AQS "AA" data table
[2] Data field in the AQS "AB" data table

A number of useful metadata are provided related to site location, monitoring programs, demographic/population activities, and regulatory applicability. A total of 2,629 records are in this data dictionary.

5.2 Monitor Information

Table 5-2 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 NATTS, UATMP, Photochemical Assessment Monitoring Sites (PAMS), and the Interagency Monitoring of Protected Visual Environments (IMPROVE) network. A total of 426,508 records are in this data dictionary.

Table 5-2. 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[1]	Parameter Occurrence Code
*AQS_PARAMETER_CODE[1]	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[1]	Program associated with each monitor, if available

Table 5-2. Monitor Information Data Fields

Data Field	Data Description
MONITOR_OBJECTIVE	Sampling Objective of the Monitor, which may be populated in AQS
MONITOR_TYPE[1]	Type of Monitor, which may be populated in AQS
MONITOR_DESIGNATION	Indicates whether the monitor is the primary, secondary, or not determined
EPA_PQAO[1]	AQS identifier for the Primary Quality Assurance Organization
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 for each AMA_SITE_CODE, AQS_PARAMETER_CODE, and SAMPLE_YEAR combination
AQS_METHOD_CODE	AQS Method Code(s) per monitor
PROGRAM_RANK	Ranking of the PROGRAM

* = primary key field
[1] Data field in the AQS "MN" and Monitors data table

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, benzene concentrations from the NATTS Program will generally have a higher priority ranking than benzene concentrations from the PAMS Program. Appendix F presents the ranking for each PROGRAM type.

5.3 Pollutant Information

Table 5-3 presents data fields for a comprehensive list of HAP parameter codes listed in the AMA_POLLUTANT_CODES_DICTIONARY. The "AQS_PARAMETER_CODE" is the only primary key field in this data dictionary (denoted by a "*"). This data dictionary table includes physical information and alternative pollutant identifiers. A total of 378 records are in the master data dictionary.

Table 5-3. 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[1]	AQS Pollutant Identifier
AQS_PARAMETER_NAME[1]	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[2]
HAP_FLAG	Indicates pollutant is a HAP
CAP_FLAG	Indicates pollutant is a criteria air pollutant (only parameters representing lead are flagged)
GHG_FLAG	Indicates pollutant is a greenhouse gas air pollutant
TO15_FLAG	Indicates pollutant can be measured using the TO-15 method[3]
TO11A_FLAG	Indicates pollutant can be measured using the TO-11A method[4]
IO3_5_FLAG	Indicates pollutant can be measured using the IO3.5 method[5]
TO13_FLAG	Indicates pollutant can be measured using the TO-13A method[6]
8270C_FLAG	Indicates pollutant can be measured using the 8270 method[7]
SNMOC_FLAG	Indicates pollutant can be measured using the SNMOC method[8]
ERG_HEX_FLAG	Indicates pollutant can be measured using the ASTM D7614 method[9]
PAMS_FLAG	Indicates pollutant can be measured using the PAMS method[10]
COMMENT	General comment

* = primary key field

1. Data field in the AQS "All Parameters" data table (https://aqs.epa.gov/aqsweb/documents/codetables/methods_all.html)
2. The list of urban-33 pollutants are listed at https://www.epa.gov/urban-air-toxics/urban-air-toxic-pollutants
3. The TO-15 pollutants are posted at: https://www.epa.gov/sites/production/files/2019-11/documents/to-15r.pdf
4. The TO-11A pollutants are posted at: https://www.epa.gov/sites/production/files/2019-11/documents/to-11ar.pdf
5. The IO-3.5 pollutants are posted at: https://www.epa.gov/sites/production/files/2019-11/documents/mthd-3-5.pdf
6. The TO-13A pollutants are posted at: https://www.epa.gov/sites/production/files/2019-11/documents/to-13arr.pdf
7. The 8270C pollutants are posted at: http://www.caslab.com/EPA-Methods/PDF/8270c.pdf
8. The SNMOC pollutants are posted at: https://www3.epa.gov/ttnamti1/archive/files/ambient/criteria/reldocs/r-99-053.pdf
9. The ASTM D7614 pollutants are posted at: https://www.astm.org/Standards/D7614.htm
10. The PAMS pollutants are listed at: https://www3.epa.gov/ttnamti1/pamsmain.html

5.4 Sampling Method Information

Table 5-4 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 original units, as well as converted to µg/m^3 (either in standard or local conditions in relation to the original units). A total of 4,074 records are in this data dictionary.

