



EPA's Hazardous Air Pollutant (HAP) Ambient Monitoring Archive

Regi Oommen, Karla Faught, Jaime Hauser, Tyler Richman, and Steve Mendenhall,
Eastern Research Group, Inc. (ERG)

Jeanette Reyes, Doris Chen, and Neilson Watkins, U.S. Environmental Protection Agency

National Ambient Air Monitoring Conference - New Orleans

August 2024

Criteria vs. Toxics Pollutant Monitoring Data

- ▶ Criteria Air Pollutant (CAP) measurements are often required, routine, use reference and equivalent methods, and have dedicated, consistent quality assurance (QA).
- ▶ Ambient Hazardous Air Pollutants (HAPs), a.k.a. Air Toxics, measurements are all voluntary (from a federal perspective). These data are spatially and temporally variable, can be collected via different methodologies, and do not always have the same or consistently applied QA.
- ▶ Unlike criteria air pollutants (CAPs) which are typically required to report to AQS, HAP data are typically not required to report to AQS (unless stipulated in the funding requirements, such as NATTS).
- ▶ The primary goal of the Archive is capture as much ambient air toxics monitoring data that exists in the public domain and create a comprehensive centralized database.

What is the Archive?

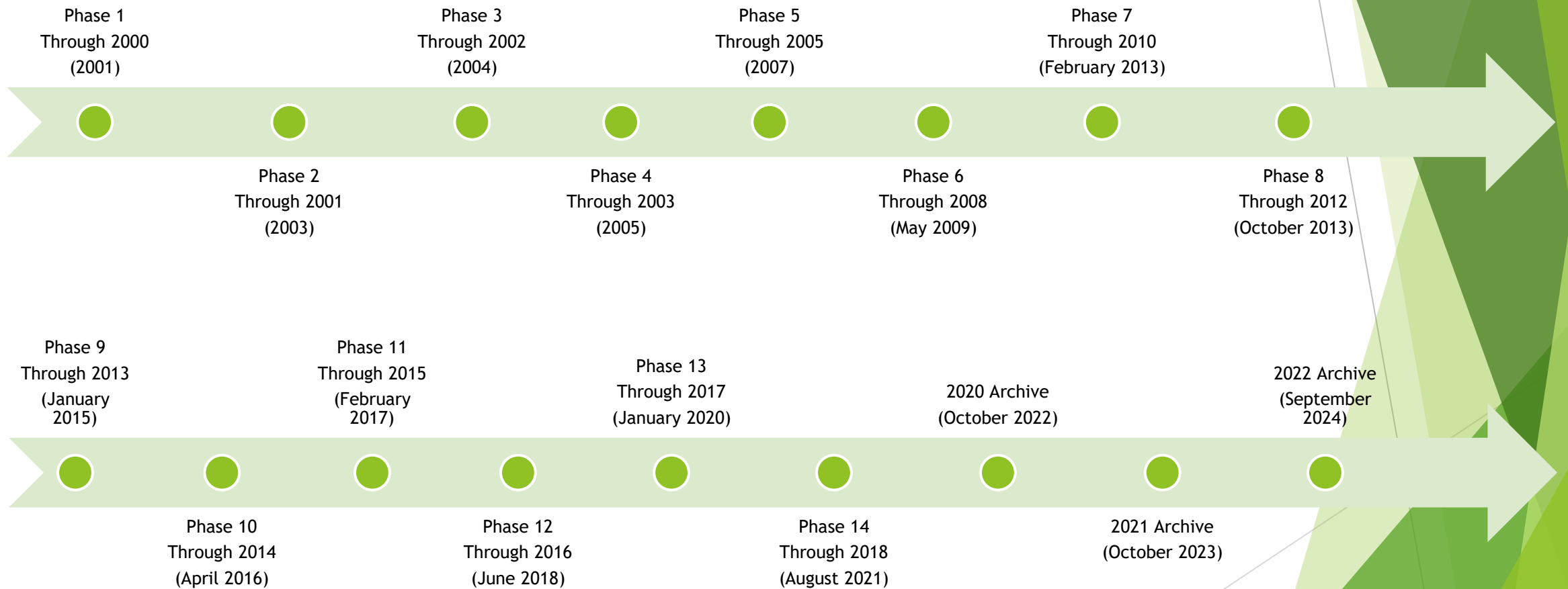
- ▶ SQL Server database maintained by ERG, containing air toxics, criteria pollutants, and meteorological data.
- ▶ Air toxics data includes monitored air pollutants at the finest level of the data (e.g., 5-minute, hour, 24-hour)
- ▶ Air toxics are from AQS and non-AQS data sources
- ▶ Air toxics are QA'd and made analysis-ready
- ▶ Meta data added or modified from AQS (corrections)
- ▶ Air toxics extracted and converted to ACCESS, TXT and R-data files for EPA - these are posted on EPA's website
- ▶ The Archive is a “one-stop-shop” for ambient air toxics monitoring data

Archive Applications - How are/has the Archive data used?

