COMPILATION AND QUALITY ASSURANCE SUMMARY REPORT FOR THE 2023 AMBIENT MONITORING ARCHIVE FOR THE HAZARDOUS AIR POLLUTANTS

EPA Contract No. 68HERH22D0002, Delivery Order # 68HERH23F0306

Prepared for:

Jeanette Reyes and Xi (Doris) Chen
U.S. Environmental Protection Agency
Office of Air Quality Planning and Standards
109 T.W. Alexander Drive
Research Triangle Park, NC 27711

Prepared by:

Regi Oommen, Karla Faught, Jaime Hauser, and Steve Mendenhall
Eastern Research Group, Inc.
3800 Gateway Centre Blvd,
Suite 307
Morrisville, North Carolina 27560

November 19, 2025

This page is intentionally left blank.

Table of Contents

Tables

Common Acronyms and Abbreviations

$\mu g/m^3$	microgram(s)/cubic meter
ng/m^3	nanogram(s)/cubic meter
pg/m^3	picogram(s)/cubic meter
ACHD	Allegheny County Health Department
AGAGE	Advanced Global Atmospheric Gases Experiment
AIR	Atmosphere Innovation Research
AIRMoN	Atmospheric Integrated Research Monitoring Network
AirToxScreen	Air Toxics Screening Assessment
AMA	Ambient Monitoring Archive
AMNet	Atmospheric Mercury Network
AMoN	Ammonia Monitoring Network
API	Application Programming Interface
AQS	Air Quality System
ASTM	American Society for Testing and Materials
BLM	Bureau of Land Management
BMDL	Below Method Detection Limit
C	Celsius
CAP	Criteria Air Pollutant
CARB	California Air Resource Board
CAS	Chemical Abstracts Service
CATI	Community Air Toxics Initiative
CATS	Chromatograph for Atmospheric Trace Species
CBSA	Core-Based Statistical Area
CDPHE	Colorado Department of Public Health and Environment
CSATAM	Community-Scale Air Toxics Ambient Monitoring
DEC	Department of Environmental Conservation
DEM	Department of Environmental Management
DEP	Department of Environmental Protection
DEQ	Department of Environmental Quality
DNR	Department of Natural Resources
DOE	Department Of Energy
EOM	Enhanced Ozone Monitoring
EPA	Environmental Protection Agency
ERG	Eastern Research Group
EtO	Ethylene Oxide
FIPS	Federal Information Processing Standards
GC-ECD	Gas Chromatograph with Electron Capture Detection
GHG	Green House Gas
HAP	Hazardous Air Pollutant
HATS	Halocarbon and other Atmospheric Trace Species
Hg	mercury
IADN	Integrated Atmospheric Deposition Network
IDL	Instrument Detection Limit

IEM	Iowa Environmental Mesonet
IMPROVE	Interagency Monitoring of Protected Visual Environments
INV	Invalid
IO	Inorganic
IRIS	Integrated Risk Information System
K	Kelvin
LC	Location Conditions
LISTOS	Long Island Sound Tropospheric Ozone Study
MATES	Multiple Air Toxics Exposure Study
MDE	Maryland Department of the Environment
MDL	Method Detection Limit
MDN	Mercury Deposition Network
MIT	Massachusetts Institute of Technology
MLN	Mercury Litterfall Network
mm	millimeter
MNPCA	Minnesota Pollution Control Agency
MQO	Method Quality Objective
mw	molecular weight
NAAQS	National Ambient Air Quality Standards
NADP	National Atmospheric Deposition Program
NASA	National Aeronautics and Space Administration
NATA	National Air Toxics Assessment
NATTS	National Air Toxics Trends Sites
NCore	National Core (Multi-pollutant Monitoring Network)
ND	Non-Detect
NEI	National Emissions Inventory
NOAA	National Oceanic and Atmospheric Administration
NPS	National Parks Service
NTN	National Trends Network
OAQPS	Office of Air Quality Planning and Standards
ODEQ	Oregon Department of Environmental Quality
OHEPA	Ohio Environmental Protection Agency
ORD	Office of Research and Development
PAH	Polycyclic Aromatic Hydrocarbon
PAMS	Photochemical Assessment Monitoring Sites
PCB	Polychlorinated Biphenyls
PCP	Programmable Compressor Package
PFP	Programmable Flask Package
PM	Particulate Matter
POC	Parameter Occurrence Code
ppbC	parts per billion carbon
ppbv	parts per billion by volume
ppmC	parts per million carbon
ppmv	parts per million by volume
pptv	parts per trillion by volume

PQAO	Primary Quality Assurance Organization
PQL	Practical Quantitation Limit
press.	pressure
PTR-TOFMS	Proton-Transfer-Reaction Time-of-Flight Mass Spectrometry
QA	Quality Assurance
RFA	Request for Applications
RFG	Reformulated Gasoline
RITS	Radiatively Important Trace Species
SCAQMD	South Coast Air Quality Management District
SNMOC	Speciated Nonmethane Organic Compounds
SPod	Sensor Pod
SQL	Structured Query Language/Sample Quantitation Limit
STD	Standard Conditions
SVOCs	Semivolatile Organic Compounds
TAMIS	Texas Air Monitoring Information System
TCEQ	Texas Commission on Environmental Quality
temp.	temperature
TO	Toxic Organics
TSP	Total Suspended Particulate
UATMP	Urban Air Toxics Monitoring Program
URE	Unit Risk Estimate
USU	Utah State University
UTM	Universal Transverse Mercator
VOC	Volatile Organic Compound
WVDEP	West Virginia Department of Environmental Protection

1.0 Introduction

The purpose of this report and accompanying appendices is to detail the ambient monitoring data sources contained within the Environmental Protection Agency's (EPA's) 2023 Ambient Monitoring Archive (AMA) for the Hazardous Air Pollutants (HAPs) (otherwise known as "the Archive") as well as describe improvements and modifications since the 2022 Archive. This database contains HAP monitoring data from 1990 to 2023 collected from several federal, state, local, and tribal agencies and from academic, community, and short-term studies.

Eastern Research Group, Inc. (ERG) was tasked to develop the next version by updating the Archive through the year 2023, incorporate additional data not in the previous Archive, and provide general maintenance and cleanup of the prior Archive. All work was performed under EPA Contract No. 68HERH22D0002, Delivery Order # 68HERH23F0306. This report contains seven sections and six appendices, as presented in the table of contents:

- Section 1 – Introduction
- Section 2 – Background Information
- Section 3 – AMA Data Sources
- Section 4 – QA Fixes and Data Changes
- Section 5 – Database Structure and Processing
- Section 6 – Final Database
- Section 7 – Final Output Data Files

2.0 Background Information

EPA first developed a master HAP Archive in 2001 to consolidate HAP measurements 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 (QA), and standardize this information for a master database, which was called the Phase I Archive. (Versions were identified by "Phases" prior to the 2020 Archive.)

During that time, EPA also began implementing its Integrated Urban Air Toxics Strategy, which was finalized in 1999.[1] In response, EPA and several state and 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 some agencies to submit their data to AQS. Table 2-1 presents a timeline of the Archive releases.

Table 2-1. Summary of Prior Archive Versions

Phase/Year	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
XIII	2020	1990 – 2017
XIV	2021	1990 – 2018
2020	2022	1990 – 2020
2021	2023	1990 – 2021
2022	2024	1990 – 2022
2023	2025	1990 – 2023

EPA completed the 2022 Archive in November 2024, which contained over 115 million HAP records from 1990 to 2022. The 2022 Archive was the twelfth 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 in their native sample durations (e.g., "1 HOUR") from AQS and other sources. Additionally, the Archive

[1] https://www.epa.gov/haps/integrated-urban-air-toxics-strategy

identified possible non-detect (ND) data measurement records that were substituted as one-half the method detection limit (MDL).

For the 2023 update, EPA requests ERG:

- Retrieve 1990-2023 ambient HAP monitoring data from EPA's 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 datum has a corresponding MDL;
- Identify sample values which were entered as one-half MDL (i.e., ND);
- Identify sample values below MDL (BMDL);
- 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 (LC) using meteorological data from collocated or nearby weather stations, where applicable; and
- Prepare data files and corresponding documentation for posting to EPA's Archive webpage.[2]

[2] https://www.epa.gov/amtic/amtic-ambient-monitoring-archive-haps

3.0 AMA Data Sources

For the 2023 Archive, there are 39 primary data sources used. Table 3-1 provides a summary of the final record counts of each data source used to populate the 2023 Archive. In total, there are over 130 million data records from 1990-2023.

Table 3-1. Data Source Information for HAP Records

Data Source	Data Years	# Sites	# HAPs	HAP Data Record Count	% of Records
Air Quality System Data	1990 – 2023	2,410	367	68,257,488	52.36%
Allegheny County, PA	2011 – 2022	19	24	20,172	0.02%
Baldwin Hills Air Quality Study	2012 – 2013	1	16	7,455	0.01%
Baltimore Inner Harbor Monitoring Study	2014 – 2015	6	1	1,734	0.00%
California Pesticides Monitoring Database	2010 – 2023	20	4	12,329	0.01%
CARB Special Study	2001 – 2002	1	34	2,098	0.00%
City of Ft. Worth, TX Natural Gas Air Quality Study	2010	8	49	5,455	0.00%
Colorado Boulder AIR	2017 – 2023	30	9	1,230,497	0.94%
Colorado Department of Public Health & Environment	2018	6	47	1,729	0.00%
Denka SPod Chloroprene	2016 – 2021	6	1	2,517	0.00%
EPA Passive Sampling Tubes Study	2013 – 2015	17	9	18,675	0.01%
EPA Refineries Fenceline Data	2016 – 2023	2,794	1	352,478	0.27%
EPA Region 3	2008 – 2020	2	14	3,633	0.00%
Ethylene Oxide Special Studies	2018 – 2022	90	1	3,236	0.00%
Houston Health Department	2019 – 2020	3	1	253,434	0.19%
Integrated Atmospheric Deposition Network Data	1999 – 2010	11	89	162,836	0.12%
Long Island Sound Tropospheric Ozone Study	2018 – 2023	8	41	266,692	0.20%
Louisiana Department of Environmental Quality	2010 – 2023	55	71	620,239	0.48%
Massachusetts Institute of Technology	1995 – 2023	1	7	1,260,726	0.97%
Michigan Community-Scale Air Toxics	2016 – 2017	3	9	168,343	0.13%
Minnesota Air Toxics Data	2008 – 2015	44	61	88,058	0.07%
Missouri Community-Scale Air Toxics Monitoring	2008 – 2009	7	3	9,612	0.01%
National Atmospheric Deposition Program Data	1996 – 2023	196	4	2,800,388	2.15%
NATTS Network Assessment	2003 – 2014	5	71	11,608	0.01%
National Oceanic and Atmospheric Administration	1990 – 2023	30	10	2,108,425	1.62%
National Park Service Studies	2011; 2013-2014; 2019; 2021	75	22	240,769	0.18%
New York State DEC	2014 – 2015; 2017 – 2019	92	36	5,658	0.00%
Ohio EPA	2017 – 2023	3	15	9,219	0.01%
Oregon Department of Environmental Quality	2012 – 2017	7	3	282	0.00%
Pennsylvania Marcellus Shale Study	2012 – 2013	6	39	14,793	0.01%
Phase V/VII Archive	1991 – 2010	144	164	201,862	0.15%
School Air Toxics Ambient Monitoring Program	2011 – 2012	6	80	800	0.00%
South Coast Air Quality Management District	1999 – 2023	146	103	24,536,867	18.82%
Sublette County, WY	2009 – 2010	14	42	37,398	0.03%
Texas A&M University	2019 – 2021	5	6	884	0.00%
Texas Commission on Environmental Quality	1992 – 2023	133	83	27,094,404	20.78%
Utah State University – Vernal	2012 – 2023	7	18	44,047	0.03%
Wisconsin Department of National Resources	2019 – 2022	3	13	3,092	0.00%
XAct Monitoring Data	2011 – 2023	11	17	501,422	0.38%
Total	**1990 – 2023**	**6,047**	**384**	**130,361,354**	**100%**

Information about each data source is presented in sections 3.1 – 3.39. In the Archive, the fields DATA_SOURCE and DATA_SOURCE_PULLDATE identify the data source and the date in which the source was obtained (e.g., "AQS" and "20250201" means AQS data retrieved on February 1, 2025).