Table 5-4. Sampling Methodology Information Data Fields

Data Field	Data Description
*AQS_PARAMETER_CODE[1]	AQS Parameter Identifier
PARAMETER_DESC[1]	AQS Parameter Identifier Description
*AQS_METHODOLOGY_CODE[1]	AQS Methodology Identifier
SAMPLE_COLLECTION_DESC[1]	Sample Collection Description
SAMPLE_ANALYSIS_DESC[1]	Sample Analysis Description
*AQS_SAMPLE_DURATION_CODE	Duration Identifier
DURATION_DESC	Duration Identifier Description
*AQS_UNIT_CODE	Unit of Measure Identifier
UNIT_DESC[1]	Unit Description
AQS_FEDERAL_MDL_VALUE[1]	Federal default method detection limit
AQS_FEDERAL_MDL_UNIT	Federal default method detection limit units
FEDERAL_MDL_VALUE_STD	Federal default method detection limit standardized to µg/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

[1] Data field in the AQS "All Parameters" data table (https://aqs.epa.gov/aqsweb/documents/codetables/methods_all.html)

5.5 Date and Season Information

Table 5-5 presents data fields for every single day from 1990 to 2017 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 10,227 records are in this data dictionary.

Table 5-5. 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

5.6 Qualifier Code Information

Table 5-6 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. While the majority of the qualifier codes are from AQS, additional qualifier codes were included from non-AQS sources. A total of 184 records are in this data dictionary.

Table 5-6. Qualifier Information Data Fields

Data Field	Data Description
*AQS_QUALIFIER_CODE[1]	Qualifier Identifier
QUALIFIER_DESC[1]	Qualifier Description
QUALIFIER_TYPE[1]	Type of Qualifier
QUALIFIER_TYPE_DESC[1]	Type of Qualifier Description

* = primary key field
[1] Data field in the AQS "Qualifiers" data table (https://aqs.epa.gov/aqsweb/documents/codetables/qualifiers.html)

5.7 Sample Duration Information

Table 5-7 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 19 records are in this data dictionary.

Table 5-7. Sample Duration Information Data Fields

Data Field	Data Description
*AQS_DURATION_CODE[1]	Duration Identifier
DURATION_DESC[1]	Duration Identifier Description
DURATION_INDICATOR	Duration Indicator Identifier
DURATION_LENGTH	Length of sampling
DURATION_UNIT	Unit of length for sampling

* = primary key field
[1] Data field in the AQS "Durations" data table (https://aqs.epa.gov/aqsweb/documents/codetables/durations.html)

5.8 Unit Code Information

Table 5-8 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 18 records are in this data dictionary.

Table 5-8. Unit Information Data Fields

Data Field	Data Description
*AQS_UNIT_CODE[1]	Unit of Measure Identifier
UNIT_DESCRIPTION[1]	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

[1] Data field in the AQS "Durations" data table (https://aqs.epa.gov/aqsweb/documents/codetables/units.html)

5.9 Collection Frequency Code Information

Table 5-9 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 28 records are in this data dictionary.

Table 5-9. Frequency Code Data Fields

Data Field	Data Description
*AQS_COLLECTION_FREQUENCY_CODE1	Collection Frequency Code Identifier
COLLECTION_FREQUENCY_DESCRIPTION1	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

[1] Data field in the AQS "Durations" data table (https://aqs.epa.gov/aqsweb/documents/codetables/collection_frequencies.html)

5.10 Data Source Code Information

Table 5-10 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 75 records are in this data dictionary.