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.[3] Additionally, ERG reviewed Community-Scale Air Toxics Ambient Monitoring (CSATAM) projects and checked to determine if the monitoring data were uploaded to AQS if appropriate.[4] Finally, ERG reviewed conference proceedings to identify data from air toxics projects that were not uploaded to AQS. In each of these situations, the project sponsor/awardees were contacted to obtain the data.

3.1 Air Quality System Data

AQS is EPA's official repository for ambient monitoring data. Users of AQS can download data from pre-generated data files,[5] monitor values reports,[6] the AQS Application Programming Interface (API),[7] or using standard/ad-hoc queries within the AQS data portal (which requires a user account).[8] Although not required for most air toxic programs, state, local, and tribal 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 and the Urban Air Toxics Monitoring Program (UATMP) are required to submit data to AQS. NATTS data are required to be submitted within 180 days at the end of the calendar quarter in which samples were collected (updated from 120 days at the end of the calendar quarter prior to 2018).[9]

AQS data for the 2023 data year were initially retrieved from the AQS data portal in February 2025 from the AMP501 ("Extract Raw Data") report. By using this report, the original data were obtained but not standardized. Additionally, data from 1990 – 2022 were also retrieved

[3] State and Local Monitoring Plans found here: https://www.epa.gov/amtic/state-monitoring-agency-annual-air-monitoring-plans-and-network-assessments

[4] More information on CSATAM projects is found here: https://www.epa.gov/amtic/community-scale-air-toxics-ambient-monitoring-csatam

[5] Pre-generated data files from AQS are found here: https://aqs.epa.gov/aqsweb/airdata/download_files.html

[6] Monitor Values Report for HAPs are found here: https://www.epa.gov/outdoor-air-quality-data/monitor-values-report-hazardous-air-pollutants

[7] More information about the AQS API is found here: https://aqs.epa.gov/aqsweb/documents/data_api.html

[8] The AQS Launch Web Application is found here: https://www.epa.gov/aqs

[9] As reported in Section 3.3.1.3.15 in the Technical Assistance Document for the NATTS Program, Revision 4. (https://www.epa.gov/system/files/documents/2022-08/NATTS-TAD-Revision-4-Final-July-2022-508.pdf)

to replace data found in the 2022 Archive (published November 2024). Over 68 million HAP records from 2,410 sites and 367 parameters were incorporated into the Archive. MDLs were populated for approximately 32% of all the HAP data records.

3.2 Allegheny County, PA

The Allegheny County Health Department (ACHD) in Pittsburgh, PA conducts metals and Volatile Organic Compound (VOC) sampling in the Pittsburgh area that are not uploaded to AQS. As such, ERG coordinates with ACHD to obtain these data and site metadata.[10] More information on the ACHD and their monitoring program can be found at: https://www.alleghenycounty.us/Services/Health-Department/Air-Quality/Monitored-Data. A total of 20,172 records from 2011through 2022 for nineteen sites[11] and 24 parameters were incorporated into the Archive. MDLs were provided for all records.

3.3 Baldwin Hills Air Quality Study

Los Angeles County, California, 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. These data were sent to ERG from the SCAQMD contractor for inclusion into the Archive, as they are not housed in AQS.[12] A total of 7,455 records from one site[13] and 16 parameters were incorporated into the Archive. Pollutant-specific MDLs were provided for all the metals data. However, the pollutant MDLs obtained from the Proton-Transfer-Reaction Time-of-Flight Mass Spectrometry (PTR-TOFMS) were obtained from the manufacturer.[14] More information on this study can be found at: https://www.sonomatech.com/projects/4111.

[10] Monitoring results provided by ACHD directly to EPA via e-mail from Mr. Darrell Stern, ACHD on 5/7/2019.
[11] The three sites are: Avalon (420030002); Lawrenceville (420030008); and Liberty (420030064).
[12] Email from Mr. Mike McCarthy, Sonoma Technology to Mr. Regi Oommen, ERG on 4/25/2016.
[13] 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."
[14] Per the manufacturer (https://www.ionicon.com/products/details/ptr-tof-6000-x2), the detection limit for the pollutants of interest (2,4-dinitrotoluene, benzene, naphthalene, 1,3-butadiene, acrolein, and total xylenes) is less than 1 pptv.

3.4 Baltimore Inner Harbor Monitoring Study

The Maryland Department of the Environment (MDE) and US EPA Region 3 oversaw a special hexavalent chromium monitoring study in the Baltimore Inner Harbor from 2014 to 2015. The study focused on establishing baseline air quality concentrations for Phase 1 construction activities.[15] These data were sent to ERG from the MDE contractor for inclusion into the Archive, as they are not housed in AQS.[16,17] A total of 1,734 records from six sites[18] and one parameter were incorporated into the Archive. Pollutant-specific MDLs were provided for all records.

3.5 California Pesticides Monitoring Database

The California Department of Pesticide Regulation maintains a Pesticide Air Monitoring Results database containing both preliminary and published data from pesticide air monitoring studies conducted throughout California.[19] This network consists of 20 monitoring sites measuring four specialized HAPs: bromomethane, carbon disulfide, 1,3-dichloropropene, and trifluralin. There were 12,329 records from 2010 through 2023 uploaded to the Archive. Pollutant-specific MDLs were provided for all records.

3.6 CARB Special Study

The California Air Resources Board (CARB) conducted an air toxics monitoring study from 2001-2002 at a school near large industrial sources in the community of Wilmington in Los Angeles, CA. This study was part of a larger statewide evaluation of the adequacy of the state's air quality monitoring network as required by the Children's Environment Health Protection

[15] https://mde.maryland.gov/programs/LAND/HazardousWaste/Documents/Allied/Allied%20Airmon%20Data%202014/DVR%2032172_07-09-14_2014-07-16.pdf

[16] Email from Mr. Ed Dexter, MDE to Mr. Regi Oommen, ERG on 2/2/2016.

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

[18] 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."

[19] https://www.cdpr.ca.gov/environmental-monitoring/air-monitoring/. 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 "06047309A" site is located in Merced County, CA (FIPS = 06047) and the site identifier is "309A."

Act.[20] There were 2,098 records collected for 34 pollutants uploaded to the Archive. Pollutant-specific MDLs were provided for all records.

3.7 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.[21] During this two-month study, 5,455 records were generated at eight monitoring sites[22] for 49 parameters. ERG, as the contract lab, received permission from DEM to include the data in the Archive. Pollutant-specific MDLs were provided for all records.

3.8 Colorado Boulder AIR

VOC HAP monitoring was conducted at 30 sites in Colorado near oil and gas activities. These counties contracted the monitoring and laboratory support services to Boulder AIR (Atmosphere Innovation Research), Inc. to evaluate concentrations of nine VOC HAPs for 10-minute sample durations every hour. More information about the sites and data collection can be found at: https://bouldair.com/. A total of 1,230,497 records from 2017 – 2023 were incorporated into the Archive. When using the data, the following disclaimer is made by City officials for the 2017 through 2020 datasets: "*Use of the City of Longmont, Broomfield, and Boulder air quality monitoring data is at the user's discretion and should be done with caution. The Cities provides no guarantee, either express or implied, as to the accuracy, reliability or completeness of raw data furnished. Further, the Cities shall not be liable under any circumstances for any direct, special, incidental or consequential damages with respect to any claim by any user or third party as a result of, or arising from, the use of the raw data.*" Pollutant-specific MDLs were provided for all records.

---

[20] More information can be found here: https://oehha.ca.gov/risk-assessment/report/childrens-environmental-health-program-report-legislature. 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 "06037WILM" site is located in Los Angeles County, CA (FIPS = 06037) and the site identifier is "WILM."

[21] The final report is found here: https://www.fortworthtexas.gov/files/assets/public/development-services/documents/gaswells/air-quality-study-final.pdf.

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

3.9 Colorado Department of Public Health & Environment

In 2015, the Colorado Department of Public Health and Environment (CDPHE) won a CSATAM grant to evaluate air toxics concentration gradients near roadways in Denver.[23] HAPs monitored included measurements of benzene, toluene, ethylbenzene, xylene, formaldehyde, and acetaldehyde during a 4-week study in 2018. These data were not available in AQS and were sent directly to ERG. A total of 1,729 records from 47 parameters and six existing sites were incorporated into the Archive. Pollutant-specific MDLs were provided for the data records.

3.10 Denka SPod Chloroprene

In response to concerns of elevated chloroprene concentrations from EPA modeling St. John The Baptist Parish, EPA began investigation the Denka Performance Elastomer ("Denka") facility in LaPlace, LA.[24] As part of the investigation, EPA began air sampling for chloroprene using sensor pod (SPod) and canister technologies.[25] A total of 2,517 records from six sites from 2016 through 2021 were incorporated into the Archive.[26] Pollutant-specific MDLs were provided for the data records.

3.11 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://www.tandfonline.com/doi/full/10.1080/10962247.2016.1184724. These data were sent to ERG from the City of Philadelphia for inclusion into the Archive, as they are not in AQS.[27] A total of 18,675 records from 17 sites and nine parameters were incorporated into the Archive. [28] Pollutant-specific MDLs were provided for all records.

---

[23] https://www.epa.gov/sites/default/files/2020-01/documents/colorado_project_narrative.pdf
[24] https://www.epa.gov/la/laplace-louisiana-frequent-questions
[25] https://www.epa.gov/la/denka-air-monitoring-data-summaries
[26] 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 Railroad site (Site ID = RAIL), located in St. John the Baptist Parish, LA (FIPS code = 22095) is assigned 22095RAIL.
[27] Email from Ms. Hallie Weiss, City of Philadelphia to Mr. Regi Oommen, ERG on 12/12/2017.
[28] 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."

3.12 EPA Refineries Fenceline Data

In 2015, the U.S. EPA issued the Petroleum Refinery Sector Rule, an air toxics regulation that required, among other things, refineries to continually monitor the concentration of benzene emissions along their property boundaries (i.e., fenceline).[29] For this reason, these data are not reported in the final output data files. Refineries began formerly reporting monitoring data to EPA in May 2019 (although some informally reported data as early as 2016) and continue to report on a quarterly basis.[30] A total of 352,478 records from 2,794 site locations were incorporated into the Archive.[31] Pollutant-specific MDLs were not provided with the records.

3.13 EPA Region 3

The West Virginia Department of Environmental Quality (WVDEP) conducted multi-year (2008 through 2020) metals measurements at two sites in West Virginia targeting specific sources of interest. Filter samples were sent for analysis to the EPA Region 3 lab, who also coordinated these data to be sent to ERG for inclusion into the Archive, as they are not housed in AQS.[32] A total of 3,633 records from two sites and 14 parameters were incorporated into the Archive. Pollutant-specific MDLs were provided for all records.

3.14 Ethylene Oxide Special Studies

In December 2016, EPA's Integrated Risk Information System (IRIS) program released an updated assessment of the carcinogenicity of inhaled ethylene oxide (EtO).[33] The new Unit Risk Estimate (URE) was integrated into EPA's National Air Toxics Assessment (NATA) in 2018 (now AirToxScreen). As a result, EtO cancer risk results were elevated (i.e., greater than 100-in-1-million) at 25 areas of the country.[34]

Special ambient air monitoring EtO studies that began in 2018 are presented below.

---

[29] https://www.epa.gov/stationary-sources-air-pollution/petroleum-refinery-sector-rule-risk-and-technology-review-and-new

[30] https://awsedap.epa.gov/public/extensions/Fenceline_Monitoring/Fenceline_Monitoring.html

[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 fenceline monitor at location 1 (Site ID = SHEL_01), located in Mobile County, AL (FIPS code = 01097) is assigned 01097SHEL_01.

[32] Email from Mr. Howard Schmidt, EPA Region 3 to Mr. Regi Oommen, ERG on 2/27/2018.

[33] https://cfpub.epa.gov/ncea/iris2/chemicalLanding.cfm?substance_nmbr=1025.

[34] "Locations and names of sterilizers where there are elevated risks at or above 100/million to nearby communities" found here: https://www.epa.gov/hazardous-air-pollutants-ethylene-oxide/forms/ethylene-oxide-risk-commercial-sterilizers#facility-list

- Lakewood, CO: The Colorado Department of Public Health and Environment (CDPHE) monitored outdoor air quality at 12 locations in the vicinity of the Terumo BCT sterilization facility. Air monitoring was conducted before and after additional controls were installed by the facility.[35]

- Georgia (Multiple Counties): The Georgia Department of Natural Resources (DNR) monitored outdoor air at several locations in the vicinity of EtO emissions sources in Cobb (Sterigenics), Fulton (Sterilization Services of Georgia), and Newton (BD Covington and Global Distribution Center) Counties, as well as a background site in Coffee County. Air monitoring was conducted before and after additional controls were installed by some of the facilities. More information can be found at: https://epd.georgia.gov/ethylene-oxide-information.

- Willowbrook, IL: The US EPA monitored near the Sterigenics facility to better understand the levels of EtO in the outdoor air. The first monitors began collecting air samples on November 13, 2018. Air samples were collected every three days with a 24-hour sampling duration for 4.5 months. More information can be found at: https://www.epa.gov/il/outdoor-air-monitoring-data-willowbrook-community.

- Grand Rapids, MI: The Michigan Department of Environmental Quality (DEQ) monitored outdoor air near Viant Medical. Phase 1 sampling took place at the facility, while Phase 2 sampling took place in the community near the facility.[36] More information on the sampling can be found at: https://www.michigan.gov/egle/about/organization/Air-Quality/facility-specific-info/viant-medical.

- Utah: Under a Community-Scale Air Toxics Ambient Monitoring (CSATAM) grant for RFA 2020, the Division of Air Quality monitored at eighteen locations for EtO near commercial sterilizers in the Salt Lake City, UT area. A total of 559 records from January through September 2022 were sent to EPA for inclusion into the Archive.[37]

- West Virginia (Multiple Locations): The WVDEP conducted short-term EtO monitoring at nine locations in the Charleston and Institute areas near EtO emissions sources from January through July 2022. More information can be found at: https://dep.wv.gov/key-issues/Pages/EtO.aspx

A total of 3,236 records from 90 sites for EtO from 2018 through 2022 were incorporated into the Archive.[38] MDLs were provided for all records.

[35] https://drive.google.com/file/d/173g_kSWWXmZnH0q2Qlxg1g0qXNmeU1KO/view

[36] E-mail with EtO measurements sent from Ms. Amy Robinson, MI EGLE to Mr. Regi Oommen, ERG on 11/10/2025.

[37] E-mail with EtO measurements sent from Ms. Isabel Jaramillo, UT DAQ to Mr. Regi Oommen, ERG on 11/23/2023.

[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 "26081VIA2" site is located in Kent County, MI (FIPS = 26081) and the site identifier is "VIA2."

3.15 Houston Health Department

The Houston Health Department received a CSATAM award in 2017 to characterize formaldehyde concentrations in the Houston Ship Channel.[39] The project used an emerging technology called FluxSense, which provided real-time continuous formaldehyde concentrations at three locations from 2019 through 2020. A total of 253,434 records at three sites were incorporated into the Archive.[40] The MDL was provided for all records.

3.16 Integrated Atmospheric Deposition Network 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 several satellite stations to provide more spatial detail for deposition. The master stations allow for 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 1999 – 2010 for organic, polycyclic aromatic hydrocarbon (PAH), and polychlorinated biphenyls (PCB) compounds were retrieved from the IADN website.[41] Recent data (2011 – present) only covers sites in Canada. A total of 162,836 records from 11 sites and 89 parameters were incorporated into the Archive.[42] Pollutant-specific MDLs were provided for all records.

3.17 Long Island Sound Tropospheric Ozone Study

The Long Island Sound Tropospheric Ozone Study (LISTOS) is a multi-agency collaborative study focusing on the Long Island Sound and the surrounding coastlines.[43] A total of 266,692 records from 8 sites and 41 parameters from 2018 through 2023 were incorporated into the Archive. Pollutant-specific MDLs were provided for approximately 30% of the data

[39] Project Plan: https://www.epa.gov/sites/default/files/2020-01/documents/city_of_houston_project_plan.pdf.
[40] Concentration data were provided by Ms. Lilian Mojica/Houston Health Department to Mr. Regi Oommen/ERG via e-mail on August 25, 2021.
[41] https://www.epa.gov/great-lakes-monitoring/integrated-atmospheric-deposition-network
[42] 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."
[43] https://www-air.larc.nasa.gov/missions/listos/index.html

records, and the remainder were populated with default federal MDLs based on the method code. The data are maintained by the National Aeronautics and Space Administration (NASA).

3.18 Louisiana Department of Environmental Quality

The Louisiana DEQ collects canister data for VOCs and PAHs analysis not uploaded to AQS.[44] A total of 620,239 records from 55 sites and 71 parameters from 2010 through 2023 were incorporated into the Archive. Pollutant-specific MDLs were not provided and were populated with default federal MDLs based on the method code.

3.19 Massachusetts Institute of Technology

The Advanced Global Atmospheric Gases Experiment (AGAGE) is a network of global sites measuring the composition of the global atmosphere since 1978.[45] One site in Trinidad Head, CA measures seven HAPs: bromomethane, carbon tetrachloride, chloroform, chloromethane, methyl chloroform, methylene chloride, and tetrachloroethylene. The data are maintained by the Massachusetts Institute of Technology (MIT). A total of 1,260,726 records were retrieved from 1995 through 2023. Although pollutant-specific MDLs were not provided, it was assumed that values reported as "-99.990" were non-detects; null values were also reported. As such, these concentrations were reported as zero and flagged accordingly. Additionally, a value of 5 ppt was assigned as a default MDL, which was half of the lowest reported concentration in the entire dataset.

3.20 Michigan Community-Scale Air Toxics

In 2015, the Michigan DEQ won a CSATAM grant to evaluate air toxics concentrations near roadways in Detroit, MI.[46] The HAPs monitored included continuous acrolein, benzene, toluene, ethylbenzene, *m,p*-xylene, and *o*-xylene. In addition to this effort, a 3-month intensive study was conducted to include a collection of carbonyl HAPs (i.e., acetaldehyde, formaldehyde, and propanal). These data were not available in AQS and were sent directly to ERG. A total of 168,343 records from three existing sites and nine parameters from 2016 and 2017 were

[44] https://deq.louisiana.gov/page/ambient-air-monitoring-data-reports
[45] https://agage.mit.edu/
[46] https://www.epa.gov/sites/default/files/2020-01/documents/michigan_csatmg_near-road_narrative.pdf

incorporated into the Archive. Pollutant-specific MDLs were not provided for the data records and federal MDL values for the same method code were used as a default.

3.21 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, ERG was alerted to data reporting issues that occurred when reporting to AQS, such as truncation of concentrations, missing MDLs, and revised data. As such, MNPCA removed that data from AQS and provided their entire dataset from 2008 – 2015 to ERG for inclusion in the Phase XIV Archive.[47] More information on the MNPCA air toxics monitoring program can be found at: https://www.pca.state.mn.us/air-water-land-climate/air-quality-monitoring. A total of 88,058 records from 44 sites and 61 parameters were incorporated into the Archive. Pollutant-specific MDLs were provided for all records.

3.22 Missouri Community-Scale Air Toxics Monitoring

In 2007, the Missouri DNR won a CSATAM grant to evaluate air toxics concentrations in the St. Louis, MO-IL area. The monitored HAPs included 24-hour measurements of arsenic (PM_{10}), lead (PM_{10}), and selenium (PM_{10}) at four locations and continuous measurements using multi-metals continuous measurements systems at seven locations.[48] These data were not available in AQS and were sent directly to ERG. A total of 9,612 records from 2008 through 2009 were incorporated into the Archive. Pollutant-specific MDLs were provided for all records.

3.23 National Atmospheric Deposition Program Data

The National Atmospheric 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); 5) the National Trends Network (NTN); and 6) the Mercury Litterfall Network (MLN). Data from 1996 through 2023 from the MDN and AMNet networks were downloaded from https://nadp.slh.wisc.edu/networks/. A total

[47] Email from Ms. Kellie Gavin, MNPCA to Mr. Regi Oommen, ERG on 3/5/2018.
[48] https://www.epa.gov/sites/default/files/2020-01/documents/114modnr.pdf.

of 2,800,388 records from 196 sites and 4 parameters were incorporated into the Archive.[49] Pollutant-specific MDLs were provided for all records.

3.24 NATTS Network Assessment

In Fall 2017, ERG, under contract to EPA, prepared a final report for the National Air Toxics Trends Sites (NATTS) Network. As per the requirements of the NATTS Network, participating sites are to report data to AQS. During this data review, several concentrations reported to AQS were identified as incorrect and were 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 ERG for inclusion into this Archive. The NATTS Network Assessment covers measurements from the 2003 through 2014. More information on the NATTS program can be found at: https://www.epa.gov/amtic/air-toxics-ambient-monitoring#natts. A total of 11,608 records from five sites and 71 parameters from 2003 through 2014 were incorporated into the Archive. Pollutant-specific MDLs were provided for all records.

3.25 National Oceanic and Atmospheric Administration

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

- Chromatograph for Atmospheric Trace Species (CATS): Long-term in-situ hourly measurements for halocarbons, including carbon tetrachloride, chloromethane, and methyl chloroform from 1998 through 2020 at three US 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/.