Table 5-10. Data Source Code Data Fields

Data Field	Data Description
*DATA_SOURCE	Data Source Code Identifier
DATA_SOURCE_DESCRIPTION	Data Source Code Description
DATA_SOURCE_GROUP	Data Source Grouping

Table 5-10. Data Source Code Data Fields

Data Field	Data 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 parameter codes (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

Approximately 22% of the raw data records are non-detects, while less than 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 8% of the reported HAP records were below the MDL (BMDL). Table 6-1 provides a summary of these counts by year.

Table 6-1. 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	63,153	44.4%	6,924	4.9%	10,809	7.6%
1991	175,668	79,601	45.3%	6,507	3.7%	13,679	7.8%
1992	207,994	90,173	43.4%	12,171	5.9%	15,093	7.3%
1993	286,436	107,219	37.4%	20,009	7.0%	19,444	6.8%
1994	488,778	147,808	30.2%	32,733	6.7%	23,216	4.7%
1995	881,149	225,914	25.6%	91,195	10.3%	26,380	3.0%
1996	1,129,989	271,310	24.0%	159,185	14.1%	41,318	3.7%
1997	1,326,936	296,108	22.3%	168,547	12.7%	40,982	3.1%
1998	1,542,253	329,064	21.3%	228,729	14.8%	43,187	2.8%
1999	1,690,569	371,909	22.0%	313,860	18.6%	43,007	2.5%
2000	1,839,739	443,352	24.1%	279,033	15.2%	79,251	4.3%
2001	2,184,936	521,240	23.9%	354,997	16.2%	112,894	5.2%
2002	2,277,669	572,831	25.1%	345,342	15.2%	152,261	6.7%
2003	2,506,069	567,619	22.6%	401,036	16.0%	161,972	6.5%
2004	3,002,274	653,665	21.8%	497,420	16.6%	198,523	6.6%
2005	3,464,413	729,130	21.0%	591,403	17.1%	263,148	7.6%
2006	3,496,342	771,179	22.1%	544,908	15.6%	237,319	6.8%
2007	3,675,595	785,485	21.4%	494,624	13.5%	235,273	6.4%
2008	3,660,109	784,218	21.4%	564,282	15.4%	228,064	6.2%
2009	3,921,448	859,210	21.9%	553,326	14.1%	315,625	8.0%
2010	4,057,556	892,239	22.0%	617,606	15.2%	364,847	9.0%
2011	4,208,242	961,193	22.8%	657,710	15.6%	393,068	9.3%
2012	4,484,830	927,817	20.7%	650,057	14.5%	442,014	9.9%
2013	4,820,584	1,041,655	21.6%	721,101	15.0%	469,005	9.7%
2014	5,311,505	1,105,402	20.8%	839,574	15.8%	477,553	9.0%
2015	5,175,515	1,038,161	20.1%	820,287	15.8%	481,846	9.3%
2016	5,156,286	1,057,603	20.5%	714,531	13.9%	503,647	9.8%
2017	3,936,127	889,272	22.6%	243,119	6.2%	403,185	10.2%
Totals	**75,051,137**	**16,583,530**	**22.1%**	**10,930,216**	**14.6%**	**5,796,610**	**7.7%**

Of the 16,583,530 HAP non-detects in the master database, approximately 5% (857,671 records) were suspected as being non-detects in which a concentration equal to ½ MDL were either intentionally or mistakenly substituted. Table 6-2 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 6-2. Non-Detect Records Populated with ½ MDL by State