- Halocarbon and other Atmospheric Trace Species (HATS): 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, CO. This system uses two standard reference gases for calibration and has

[49] 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."

been in operation since 1995. Weekly, 5-minute measurement data of seven HAPs (benzene, bromomethane, carbonyl sulfide, chloromethane, methyl chloroform, methylene chloride, and tetrachloroethylene) from eight sites from 1991 – 2023 for all pollutants, except benzene, were retrieved at: https://gml.noaa.gov/aftp/data/hats/solvents/. The benzene results were sent directly from the Principal Investigator to ERG.[50] More information can be found at: http://www.esrl.noaa.gov/gmd/hats/flask/flasks.html.

- Radiatively Important Trace Species (RITS): The data reported were from samples collected every day in concurrent, flask pairs that were later analyzed on a gas chromatograph with GC-ECD located in Boulder, CO. Hourly measurements of carbon tetrachloride at three US sites (Mauna Loa, HI; Niwot Ridge, CO; and Pt. Barrow, AK) from 1990-2001 were retrieved at: https://gml.noaa.gov/dv/data/. More information can be found at: https://gml.noaa.gov/hats/insitu/insitu.html.

- SURFACE: Five-minute data for five HAPs were reported from samples collected in programmable flask packages (PFP) using programmable compressor packages (PCP) at 23 sites from 2015-2023. Over 108,000 data records were coalesced from data files obtained from: https://gml.noaa.gov/dv/data/index.php?type=Surface%2BPFP.

A total of 2,108,425 records from 1990 through 2023 for 30 sites and ten parameters were incorporated into the Archive.[51] Pollutant-specific MDLs were provided for all records.

3.26 National Park Service Studies

The National Park Service (NPS) has sponsored several air toxics studies since 2011, primarily in remote locations in Colorado, New Mexico, and North Dakota.[52] These data were not available in AQS and were obtained by ERG via the project lead. A total of 240,769 records at 75 sites for twenty-two pollutants from 2011, 2013-2014, 2019, and 2021 were incorporated into the Archive.[53] Pollutant-specific MDLs were provided for all records.

[50] Benzene data were provided by Stephen Montzka, A. NOAA/Earth System Research Laboratory/Global Monitoring Division to Regi Oommen/ERG. February 18, 2022, via e-mail.

[51] 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.

[52] Emails from Dr. Barkley Sive, National Park Service to Mr. Regi Oommen, ERG on 1/12/2023 and 3/9/2023.

[53] 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 Bio Building sampling site at the Carlsbad Caverns site (Site ID = BIOB), located in Eddy County, NM (FIPS code = 35015) is assigned 35015BIOB.

3.27 New York State DEC

In 2014, the New York State Department of Environmental Conservation (DEC) and local community groups conducted a special study to determine whether the levels of air pollutants from motor vehicles were a public health concern in the residential neighborhood near the Peace Bridge in Buffalo, NY.[54] These data were obtained by ERG via the project website.

New York State DEC also completed a community air quality study in Albany, NY in response to community concerns.[55] Sorbent material contained within a stainless-steel tube were deployed at 2-week intervals at different locations within the study area. The tubes were analyzed for select VOC HAPs at 91 locations.[56]

A total of 5,658 records at 92 locations for 36 pollutants from 2014 through 2015 and 2017 through 2019 were incorporated into the Archive. Pollutant-specific MDLs were provided for all records.

3.28 Ohio Environmental Protection Agency (OHEPA)

Since 2017, Ohio EPA (OHEPA) has conducted ambient air monitoring near the Republic Steel manufacturing facility in Canton, Ohio.[57] A total of 9,219 records at three locations for 15 pollutants from 2017 through 2023 were incorporated into the Archive.[58] Pollutant-specific MDLs were not provided for the data records and federal MDL values for the same method code were used as a default.

3.29 Oregon Department of Environmental Quality

In Summer 2019, EPA was alerted by the Oregon Department of Environmental Quality (ODEQ) of incorrectly carbonyl compound concentrations (i.e., acetaldehyde, formaldehyde, and propionaldehyde) submitted to AQS from 2012 through 2017. A total of 282 revised concentrations from 7 sites and three parameters were sent by ODEQ and incorporated into the Archive.[59] Pollutant-specific MDLs were provided for all records.

---

[54] https://extapps.dec.ny.gov/docs/air_pdf/pbfinalreport.pdf.
[55] https://extapps.dec.ny.gov/docs/air_pdf/albanysouthendreport.pdf.
[56] Email from Mr. Dirk Felton to Mr. Regi Oommen, ERG on 3/18/2024.
[57] https://epa.ohio.gov/divisions-and-offices/air-pollution-control/reports-and-data/special-sampling-projects
[58] A unique AMA_SITE_CODE identifier was assigned to one site based on the 2-digit state code, 3-digit county code, and the unique site code. The Youtz sampling site (Site ID = YOUT), located in Stark County, OH (FIPS code = 39151) is assigned 39151YOUT.
[59] Email from Mr. Chris Moore, ODEQ to Mr. Regi Oommen, ERG on 9/27/2019.

3.30 Pennsylvania Marcellus Shale Study

The Pennsylvania Department of Environmental Protection (DEP) evaluated the impacts from oil and gas wells in the Marcellus Shale area of Pennsylvania through HAP measurements from 2012 through 2013. The sampling results provided basic information about the types of pollutants emitted into 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 can be found at: https://www.dep.pa.gov/Business/Energy/OilandGasPrograms/OilandGasMgmt/Oil-and-Gas-Related-Topics/Pages/Air.aspx. A total of 14,793 records for six sites and 39 parameters were incorporated into the Archive.[60] Pollutant-specific MDLs were provided for all records.

3.31 Phase V/VII Archive

The Phase V Archive originally consisted of over nine million daily concentration records for HAPs. The initial compilation of this air toxics Archive began in the mid-1990s, consisting of datasets from several state and local agencies culminating in a 2001 release. Many of these datasets were eventually placed into AQS or were subsequently deleted. A portion of Phase V data records were never placed in AQS and remain in the Archive. The Phase VII Archive consists of historical data that have been invalidated and are no longer in AQS. Retained for posterity, nearly all these records are invalidated VOC data originally submitted by the Kentucky Department of Environmental Services. A total of 201,862 records from 1991 through 2010 from 144 sites and 164 parameters were incorporated into the Archive. Pollutant-specific MDLs were provided for most records.

3.32 School Air Toxics Ambient Monitoring Program

As part of an air toxics monitoring initiative in 2009, EPA, state, and local air pollution control agencies monitored air toxics in the outdoor air around schools. EPA selected schools after evaluating several 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

[60] 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.

series on air toxics at schools, and information from state and local air pollution agencies. Phase 1 sampling took place in 2009 – 2010 at 59 schools across the US, while Phase 2 sampling followed up at 22 schools in 2010 – 2012. Nearly all the data resides in AQS, except for 1) special VOC measurements taken at two schools during the Phase 2 sampling: Enterprise High School in Enterprise, MS and Temple Elementary in Diboll, TX and 2) some records from the four Alabama schools. These missing data from 2011-2012 were retrieved by EPA and formatted for inclusion into the Archive. More information can be found at: https://www3.epa.gov/air/sat/. A total of 800 records from six sites and 80 parameters were incorporated into the Archive. Pollutant-specific MDLs were provided for all records.

3.33 South Coast Air Quality Management District

The SCAQMD sponsors the Multiple Air Toxics Exposure Study (MATES) which characterizes air quality data. MATES-II (1999), MATES-III (2004 – 2007), MATES-IV (2012 – 2013), and MATES-V (2018 – 2019) data were obtained from SCAQMD. Over the course of these studies, a total of 193,167 records from 95 pollutants measured at 23 sites were incorporated into the Archive. More information can be found at: http://www.aqmd.gov/home/air-quality/air-quality-studies/health-studies.

SCAQMD also launched additional air quality studies described below.

- Community Air Toxics Initiative (CATI): SCAQMD measured levels of hexavalent chromium in ambient air near several industrial facilities in the Compton area from June 2017 to November 2018. This monitoring effort at 12 sites identified and prioritized high-risk facilities with the potential to emit hexavalent chromium, then used additional technology to confirm specific sources of emissions.[61] A total of 1,278 records were incorporated into the Archive. More information can be found at: http://www.aqmd.gov/docs/default-source/air-toxics-initiative/compton/updated-air-monitoring-plan.pdf?sfvrsn=14.

- Ethylene Oxide: SCAQMD began investigating facilities that emit ethylene oxide in March 2022. Fourteen monitoring locations were placed downwind of facilities emitting ethylene oxide, totaling 1,131 records.[62] More information can be found at: https://www.aqmd.gov/home/eto.

[61] 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.

[62] 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, #1 Kingsview Ave), located in Los Angeles County, CA (FIPS code = 06037) is assigned 06037PS01.

- Exide Technologies: Since 2006 SCAQMD has monitored ambient arsenic (TSP [Total Suspended Particulate]) and lead (TSP) near Exide Technologies, a facility that recovers lead from recycled automotive batteries. A total of 17,581 records from five sites were incorporated into the Archive.[63] More information can be found at: https://www.aqmd.gov/home/news-events/community-investigations/exide-updates/compliance-permitting-toxics.

- Jordan Downs: SCAQMD measured levels of ambient arsenic (PM_{10}) and lead (PM_{10}) every three days during the Jordan Downs Redevelopment Cleanup, taking place from April through May 2016. A total of 30 records at two locations were processed for inclusion in the Archive.

- Paramount: As part of the ongoing investigation to identify and address sources of hexavalent chromium in the City of Paramount, the SCAQMD, with assistance from CARB, conducted mobile air sampling for hexavalent chromium, other TSP metals, and $PM_{2.5}$ metals at schools in Paramount, CA. This study assessed potential elevated levels of hexavalent chromium at local schools. Sampling began in 2013 at two sites and increased to 49 sites by 2020. A total of 16,512 records for 24 pollutants were incorporated into the Archive. More information can be found at: http://www.aqmd.gov/home/news-events/community-investigations/air-monitoring-activities.

- State Rule 1180 Community Air Monitoring Program: Rule 1180 mandates the implementation of real-time observations of air quality at or near the fenceline of all major refineries in the South Coast Basin, and in nearby communities. Sampling began in 2020 at eleven sites in community locations. A total of 24,302,406 records for 14 pollutants were incorporated into the Archive.[64] More information can be found at: http://www.aqmd.gov/docs/default-source/rule-book/support-documents/1180/rule-1180-guidelines.pdf?sfvrsn=8.

- West Dominguez: SCAQMD monitored hexavalent chromium in ambient air near several industrial facilities in the West Dominguez area from June 2019 to June 2022. This monitoring effort at 14 sites identified and prioritized high-risk facilities with the potential to emit hexavalent chromium, then used additional technology to confirm specific sources of emissions. A total of 1,998 records were incorporated into the Archive.[65] More information can be found at: http://www.aqmd.gov/home/news-events/community-investigations/west-rancho-dominguez-emissions-investigations/reports-data-assessments

[63] 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 "Exide Mid" located in Los Angeles County, CA (FIPS code = 06037) is assigned 06037EMID.

[64] 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 Hudson Air Monitoring Station site, located in Los Angeles County, CA (FIPS code = 06037) is assigned 06037CHUD.

[65] 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 #1 (WD01), located in Los Angeles County, CA (FIPS code = 06037) is assigned 06037WD01.

- Western Riverside County: In 2008, SCAQMD identified cement production as a source of elevated levels of hexavalent chromium in the western areas of Riverside and San Bernardino Counties. SCAQMD sampled at 17 sites in those areas from 2008 – 2011.[66] A total of 2,764 records were incorporated into the Archive. 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.

A total of 24,536,867 records from 146 sites and 103 parameters were incorporated into the Archive. Pollutant-specific MDLs were provided for all records.

3.34 Sublette County, WY

The Wyoming DEP monitored near oil and gas wells from February 2009 to February 2010. A total of 37,398 records from 14 sites and 42 parameters were incorporated into the Archive.[67,68] Pollutant-specific MDLs were provided for all records. More information on the sampling design and analysis of the measurements can be found at: https://fossil.energy.gov/ng_regulation/sites/default/files/programs/gasregulation/authorizations/2013/applications/sierra_club_13-69_venture/exhibits_62_76.pdf.

3.35 Texas A&M University

Ambient air measurements were conducted by Texas A&M University near oil and gas wells in Texas from 2019 to 2021. The study tracked selected non-methane hydrocarbons at locations throughout a busy central production area of the Eagleford Shale. A total of 884 records from five sites and six parameters were incorporated into the Archive.[69,70] Pollutant-specific MDLs were provided for all records.

---

[66] 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.

[67] 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."

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

[69] 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 "48123NORD" site is located in DeWitt County, TX (FIPS = 48123) and the site identifier is "NORD."

[70] https://www.mdpi.com/2073-4433/14/4/744.

3.36 Texas Commission on Environmental Quality

The Texas Commission on Environmental Quality (TCEQ) maintains a large database of ambient HAP measurements on its Texas Air Monitoring Information System (TAMIS) website (http://www17.tceq.texas.gov/tamis/index.cfm?fuseaction=home.welcome), which allows for ad-hoc queries. 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 AQS data over TAMIS for non-identical overlaps. A total of 27,094,404 records from 1992 through 2023 for 133 sites and 83 parameters were incorporated into the Archive. The pollutant-method specific MDLs were pulled from the TAMIS website.

3.37 Utah State University – Vernal

Utah State University (USU) in Vernal, UT collects HAP measurements during wintertime in and around oil and gas wells in northeastern Utah. This is a cooperative effort with Uintah and Duchesne Counties, local industry, the Utah Division of Air Quality, the Ute Indian Tribe, the Tri-County Health Department, research teams at other Utah universities and universities around the U.S., and federal agencies (i.e., Bureau of Land Management [BLM], EPA, and Department of Energy [DOE]). A total of 44,047 HAP concentrations from seven sites and 18 parameters from 2012 to 2023 were incorporated into the Archive.[71,72] Pollutant-specific MDLs were provided for all records. More information on the sampling program can be found at: https://www.usu.edu/binghamresearch/cumulative-research-summary.

3.38 Wisconsin Department of Natural Resources

The Wisconsin DNR are federally required to conduct Enhanced Ozone Monitoring (EOM) to monitor for ozone and precursors at locations along the Lake Michigan shoreline.[73] This additional monitoring included three stationary sites and one portable site for 13 HAPs not in AQS. A total of 3,092 records for 2019 through 2022 were incorporated into the Archive. Pollutant-specific MDLs were provided for all records.

---

[71] 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."

[72] Emails from Mr. Seth Lyman, USU to Mr. Regi Oommen, ERG on 4/19/2019 and 3/5/2020.

[73] https://wi-dnr.widencollective.com/portals/iwvftorq/AirMonitoringData.

3.39 XAct Monitoring Data

The U.S. EPA purchased XAct Monitoring Measurement Systems from the School Air Toxics Ambient Monitoring Program. The purpose of these continuous, multi-metal measurement systems is to aid EPA, state, and local air agencies to target and identify source characterization signatures of HAP metal-emitting facilities. ODEQ used the XAct system in a 2011 study of PM metals. Data were sent by ODEQ to ERG and were processed for the Archive.[74] After this study, EPA Region 5 conducted several monitoring campaigns, ranging from two- to six-months from 2012 to 2023 in Illinois, Indiana, Michigan, and Ohio using XAct for targeting specific sources. A total of 501,422 records from eleven sites and 17 parameters were incorporated into the Archive.[75] Pollutant-specific MDLs were provided for all records.

[74] Email from Ms. Aida Biberic, ODEQ to Mr. Dave Shelow, EPA on 6/24/2013.

[75] 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."

4.0 QA Fixes and Data Changes

After an initial assessment of all available data, 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)."

- Non-Detects (ND): NDs are to be reported in AQS as zeroes, with the appropriate flag of "ND" populated. However, several sample concentration values in AQS were surrogate values which equated to one-half MDL. The concentrations for these records were changed to 0, the SAMPLE_VALUE_FLAG field was populated with "ND", and the COMMENT field was populated documenting the record update. The following approach was used to identify these records:

    1. Identify all records in which the concentration is one-half MDL.
    2. By site code, pollutant, and year, summarize counts of sample dates, sample values, ND flags, one-half MDLs, and BMDL flags.
    3. Identify site code, pollutant, and year combinations in which all the BMDL flag counts is equal to the count of one-half MDL.
    4. For the records in (3), if the count of BMDL flags is equal to the counts of one-half MDL records AND if NDs are not reported, mark as being an incorrectly substituted record for NDs.

- Negative Concentrations: Over 765,107 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_10 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" (i.e., "Miscellaneous Void") in the AQS_NULL_DATA_CODE data field and as "INV" (i.e., "Invalidated") in AQS_QUALIFIER_07 data field. Similarly, the State of Kentucky invalidated all VOC measurements analyzed by their laboratory since 1995 due to laboratory error ("AR" code). All hexavalent chromium concentrations prior to 2005, all PAHs (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 conditions (STD) and LC for the same measurement. Both conditions were retained in the Archive, while STD were invalidated.

- Revised Concentrations: Through the NATTS Network Assessment and UATMP, a 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 the sampling frequency code 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:

    1. 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.
    2. MS: This qualifier code is used to designate reported concentrations that are substituted with one-half MDL. Concentration records were deemed "inconsistent" if they were assigned "MS," but the reported values were not equal to one-half MDL. As such, the qualifier code flag was removed.
    3. 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.
    4. 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 five times the MDL. As such, the qualifier code flag was removed.
    5. SQ: This qualifier code is used to designate reported concentrations compared to the Sample Quantitation Limit (SQL), which is 3.18 times the MDL.[76] 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 through the quality checks:

- 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 an LC and an STD. 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 LC record was retained, and the STD 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 AQS_QUALIFIER_06 field to identify these invalidated records. Table 4-1 summarizes

[76] As reported in Section 3.3.1.3.15 in the Technical Assistance Document for the NATTS Program, Revision 4. (https://www.epa.gov/system/files/documents/2022-08/NATTS-TAD-Revision-4-Final-July-2022-508.pdf)

the fate of multiple reporting for the xylene records, where "X' indicates there is a valid concentration. Appendix A. Overlapping Records presents the records that were invalidated.

Table 4-1. Xylene Overlap Scenarios

Overlap Scenarios					
xylene(s) (45102)	***m,p*-xylene (45109)**	***o*-xylene (45204)**	***m*-xylene (45205)**	***p*-xylene (45206)**	**Fate**
X	X	X	X	X	Invalidate 45102, 45109
X	X	X	X		Invalidate 45102, 45205
X	X	X			Invalidate 45102
X	X		X		Invalidate 45109, 45205
X	X			X	Invalidate 45109, 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	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 because 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. Invalidated Records 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. Sampling Frequency Code Corrections 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 one-half 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 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, the Office of Air Quality Planning and Standards (OAQPS) 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 recategorized as verified and moved or 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.[77] Lastly, "PC" ("potential calculation error") is assigned in this field. Appendix D. Questionable Values and Incorrectly Submitted One-Half MDL Concentrations 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. Negative Concentrations and Incorrectly Assigned Qualifier Codes for "MD," "ND," and "SQ" presents the records that were identified.

[77] "Data Quality Evaluation Guidelines for Ambient Air Acrolein Measurements." OAQPS. December 17, 2010. Found at: https://www.epa.gov/sites/default/files/2021-03/documents/20101217acroleindataqualityeval.pdf

5.0 Database Structure and Processing

All data were uploaded into Microsoft Structured Query Language (SQL) Server for pre-processing and setting data field conventions. The 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 the HAP measurements from the TAMIS website were uploaded to the SQL Server and compared to the AQS data to identify missing and overlapping data. The SQL Server also offers the ability to create primary key constraints on tables to ensure no duplication of records. In total, over 130 million HAP records are in the blended master database.

After merging the data, ERG calculated the "SAMPLE_VALUE_REPORTED" to a standardized concentration in $\mu g/m^3$, using the following procedures outlined in Table 5-1:

Table 5-1. Unit Conversion to $\mu g/m^3$

AQS_UNIT_CODE	Description	Conversion to $\mu g/m^3$
001	$\mu g/m^3$, STD	no change
002	$\mu g/m^3$, 0° C	$(\mu g/m^3, 0°\ C * 273K)/298K$
003	ng/m^3, STD	$ng/m^3, STD * 10^3$
004	ng/m^3, 0° C	$(ng/m^3, 0°\ C * 273K * 10^3)/(298K)$
007	$ppmv$	$(ppmv * mw)/(10^3 * 24.45)$
008	$ppbv$	$(ppbv * mw)/24.45$
074	pg/m^3, STD	$pg/m^3, STD * 10^6$
078	$ppbC$	$(ppbC * mw)/(24.45 * \#\ of\ carbons)$
101	$ppmC$	$(ppmC * mw)/(24.45 * \#of\ carbons * 10^3)$
105	$\mu g/m^3$, LC	$(\mu g/m^3, LC * local\ temp.in\ K * 760\ mm\ Hg\ )/(298K * local\ press.in\ mm\ Hg)$
108	ng/m^3, LC	$(\mu g/m^3, LC] * local\ temp.in\ K * 760\ mm\ Hg * 10^3)/(298K * local\ press.in\ mm\ Hg)$
121	$pptv$	$(ppbv * mw * 10^3)/(24.45)$
174	pg/m^3, 0° C	$(pg/m^3, 0°\ C * 273K * 10^6)/(298K)$

The 2023 Archive 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 Archive table. ("Foreign keys" are columns in a relational database table that provides a link between data in two tables.) To translate the data in the Archive, ERG developed 10 data dictionary tables. These dictionaries describe and standardize the raw data and provide 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 10 data dictionaries are presented in Sections 5.1 through 5.10 below.