State	Total # of ND	Total # Surrogate	# Fed MDL Surrogate	# Entity-Provided MDL Surrogates	Time Period of Surrogates
Alabama	78,878	6	6	0	1993-2002
Alaska	61,105	0	0	0	NA
Arizona	189,459	4	0	4	2007-2014
Arkansas	27,158	0	0	0	NA
California	1,055,556	448,288	265,226	183,062	1990-2017
Colorado	159,485	56	56	0	2002-2002
Connecticut	278,521	69	2	67	1994-2001
Delaware	84,474	240	35	205	2000-2016
District of Columbia	151,397	89	89	0	1996-2006
Florida	179,457	16,046	123	15,923	1990-2014
Georgia	593,131	1	0	1	2008-2008
Hawaii	34,825	0	0	0	NA
Idaho	54,596	10,690	0	10,690	2002-2008
Illinois	502,122	159	159	0	1991-2009
Indiana	400,277	128	75	53	1990-2013
Iowa	82,084	1	0	1	2003-2003
Kansas	134,177	2	2	0	1990-1990
Kentucky	100,627	0	0	0	NA
Louisiana	170,248	17	16	1	1994-2017
Maine	802,612	1	1	0	1991-1991
Maryland	183,421	617	528	89	1990-2017
Massachusetts	402,526	208	1	207	2000-2017
Michigan	381,385	126	120	6	1992-2014
Minnesota	481,594	23	20	3	1997-2015
Mississippi	83,458	3	0	3	2006-2006
Missouri	199,335	4	4	0	2006-2006
Montana	111,265	19	19	0	1991-1996
Nebraska	23,546	0	0	0	NA
Nevada	44,903	0	0	0	NA
New Hampshire	322,037	110	110	0	2002-2004
New Jersey	299,389	20	19	1	1990-2009
New Mexico	65,987	1	0	1	2015-2015
New York	277,961	10,644	10,643	1	1990-2002
North Carolina	159,887	1,238	1,238	0	2002-2011
North Dakota	40,641	2	2	0	2000-2000
Ohio	219,273	10	0	10	2004-2015
Oklahoma	111,638	0	0	0	NA
Oregon	170,883	39,599	2,101	37,498	1999-2017
Pennsylvania	607,002	1,406	1,133	273	1993-2013
Rhode Island	200,765	744	0	744	2003-2017
South Carolina	217,305	22	22	0	1993-1994
South Dakota	65,821	0	0	0	NA
Tennessee	65,016	185	185	0	1990-1998

Table 6-2. 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
Texas	5,934,020	321,690	353	321,337	1992-2009
Utah	97,598	0	0	0	NA
Vermont	123,961	175	8	167	1995-2016
Virginia	132,303	321	136	185	1995-2012
Washington	143,853	4,634	5	4,629	1995-2009
West Virginia	31,555	72	5	67	1997-2016
Wisconsin	134,034	0	0	0	NA
Wyoming	87,201	0	0	0	NA
Puerto Rico	14,296	0	0	0	NA
Virgin Islands	9,482	1	0	1	2012-2012
Total	**16,583,530**	**857,671**	**282,442**	**575,229**	**1990-2017**

In the Phase XIII database, data has been stored with native sample durations, as presented in Table 6-3. Over 63% of the records have a sample duration of 1-hour and nearly 27% have a sample duration of 24 hours.

Table 6-3. Phase XIII HAP Database Sample Duration Counts by Year

Year	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	4	0	0	493	0	0	0	0	0	175,161	10
1992	42	0	0	1,302	0	0	0	0	0	206,650	0
1993	78	38,579	0	21,401	0	0	872	0	0	225,506	0
1994	194	154,837	0	59,000	0	0	0	0	0	274,747	0
1995	644	482,527	0	84,192	2,088	0	133	0	0	311,565	0
1996	988	659,926	0	120,852	6,876	0	0	0	0	341,347	0
1997	1,242	837,469	0	120,476	3,843	0	0	0	0	363,906	0
1998	1,508	1,018,796	0	154,287	2,799	0	0	0	0	364,863	0
1999	2,500	1,104,572	0	154,112	0	0	0	2,130	0	425,536	1,719
2000	2,369	1,187,307	0	137,269	1,797	0	0	1,578	0	507,463	1,956
2001	2,223	1,327,335	0	135,038	5,879	0	0	0	6,092	705,712	2,657
2002	3,027	1,258,598	0	134,088	10,664	0	0	0	4,290	863,847	3,155
2003	3,621	1,428,373	0	116,193	9,641	0	0	0	2,262	942,331	3,648
2004	93,243	1,720,676	0	100,965	17,659	0	0	2,313	1,108	1,062,234	4,076
2005	105,189	2,038,832	0	104,265	14,526	0	0	10,475	0	1,186,845	4,281
2006	104,171	2,184,692	0	113,262	5,073	0	0	3,324	0	1,081,144	4,676
2007	324,307	2,184,749	0	125,786	0	0	2,020	0	0	1,033,703	5,030
2008	389,269	2,119,019	6,072	111,063	18	9	2,015	0	1,975	1,025,066	5,603
2009	390,976	2,235,947	95,412	114,119	1,140	369	3	0	1,089	1,076,705	5,688
2010	419,456	2,390,815	97,776	118,167	1,149	276	0	0	1,134	1,023,202	5,581
2011	300,462	2,736,937	93,931	110,856	858	225	26	0	0	959,574	5,373
2012	267,464	3,072,301	53,664	105,195	933	30	218	0	0	979,550	5,475
2013	190,423	3,435,766	60,537	99,883	327	48	0	0	0	1,029,289	4,311
2014	306,967	3,765,077	122,106	101,120	579	69	0	0	0	1,002,654	12,933