5.1 Site Information

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

Table 5-2. Site Information Data Fields

Data Field	Data Description
AMA_SITE_INFORMATION*	Site identifier comprised of STATE_FIPS, COUNTY_FIPS, and LOCAL_SITE_ID
STATE_FIPS[1]	Federal Information Processing System (FIPS) 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_2010	U.S. census tract identifier for year 2010
CENSUS_TRACT_ID_2020[1]	U.S. census tract identifier for year 2020
CENSUS_TRACT_POPULATION_2010	U.S. census tract population for year 2010
CENSUS_TRACT_POPULATION_2020	U.S. census tract population for year 2020
CENSUS_BLOCK_GROUP_ID_12_2010	U.S. census block group identifier for year 2010
CENSUS_BLOCK_GROUP_ID_12_2020[1]	U.S. census block group identifier for year 2020
CENSUS_BLOCK_GROUP_POPULATION_2010	U.S. census block group population for year 2010
CENSUS_BLOCK_GROUP_POPULATION_2020[1]	U.S. census block group population for year 2020
CENSUS_BLOCK_ID_14_2010	U.S. census block identifier for year 2010
CENSUS_BLOCK_ID_14_2020[1]	U.S. census block identifier for year 2020
CENSUS_BLOCK_POPULATION_2010	U.S. census block population for year 2010
CENSUS_BLOCK_POPULATION_2020[1]	U.S. census block population for year 2020
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
SUPPORT_AGENCY_CODE[1]	Code for the support agency
SUPPORT_AGENCY[1]	Support agency name
NATTS_SITE_FLAG	Identifies the site as a NATTS
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 a CASTNET site

Data Field	Data Description
PM_SUPERSITES_SITE_FLAG	Identifies the site as a 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 CSATAM site
CANDIDATE_NCORE_SITE_FLAG	Identifies the site as a potential National Core (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
REFINERIES_FLAG	Identifies the site as a Refineries fenceline monitoring site
MONITOR_LATITUDE[1]	Latitude coordinates of the monitoring site
MONITOR_LONGITUDE[1]	Longitude coordinates of the monitoring site
DATUM[1]	Coordinate data system
UTM_NORTHING[1]	UTM projection Y-coordinate
UTM_EASTING[1]	UTM projection X-coordinate
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]	Date in which the site was operational
DATE_SITE_CLOSED[1]	Date in which the site ceased operations
CBSA_NAME	Consolidated Business Statistical Area (CBSA) name
CBSA_TYPE	CBSA type (metropolitan or micropolitan)
URBAN_AREA_NAME	Alternate CBSA 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 (RFG) mandated regulated area
RFG_OPT_IN_AREA_FLAG	Indicates the site is in an RFG opt-in regulated area
RFG_OPT_OUT_AREA_FLAG	Indicates the site is in an RFG opt-out regulated area
WINTER_OXYGENATED_AREA_FLAG	Indicates the site is in a winter oxygenated regulation area
CLOSEST_IEM_STATION	Closest Iowa Environmental Mesonet (IEM) meteorological weather station
CLOSEST_IEM_STATION_WBAN	Closest IEM station identifier
CLOSEST_IEM_STATION_DISTANCE_MILES	Distance in miles between the monitoring site and the closest IEM station
CLOSEST_IEM_STATION_BEARING_FROM_NORTH	Bearing angle from the north of the monitoring site and the closest IEM station
SECOND_CLOSEST_IEM_STATION	Second closest IEM station
SECOND_CLOSEST_IEM_STATION_WBAN	Second closest IEM station identifier

Data Field	Data Description
SECOND_CLOSEST_IEM_STATION_DISTANCE_MILES	Distance in miles between the monitoring site and the second closest IEM station
SECOND_CLOSEST_IEM_STATION_BEARING_FROM_NORTH	Bearing angle from the north of the monitoring site and the second closest IEM station
THIRD_CLOSEST_IEM_STATION	Third closest IEM station
THIRD_CLOSEST_IEM_STATION_WBAN	Third closest IEM station identifier
THIRD_CLOSEST_IEM_STATION_DISTANCE_MILES	Distance in miles between the monitoring site and the third closest IEM station
THIRD_CLOSEST_IEM_STATION_BEARING_FROM_NORTH	Bearing angle from the north of the monitoring site and the third closest IEM 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

Several useful metadata are provided related to site location, monitoring programs, demographic/population activities, and regulatory applicability. A total of 6,064 records are in this data dictionary.

5.2 Monitor Information

Table 5-3 presents data fields for the monitors 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 "*"). 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 which EPA programs, such as NATTS, UATMP, Photochemical Assessment Monitoring Sites (PAMS), and the Interagency Monitoring of PROtected Visual Environments (IMPROVE) network. A total of 553,630 records are in this data dictionary.

Table 5-3. Monitor Information Data Fields

Data Field	Data Description
AMA_SITE_CODE*	Site identifier comprised of STATE_FIPS, COUNTY_FIPS, and LOCAL_SITE_ID
AQS_POC*[1]	Parameter Occurrence Code (POC)
AQS_PARAMETER_CODE*[1]	AQS pollutant identifier
SAMPLE_YEAR*	Year of sampling
MIN_DATE	Start date of measurements for SAMPLE_YEAR
MAX_DATE	End date of measurements for SAMPLE_YEAR
MONITOR_CODE	Site identifier comprised of AMA_SITE_CODE, AQS_POC, and AQS_PARAMETER_CODE

Data Field	Data Description
PROGRAM[1]	Program associated with each monitor, if available
MONITOR_OBJECTIVE	Sampling objective of the monitor, primarily populated in AQS or by ERG if not in AQS
MONITOR_TYPE[1]	Type of monitor, which is primarily populated in AQS or by ERG if not 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 (PQAO)
COUNT_RECORD	Number of AMA HAP records
COUNT_CONCENTRATION	Number of AMA HAP concentrations
ERG_COMMENT	Comment field
SAMPLING_FREQUENCY_DESCRIPTION	Description of the sampling 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
AQS_UNIT_CODE	Unit of measure identifier
PROGRAM_RANK	Ranking of 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 ([MQOs]; 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 MQOs, 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. Program Ranking presents the ranking for each PROGRAM type.

5.3 Pollutant Information

Table 5-4 presents data fields for a comprehensive list of HAP parameter codes listed in the AMA_POLLUTANT_CODE_DICTIONARY. AQS_PARAMETER_CODE is the only primary key field in this data dictionary (denoted by "*"). This data dictionary table includes physical information and alternative pollutant identifiers. There is a total of 385 records in this data dictionary.

Table 5-4. Pollutant Information Data Fields

Data Field	Data Description
REPORTED	Flag to identify if the 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 Chemical Abstract Service (CAS) number, if available
NEI_POLLUTANT_ID	National Emissions Inventory (NEI) 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 MQO 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 (CAP); only lead is flagged
GHG_FLAG	Indicates pollutant is a Greenhouse Gas (GHG) air pollutant
TO15_FLAG	Indicates pollutant can be measured using the TO-15/TO-15A 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]TO-15 pollutants are listed at: https://www.epa.gov/sites/default/files/2019-11/documents/to-15r.pdf
TO15A pollutants are listed at: https://www.epa.gov/sites/default/files/2019-12/documents/to-15a_vocs.pdf
[4]TO-11A pollutants are listed at: https://www.epa.gov/sites/production/files/2019-11/documents/to-11ar.pdf
[5]IO-3.5 pollutants are listed at: https://www.epa.gov/sites/production/files/2019-11/documents/mthd-3-5.pdf
[6]TO-13A pollutants are listed at: https://www.epa.gov/sites/production/files/2019-11/documents/to-13arr.pdf
[7]8270C pollutants are listed at: https://www.epa.gov/sites/default/files/2020-10/documents/method_8270e_update_vi_06-2018_0.pdf
[8]SNMOC pollutants are listed at: https://19january2021snapshot.epa.gov/sites/static/files/2019-11/documents/pams_technical_assistance_document_revision_2_april_2019.pdf
[9]ASTM D7614 pollutants are listed at: https://www.astm.org/Standards/D7614.htm
[10]PAMS pollutants are listed at: https://www.epa.gov/amtic/photochemical-assessment-monitoring-stations-pams

5.4 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 "*"). This data dictionary table includes the federal MDL in its original units and units converted to $\mu g/m^3$

(either in STD or LC in relation to the original units). A total of 4,866 records are in this data dictionary.

Table 5-5. Sampling Methodology Information Data Fields

Data Field	Data Description
AQS_PARAMETER_CODE*[1]	AQS pollutant 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 MDL
AQS_FEDERAL_MDL_UNIT	Federal default MDL units
FEDERAL_MDL_VALUE_STD	Federal default MDL standardized to $\mu g/m^3$
COMMENT	General comment

*primary key field
[1]Data field in the "Sampling Methods for All Parameters" table (https://aqs.epa.gov/aqsweb/documents/codetables/methods_all.html).

5.5 Date and Season Information

Table 5-6 presents data fields for every day from 1990 to 2023 listed in AMA_DATE_DICTIONARY. The primary key is DATE (denoted by "*"). This data dictionary table includes the corresponding day of the week, day type (i.e., weekday or weekend), and calendar quarter in which the month belongs (e.g., Quarter 1 = January, February, and March; Quarter 2 = April, May, and June, etc.). A total of 12,418 records are in this data dictionary.

Table 5-6. Date and Season Information Data Fields

Data Field	Data Description
DATE*	Date of the sample (YYYY-MM-DD)
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 (WD) or Weekend (WE)
YEAR	Calendar year (YYYY)
MONTH	Month (MM)
DAY	Day (DD)
DATE_FORMATTED	Date of the sample (YYYYMMDD)
DAY_NUMBER	Numeric day count within calendar year (ranges 1 – 366)
QUARTER	Identifies the quarter within the calendar year

*primary key field

5.6 Qualifier Code Information

Table 5-7 presents the data qualifier codes in the AMA_QUALIFIER_CODE_DICTIONARY data table. The primary key is AQS_QUALIFIER_CODE (denoted by "*"). This data dictionary table includes information related to QA issues, sampling problems, or information related to the concentration records. While most of the qualifier codes are from AQS, additional qualifier codes were included from non-AQS sources. For example, if the populated Collection Frequency Code in AQS is incorrect, ERG developed a qualifier code describing this error. A total of 179 records are in this data dictionary.

Table 5-7. 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 5-8 presents data fields for the sample duration codes in the AMA_SAMPLE_DURATION_CODE_DICTIONARY. The primary key is AQS_DURATION_CODE (denoted by "*"). This data dictionary table includes information related to the length of time for the sample measurements. A total of 27 records are in this data dictionary.

Table 5-8. 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 5-9 presents the unit codes in the AMA_UNIT_CODE_DICTIONARY. The primary key is AQS_UNIT_CODE (denoted by "*"). A total of 20 records are in this data dictionary.

Table 5-9. 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 "Units" data table (https://aqs.epa.gov/aqsweb/documents/codetables/units.html)

5.9 Collection Frequency Code Information

Table 5-10 presents data fields for the sampling collection frequency codes in the AMA_COLLECTION_FREQUENCY_CODE_DICTIONARY. The primary key is AQS_COLLECTION_FREQUENCY_CODE (denoted by "*"). A total of 30 records are in this data dictionary.

Table 5-10. Frequency Code Data Fields

Data Field	Data Description
AQS_COLLECTION_FREQUENCY_CODE*[1]	Collection frequency code identifier
COLLECTION_FREQUENCY_DESCRIPTION[1]	Collection frequency code description
DAILY_SAMPLE_NUMBER	Number of sub-daily measurements (PAMS only)
DAILY_INTERVAL	Numeric equivalent of the collection frequency code

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

5.10 Data Source Code Information

Table 5-11 presents data fields for the data source codes in the AMA_DATA_SOURCE_CODE_DICTIONARY. The primary key is DATA_SOURCE (denoted by "*"). A total of 154 records are in this data dictionary.

Table 5-11. Data Source Code Data Fields

Data Field	Data Description
DATA_SOURCE_CODE*	Data source code identifier
DATA_SOURCE	Data source abbreviation
DATA_SOURCE_PULLDATE	Date on which the data were pulled
DATA_SOURCE_DESCRIPTION	Data source code description
DATA_SOURCE_GROUP	Data source name
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) from the data source
NUM_SITES	Number of monitoring sites from the data source
NUM_STATES	Number of states from the data source
NUM_COUNTIES	Number of counties from the data source

*primary key field

6.0 Final Database

Approximately 25% of the raw data concentration records were NDs, while 11% were null. Another 16% of the reported HAP concentration records were BMDL. Table 6-1 provides a summary of these counts by year.

Table 6-1. HAP Summary Counts by Year

		ND Records		Null Data Records		BMDL Records	
Year	**# HAP Records**	**#**	**%**	**#**	**%**	**#**	**%**
1990	166,430	63,512	38.2%	6,561	3.9%	17,999	10.8%
1991	207,059	80,007	38.6%	6,446	3.1%	21,523	10.4%
1992	247,144	90,409	36.6%	11,955	4.8%	27,342	11.1%
1993	325,731	107,439	33.0%	20,749	6.4%	35,848	11.0%
1994	530,809	148,209	27.9%	33,130	6.2%	43,058	8.1%
1995	925,903	226,793	24.5%	92,410	10.0%	43,305	4.7%
1996	1,208,911	272,836	22.6%	164,348	13.6%	52,017	4.3%
1997	1,399,184	280,263	20.0%	172,151	12.3%	57,844	4.1%
1998	1,627,758	304,773	18.7%	236,916	14.6%	68,475	4.2%
1999	1,787,329	339,859	19.0%	328,407	18.4%	75,275	4.2%
2000	1,927,823	413,795	21.5%	287,398	14.9%	109,246	5.7%
2001	2,273,075	481,045	21.2%	371,803	16.4%	151,973	6.7%
2002	2,361,217	533,503	22.6%	368,793	15.6%	184,681	7.8%
2003	2,589,366	545,361	21.1%	415,100	16.0%	183,480	7.1%
2004	3,080,988	642,949	20.9%	502,737	16.3%	211,395	6.9%
2005	3,555,018	725,070	20.4%	602,172	16.9%	277,638	7.8%
2006	3,585,258	749,368	20.9%	562,481	15.7%	270,911	7.6%
2007	3,744,054	765,891	20.5%	504,186	13.5%	271,194	7.2%
2008	3,745,371	784,902	21.0%	579,086	15.5%	241,817	6.5%
2009	3,979,890	860,122	21.6%	532,191	13.4%	327,496	8.2%
2010	4,145,754	913,034	22.0%	596,196	14.4%	369,887	8.9%
2011	4,325,168	985,852	22.8%	641,174	14.8%	407,051	9.4%
2012	4,592,801	962,639	21.0%	647,092	14.1%	461,008	10.0%
2013	4,952,051	1,088,893	22.0%	731,647	14.8%	499,755	10.1%
2014	5,453,079	1,164,774	21.4%	745,480	13.7%	526,507	9.7%
2015	5,196,053	1,076,262	20.7%	709,209	13.6%	522,842	10.1%
2016	5,451,964	1,159,377	21.3%	738,053	13.5%	580,172	10.6%
2017	4,242,360	990,147	23.3%	262,158	6.2%	477,561	11.3%
2018	4,140,242	922,363	22.3%	307,657	7.4%	473,942	11.4%
2019	4,444,657	943,102	21.2%	414,725	9.3%	516,089	11.6%
2020	5,592,710	1,512,869	27.1%	774,009	13.8%	1,360,208	24.3%
2021	12,460,194	3,945,405	31.7%	938,884	7.5%	3,765,450	30.2%
2022	12,707,245	4,001,059	31.5%	870,884	6.9%	3,875,059	30.5%
2023	13,388,758	4,395,594	32.8%	605,434	4.5%	4,355,301	32.5%
Total	**130,361,354**	**32,477,476**	**24.9%**	**14,781,622**	**11.3%**	**20,863,349**	**16.0%**

Of the 32,477,476 NDs in the master database, approximately 2% (715,086 records) were suspected as being NDs in which a concentration equal to one-half MDL were either intentionally or mistakenly substituted. Table 6-2 provides an overview of these records by state

and counts of the MDLs that were provided by the data owner versus using the default federal MDL.

Table 6-2. Non-Detect Records Populated with One-Half MDL by State

State	# of ND	# Half-MDL Surrogates	# Fed. MDL Surrogates	# Entity-Provided MDL Surrogates	Time Period of Surrogates
Alabama	86,325	24	24	0	1992-2012
Alaska	77,253	0	0	0	-
Arizona	240,972	6	0	6	2007-2023
Arkansas	32,101	0	0	0	-
California	11,517,489	503,877	269,828	234,049	1990-2023
Colorado	275,863	60,682	58	60,624	1996-2023
Connecticut	283,167	80	13	67	1992-2001
Delaware	88,962	240	35	205	2000-2016
District of Columbia	188,749	155	101	54	1992-2016
Florida	214,834	15,965	124	15,841	1990-2023
Georgia	626,320	2	0	2	2008-2014
Hawaii	43,057	0	0	0	-
Idaho	61,419	10,690	0	10,690	2002-2008
Illinois	582,518	487	484	3	1991-2020
Indiana	520,349	256	199	57	1990-2019
Iowa	86,803	1	0	1	-
Kansas	138,910	22	22	0	1990-1992
Kentucky	194,263	4	0	4	2013-2014
Louisiana	534,935	205	204	1	1994-2023
Maine	897,417	4	4	0	1991-2022
Maryland	242,366	644	555	89	1990-2017
Massachusetts	429,076	344	14	330	1993-2021
Michigan	527,719	139	130	9	1991-2021
Minnesota	520,063	46	43	3	1995-2015
Mississippi	93,782	7	0	7	2006-2008
Missouri	215,025	8	4	4	2006-2014
Montana	132,095	156	156	0	1990-1996
Nebraska	27,863	0	0	0	-
Nevada	75,768	0	0	0	-
New Hampshire	528,967	114	114	0	2002-2007
New Jersey	383,398	31	29	2	1990-2011
New Mexico	83,785	1	0	1	-
New York	415,991	15,101	15,096	5	1990-2022
North Carolina	223,186	1,243	1,243	0	2002-2011
North Dakota	47,099	6	2	4	2000-2014
Ohio	336,870	11	0	11	2002-2015
Oklahoma	145,437	3	0	3	2009-2016
Oregon	252,016	40,358	1,472	38,886	1999-2017
Pennsylvania	789,310	1,410	1,135	275	1993-2016
Rhode Island	239,220	795	3	792	1999-2017
South Carolina	237,086	22	22	0	1993-1994
South Dakota	71,656	0	0	0	-
Tennessee	70,070	190	190	0	1990-1998
Texas	8,724,647	56,521	56,521	0	1990-2010
Utah	162,309	10	0	10	2015-2022
Vermont	145,959	182	13	169	1995-2016
Virginia	159,179	321	136	185	1995-2012
Washington	175,498	4,638	8	4,630	1995-2013

State	# of ND	# Half-MDL Surrogates	# Fed. MDL Surrogates	# Entity-Provided MDL Surrogates	Time Period of Surrogates
West Virginia	39,469	72	5	67	1997-2016
Wisconsin	166,282	4	3	1	1996-2015
Wyoming	99,144	0	0	0	-
Puerto Rico	14,848	8	8	0	-
Virgin Islands	10,587	1	0	1	-
Total	**32,477,476**	**715,086**	**347,998**	**367,088**	**1990-2023**

In the 2023 Archive, data have been stored with native sample durations, as presented in Table 6-3. Approximately 53% of the records have a sample duration of 1-hour and 20% have a sample duration of 24 hours.

Table 6-3. The 2023 Archive Sample Duration Counts by Year

Year	Sub-Hourly	1-Hr	2-Hr	3-Hr	4-Hr	5-Hr	6-Hr	8-Hr	12-Hr	15-Hr	Daily	>Daily
1990	0	24,304	0	756	0	0	0	0	400	0	140,890	80
1991	12	31,383	0	493	0	0	0	0	0	0	175,161	10
1992	51	39,141	0	1,302	0	0	0	0	0	0	206,650	0
1993	137	77,815	0	21,401	0	0	872	0	0	0	225,506	0
1994	162	196,900	0	59,000	0	0	0	0	0	0	274,747	0
1995	9,355	518,570	0	84,192	2,088	0	133	0	0	0	311,565	0
1996	36,894	702,347	0	120,852	6,876	0	0	0	0	0	341,347	595
1997	35,781	874,096	0	120,476	3,843	0	0	0	0	0	363,906	1,082
1998	36,295	1,068,042	0	154,287	2,799	0	0	0	0	0	364,863	1,472
1999	37,648	1,166,184	0	154,112	0	0	0	2,130	0	0	425,536	1,719
2000	37,216	1,240,544	0	137,269	1,797	0	0	1,578	0	0	507,463	1,956
2001	36,925	1,379,758	0	135,038	5,879	0	0	0	6,092	0	706,726	2,657
2002	37,734	1,306,352	0	134,088	10,664	0	0	0	4,290	0	864,934	3,155
2003	37,637	1,477,927	0	116,193	9,641	0	0	0	2,262	0	942,058	3,648
2004	127,065	1,765,569	0	100,965	17,659	0	0	2,313	1,108	0	1,062,233	4,076
2005	147,210	2,087,503	0	104,265	14,526	0	0	10,475	0	0	1,186,758	4,281
2006	149,823	2,226,262	0	113,262	5,073	0	0	3,324	0	0	1,082,820	4,694
2007	361,034	2,214,846	0	125,786	0	0	2,020	0	0	0	1,035,286	5,082
2008	420,952	2,168,563	6,756	111,063	18	9	2,015	0	1,975	0	1,028,364	5,656
2009	416,535	2,268,448	94,413	113,950	1,029	306	0	0	1,089	0	1,078,380	5,740
2010	451,699	2,419,624	97,713	118,116	1,050	234	0	0	1,134	0	1,050,550	5,634
2011	350,886	2,767,553	93,841	110,817	777	201	26	0	0	0	995,641	5,426
2012	305,846	3,104,629	53,649	105,192	933	30	218	0	0	0	1,016,774	5,530
2013	232,141	3,481,703	59,472	99,868	324	48	0	0	0	0	1,068,758	9,737
2014	468,839	3,734,338	74,148	100,799	324	45	0	0	0	9,823	1,046,083	18,680
2015	392,326	3,627,670	80,022	80,213	9	3	0	0	0	0	1,008,746	7,064
2016	493,632	3,875,980	44,256	41,463	0	0	0	456	352	0	990,591	5,234
2017	420,011	2,756,097	46,494	33,378	0	0	0	666	348	0	979,286	6,080
2018	461,620	2,587,411	23,568	32,643	0	0	0	1,981	0	0	971,608	61,411
2019	625,910	2,800,366	28,233	30,512	21	28	42	4,152	154	14	889,615	65,610
2020	2,932,946	1,685,474	13,392	30,012	0	14	41	4,800	126	154	861,111	64,640
2021	7,157,136	4,289,162	26,790	29,849	0	0	0	7,976	14	21	884,333	64,913
2022	7,170,871	4,596,994	19,194	33,542	559	0	0	9,325	0	0	810,674	66,086
2023	7,732,338	4,693,729	46,494	34,598	0	0	0	9,093	0	0	813,114	59,392
Total	**31,124,667**	**69,255,284**	**808,435**	**2,789,752**	**85,889**	**918**	**5,367**	**58,269**	**19,344**	**10,012**	**25,712,077**	**491,340**
% Total	**23.9%**	**53.1%**	**0.6%**	**2.1%**	**0.1%**	**<0.1%**	**<0.1%**	**<0.1%**	**<0.1%**	**<0.1%**	**19.7%**	**0.4%**

[1]>Daily = Weekly, Bi-weekly, or Monthly

7.0 Final Output Data Files

The raw ambient monitoring data are housed in the 2023 Archive data table. For the public release files, the key data fields in the raw table are presented in Table 7-1. Primary key fields are denoted by "*."