Table 6-3. Phase XIII HAP Database Sample Duration Counts by Year

Year	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		
2015	339,438	3,654,542	133,863	80,321	15	6	0	0	0	965,899	1,431
2016	434,567	3,738,055	44,256	41,288	0	0	0	456	352	897,312	0
2017	330,087	2,637,643	46,488	33,134	0	0	0	666	348	887,761	0
Totals	**4,014,459**	**47,413,370**	**754,105**	**2,598,883**	**85,864**	**1,032**	**5,287**	**20,942**	**19,050**	**20,060,462**	**77,683**
%Total	**5.3%**	**63.2%**	**1.0%**	**3.5%**	**0.1%**	**<0.01%**	**0.01%**	**0.03%**	**0.03%**	**26.7%**	**0.1%**

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 XIII raw table are presented in Table 7-1. Primary key fields are denoted by a "*".

Table 7-1. Ambient Monitoring Archive Output Fields

Data Field	Data Description
STATE_ABBR	Two-letter abbreviation for the state with the monitoring site
*AMA_SITE_CODE	Ambient Monitoring Archive (AMA) Site Code
*AQS_POC	Parameter Occurrence Code (POC)
PROGRAM	Identifies Monitoring Program, if available
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=Every14Days; 18=Every18Days; 90=Every90Days; A, B, or E=PAMS Daily; H, I, J, or L=PAMS 3rdDay; O=Every10Days; P=PAMS 6thDay; Q=Every8Days; R=Every13Days; S=Seasonal; Y=TwicePerWeek; Z=Every9Days)
COMMENT	Reserved for comments
SAMPLE_VALUE_STD	Concentration value standardized to µg/m^3, standard conditions
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 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

Table 7-1. Ambient Monitoring Archive Output Fields

Data Field	Data Description
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
MONITOR_LATITUDE	Y-Coordinate Value in decimal degrees
MONITOR_LONGITUDE	X-Coordinate Value in decimal degrees
PRIORITY_TRENDS	Ranking of monitor datasets for each AMA_SITE_CODE, AQS_PARAMETER_CODE, and SAMPLE_YEAR combination

* = primary key field

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

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

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

- Concentration records in which the reported unit codes are local conditions, such as: 105, 108; and
- All null or zero concentration records, regardless of reported unit

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), regardless of 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	300,443	283,917	16,526	94.50%
Alaska	420,182	413,986	6,196	98.53%
Arizona	551,999	547,609	4,390	99.20%
Arkansas	66,934	66,361	573	99.14%
California	4,270,647	3,756,512	514,135	87.96%
Colorado	604,669	450,898	153,771	74.57%
Connecticut	1,241,176	1,178,453	62,723	94.95%
Delaware	268,143	243,485	24,658	90.80%
District of Columbia	580,506	576,099	4,407	99.24%
Florida	756,772	744,552	12,220	98.39%
Georgia	1,916,203	1,912,342	3,861	99.80%
Hawaii	376,358	181,371	194,987	48.19%
Idaho	109,976	109,629	347	99.68%
Illinois	1,257,524	1,203,669	53,855	95.72%
Indiana	2,398,025	2,346,586	51,439	97.85%
Iowa	194,543	194,129	414	99.79%
Kansas	222,280	187,933	34,347	84.55%
Kentucky	376,363	376,361	2	100.00%
Louisiana	776,706	574,410	202,296	73.95%
Maine	1,814,389	1,742,404	71,985	96.03%
Maryland	1,300,304	1,070,115	230,189	82.30%
Massachusetts	1,906,378	1,822,022	84,356	95.58%
Michigan	1,253,767	1,214,555	39,212	96.87%