Table 7-1. Ambient Monitoring Archive Output Fields

Data Field	Data Description
STATE_ABBR	Two-letter abbreviation for the state of the monitoring site
AMA_SITE_CODE*	Site identifier comprised of STATE_FIPS, COUNTY_FIPS, and LOCAL_SITE_ID
AQS_POC*	Parameter Occurrence Code
PROGRAM	Program associated with each monitor, 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*	AQS pollutant identifier
AQS_PARAMETER_NAME	Pollutant or parameter name
DATA_SOURCE	Identifies the source of the data record
DATA_SOURCE_PULLDATE	Identifies the date which the data were retrieved
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	Sampling frequency code (1=Daily; 2=Every Other Day; 3=Every 3 Days; 4=Every 4 Days; 5=Every 5 Days; 6=Every 6 Days; 7=Every 12 Days; 8=Stratified Random; 9=Random; 10=Every 24 Days; 11=Every 30 Days; 12=Every 7 Days; 14=Every 14 Days; 18=Every 18 Days; 90=Every 90 Days; A, B, or E=PAMS Daily; H, I, J, or L=PAMS 3 Days; O=Every 10 Days; P=PAMS 6 Days; Q=Every 8 Days; R=Every 13 Days; S=Seasonal; Y=Twice Per Week; Z=Every 9 Days)
COMMENT	Reserved for comments
SAMPLE_VALUE_STD	Concentration value standardized to $\mu g/m^3$, STD
SAMPLE_VALUE_STD_FINAL_UG_M3	Concentration value standardized to $\mu g/m^3$, LC
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 pollutant or parameter name for analysis
ALTERNATE_MDL	Reported MDL in native units
MDL_STD_UG_M3	MDL standardized to $\mu g/m^3$
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

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 non-detect
BELOW_MDL_FLAG	Identifies if the non-zero sample value is less than the MDL
CENSUS_BLOCK_ID_2010	U.S. Census (2010) block identifier in which the monitoring site is located
CENSUS_BLOCK_ID_2020	U.S. Census (2020) block identifier in which the monitoring site is located
MONITOR_LATITUDE	Latitude value in decimal degrees
MONITOR_LONGITUDE	Longitude 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 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, Archive records which have deposition units, such as nanogram per liter, are not outputted in the public release files. Archive records prior to 1990 are not output. Archive records in which there is no latitude or longitude coordinate pair are not in the public release files. Lastly, EPA is not outputting the Refineries dataset in the public release files; this dataset is not included in Table 7-2 and Table 7-3 below.

Table 7-2 presents a summary of the final counts in the output files by state. Approximately 96% of the output records are in LC. LC records are initially identified as

- Concentration records in which the reported unit codes are LCs, 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. Additionally, if the

measurement record is not hourly or daily, then the hourly meteorological data were averaged for the same duration of hours. Further, local onsite meteorological data had higher priority than IEM stations. If the closest IEM station data did not have complete information, then the second and third closest data were used if the distances between the monitoring site and meteorological station were within 50 miles. The following hierarchy is used for selecting temperature and pressure data:

- Average (daily) ambient temperature (AQS parameter code = 68105) and average (daily) ambient pressure from AQS (AQS parameter code = 68108).
- The hourly outdoor 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 Iowa Environmental Mesonet (IEM) or AQS stations were used as a surrogate.

The calculation to convert from STD to LC is:

$$concentration, LC = \frac{(concentration, STD) * (298K) * (local\ press.\ in\ mm\ Hg)}{(local\ temp.\ in\ K) * (760\ mm\ Hg)}$$

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

State	# Output Records	# LC[1] Records	# STD[2] Records	% LC Records
Alabama	333,516	321,913	11,603	96.52%
Alaska	1,008,815	730,016	278,799	72.36%
Arizona	804,357	799,888	4,469	99.44%
Arkansas	86,655	86,081	574	99.34%
California	28,774,094	27,732,773	1,041,321	96.38%
Colorado	2,555,333	2,202,326	353,007	86.19%
Connecticut	1,325,513	1,266,609	58,904	95.56%
Delaware	282,267	257,580	24,687	91.25%
District of Columbia	781,880	777,918	3,962	99.49%
Florida	912,858	896,746	16,112	98.23%
Georgia	2,074,378	2,068,001	6,377	99.69%
Hawaii	963,796	660,491	303,305	68.53%
Idaho	130,160	129,704	456	99.65%
Illinois	1,539,735	1,484,384	55,351	96.41%
Indiana	2,918,312	2,870,029	48,283	98.35%
Iowa	206,031	205,666	365	99.82%
Kansas	247,573	211,423	36,150	85.40%
Kentucky	687,193	686,361	832	99.88%
Louisiana	1,513,976	1,326,600	187,376	87.62%
Maine	2,064,188	1,988,730	75,458	96.34%
Maryland	1,601,361	1,392,097	209,264	86.93%
Massachusetts	2,067,389	1,983,222	84,167	95.93%
Michigan	1,630,397	1,589,209	41,188	97.47%
Minnesota	1,350,745	1,299,779	50,966	96.23%
Mississippi	330,066	328,697	1,369	99.59%
Missouri	958,928	944,839	14,089	98.53%

State	# Output Records	# LC[1] Records	# STD[2] Records	% LC Records
Montana	317,700	306,923	10,777	96.61%
Nebraska	80,703	75,319	5,384	93.33%
Nevada	189,395	188,929	466	99.75%
New Hampshire	1,173,295	1,171,444	1,851	99.84%
New Jersey	1,989,690	1,978,859	10,831	99.46%
New Mexico	316,719	315,780	939	99.70%
New York	2,829,542	2,742,223	87,319	96.91%
North Carolina	703,529	681,794	21,735	96.91%
North Dakota	106,566	106,028	538	99.50%
Ohio	978,419	972,950	5,469	99.44%
Oklahoma	460,649	457,312	3,337	99.28%
Oregon	627,761	625,621	2,140	99.66%
Pennsylvania	1,932,664	1,686,544	246,120	87.27%
Rhode Island	975,953	863,606	112,347	88.49%
South Carolina	602,694	575,136	27,558	95.43%
South Dakota	154,511	154,240	271	99.82%
Tennessee	219,218	207,347	11,871	94.58%
Texas	47,984,692	46,826,955	1,157,737	97.59%
Utah	758,239	755,256	2,983	99.61%
Vermont	852,136	847,076	5,060	99.41%
Virginia	710,761	686,191	24,570	96.54%
Washington	534,359	529,917	4,442	99.17%
West Virginia	170,610	165,163	5,447	96.81%
Wisconsin	5,398,381	5,375,102	23,279	99.57%
Wyoming	198,742	197,794	948	99.52%
Puerto Rico	41,976	41,128	848	97.98%
Virgin Islands	33,132	33,132	0	100.00%
Total	**127,491,552**	**122,808,851**	**4,682,701**	**96.33%**

[1]=Standard Conditions
[2]=Local Conditions

Table 7-3 presents a summary of the final counts in the output files by year. From 2001 to 2023, over 98% of the data records are in LCs.

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

Year	# Output Records	# LC[1] Records	# STD[2] Records	% LC Records
1990	165,127	84,786	80,341	51.35%
1991	205,727	103,883	101,844	50.50%
1992	245,600	124,664	120,936	50.76%
1993	325,035	154,524	170,511	47.54%
1994	528,743	269,020	259,723	50.88%
1995	924,632	462,002	462,630	49.97%
1996	1,206,994	792,733	414,261	65.68%
1997	1,394,376	974,580	419,796	69.89%
1998	1,623,442	1,330,260	293,182	81.94%
1999	1,782,369	1,545,012	237,357	86.68%
2000	1,924,469	1,742,178	182,291	90.53%
2001	2,270,542	2,067,763	202,779	91.07%
2002	2,357,589	2,127,754	229,835	90.25%
2003	2,585,517	2,355,456	230,061	91.10%
2004	3,075,789	2,845,108	230,681	92.50%
2005	3,535,259	3,271,689	263,570	92.54%

Year	# Output Records	# LC[1] Records	# STD[2] Records	% LC Records
2006	3,570,245	3,322,024	248,221	93.05%
2007	3,735,138	3,593,275	141,863	96.20%
2008	3,737,187	3,706,751	30,436	99.19%
2009	3,969,891	3,937,871	32,020	99.19%
2010	4,136,465	4,095,433	41,032	99.01%
2011	4,316,078	4,284,985	31,093	99.28%
2012	4,583,898	4,555,702	28,196	99.38%
2013	4,942,150	4,914,117	28,033	99.43%
2014	5,443,305	5,411,040	32,265	99.41%
2015	5,186,385	5,164,582	21,803	99.58%
2016	5,445,638	5,426,460	19,178	99.65%
2017	4,235,639	4,212,643	22,996	99.46%
2018	4,079,506	4,056,846	22,660	99.44%
2019	4,379,140	4,359,568	19,572	99.55%
2020	5,320,483	5,304,723	15,760	99.70%
2021	11,837,679	11,799,917	37,762	99.68%
2022	11,693,016	11,691,241	1,775	99.98%
2023	12,728,499	12,720,261	8,238	99.94%
Total	**127,491,552**	**122,808,851**	**4,682,701**	**96.33%**

[1]=Standard Conditions
[2]=Local Conditions

Appendix A. Overlapping Records

Appendix A can be found under Supporting Appendices from the Archive webpage.

Appendix B. Invalidated Records

Appendix B can be found under Supporting Appendices from the Archive webpage.

Appendix C. Sampling Frequency Code Corrections

Appendix C can be found under Supporting Appendices from the Archive webpage.

Appendix D. Questionable Values and Incorrectly Submitted One-Half MDL Concentrations

Appendix D can be found under Supporting Appendices from the Archive webpage.

Appendix E. Negative Concentrations and Incorrectly Assigned Qualifier Codes for "MD," "ND," and "SQ"

Appendix E can be found under Supporting Appendices from the Archive webpage.

Appendix F. Program Ranking

Appendix F can be found under Supporting Appendices from the Archive webpage.