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
Minnesota	1,199,077	1,150,402	48,675	95.94%
Mississippi	381,984	379,826	2,158	99.44%
Missouri	860,883	846,496	14,387	98.33%
Montana	245,640	235,922	9,718	96.04%
Nebraska	64,981	60,419	4,562	92.98%
Nevada	92,981	92,981	0	100.00%
New Hampshire	840,827	815,812	25,015	97.02%
New Jersey	1,401,449	1,389,005	12,444	99.11%
New Mexico	170,339	169,206	1,133	99.33%
New York	1,863,190	1,782,274	80,916	95.66%
North Carolina	455,007	435,441	19,566	95.70%
North Dakota	82,690	82,690	0	100.00%
Ohio	573,491	565,371	8,120	98.58%
Oklahoma	343,041	342,141	900	99.74%
Oregon	402,208	402,044	164	99.96%
Pennsylvania	1,522,502	1,276,341	246,161	83.83%
Rhode Island	774,210	645,310	128,900	83.35%
South Carolina	548,861	522,525	26,336	95.20%
South Dakota	133,882	133,882	0	100.00%
Tennessee	179,801	168,771	11,030	93.87%
Texas	33,433,136	32,042,654	1,390,482	95.84%
Utah	355,629	354,112	1,517	99.57%
Vermont	784,624	780,893	3,731	99.52%
Virginia	601,839	575,324	26,515	95.59%
Washington	415,169	412,582	2,587	99.38%
West Virginia	143,281	137,865	5,416	96.22%
Wisconsin	3,761,248	3,741,735	19,513	99.48%
Wyoming	163,706	163,638	68	99.96%
Puerto Rico	32,624	32,200	424	98.70%
Virgin Islands	28,212	28,212	0	100.00%
TOTALS	**74,820,799**	**70,963,472**	**3,857,327**	**94.84%**

Table 7-3 presents a summary of the final counts in the Output files by year. From 2001 to 2017, over 97% 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	83,794	57,029	59.50%
1991	174,336	102,612	71,724	58.86%
1992	206,450	123,641	82,809	59.89%
1993	285,740	153,559	132,181	53.74%
1994	486,712	268,289	218,423	55.12%

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
1995	879,878	459,919	419,959	52.27%
1996	1,128,072	786,714	341,358	69.74%
1997	1,322,128	975,224	346,904	73.76%
1998	1,537,937	1,329,979	207,958	86.48%
1999	1,683,890	1,536,810	147,080	91.27%
2000	1,834,429	1,736,532	97,897	94.66%
2001	2,179,746	2,056,496	123,250	94.35%
2002	2,270,886	2,111,023	159,863	92.96%
2003	2,498,572	2,348,436	150,136	93.99%
2004	2,992,999	2,843,174	149,825	94.99%
2005	3,440,373	3,266,371	174,002	94.94%
2006	3,476,642	3,310,087	166,555	95.21%
2007	3,661,649	3,596,851	64,798	98.23%
2008	3,646,322	3,535,712	110,610	96.97%
2009	3,905,761	3,799,630	106,131	97.28%
2010	4,042,686	3,948,594	94,092	97.67%
2011	4,193,816	4,077,072	116,744	97.22%
2012	4,470,489	4,409,683	60,806	98.64%
2013	4,810,744	4,744,596	66,148	98.62%
2014	5,301,746	5,236,841	64,905	98.78%
2015	5,165,877	5,125,141	40,736	99.21%
2016	5,151,293	5,097,647	53,646	98.96%
2017	3,930,803	3,899,045	31,758	99.19%
Total	**74,820,799**	**70,963,472**	**3,857,327**	**94.84%**