- ▶ NATTS Network Assessment
- ▶ EPA Reports
 - ▶ Report on Environment
 - ▶ The Trends Report
 - ▶ The Second Integrated Report to Congress
- ▶ Model Evaluations (e.g., AirToxScreen, CMAQ, AERMOD)
- ▶ AirToxScreen Mapping Tool
- ▶ EPA Office of Compliance/Civil Enforcement targeting activities
- ▶ OAQPS and ORD peer-reviewed journal publications
- ▶ EJ Considerations
- ▶ Public resource for universities and state/local air agencies



Archive Timeline (2001-2024)



What's the Latest Archive Version?

- ▶ Phase 2021 Archive (1990-2021)
 - ▶ Consolidated one-stop shop ready for data analysis
- ▶ Key features
 - ▶ 101.5 million HAP records
 - ▶ 5,756 monitoring sites
 - ▶ 3,006 non-benzene refinery locations
 - ▶ 2,750 benzene refinery locations
 - ▶ 385 parameter codes
 - ▶ 37 unique sources of data



2021 Archive Data Sources (1 of 3)

| Data Source | Data Years | # Sites | # Parameters | # Records |
|--|------------|---------|--------------|------------|
| Air Quality System (AQS) Database ¹ | 1990-2021 | 2,384 | 367 | 62,659,925 |
| TCEQ TAMIS ² | 1992-2021 | 131 | 83 | 22,116,889 |
| South Coast AQMD ³ | 1999-2021 | 130 | 101 | 9,075,330 |
| National Acid Deposition ⁴ | 1996-2021 | 189 | 4 | 2,430,025 |
| NOAA ⁵ | 1990-2021 | 8 | 8 | 1,703,880 |
| Massachusetts Institute of Technology | 1995-2021 | 1 | 3 | 934,122 |
| Louisiana DEQ | 2010-2021 | 27 | 59 | 505,272 |
| XACT Monitoring Data | 2011-2021 | 10 | 17 | 404,392 |
| Houston Health Department | 2019-2020 | 3 | 1 | 253,434 |
| EPA Refineries | 2016-2021 | 2,750 | 1 | 237,721 |
| National Park Service Studies | 2011-2019 | 75 | 20 | 228,479 |
| Phase V Archive | 1991-2020 | 144 | 164 | 201,862 |
| Colorado Boulder AIR | 2017-2020 | 4 | 7 | 192,186 |

¹ 62.7% of the Archive records; ² 21.8% of the Archive records; ³ 38.9% of the Archive records; ⁴ 2.4% of the Archive records; ⁵ 1.7% of the Archive records

2021 Archive Data Sources (2 of 3)

| Data Source | Data Years | # Sites | # Parameters | # Records |
|--|------------|---------|--------------|-----------|
| Michigan Community-Scale Air Toxics Monitoring | 2016-2017 | 3 | 9 | 168,343 |
| Integrated Atmospheric Deposition Network | 1999-2010 | 11 | 89 | 162,836 |
| Minnesota Air Toxics | 2008-2015 | 44 | 61 | 88,058 |
| Sublette County, WY | 2009-2010 | 14 | 42 | 37,398 |
| Utah State University - Vernal | 2012-2021 | 6 | 16 | 26,435 |
| EPA Passive Sampling | 2013-2015 | 17 | 9 | 18,675 |
| Pennsylvania Marcellus Shale Study | 2012-2013 | 6 | 39 | 14,793 |
| Allegheny County, PA Health Department | 2011-2021 | 4 | 22 | 14,109 |
| NATTS Network Assessment | 2003-2014 | 5 | 71 | 11,608 |
| CARB Pesticides | 2010-2021 | 20 | 4 | 10,899 |
| Missouri Community-Scale Air Toxics Monitoring | 2008-2009 | 7 | 3 | 9,612 |
| Baldwin Hills Air Quality Study | 2012-2013 | 1 | 16 | 7,455 |

2021 Archive Data Sources (3 of 3)

| Data Source | Data Years | # Sites | # Parameters | # Records |
|---|------------|---------|--------------|-------------|
| City of Ft. Worth, TX Natural Gas Air Quality Study | 2010 | 8 | 49 | 5,455 |
| EPA Region 3 | 2008-2020 | 2 | 14 | 3,633 |
| Oregon Department of Environmental Quality | 2012-2017 | 10 | 3 | 3,350 |
| EPA Denka Chloroprene | 2016-2021 | 6 | 1 | 2,517 |
| Wisconsin Department of Natural Resources | 2019-2021 | 4 | 13 | 2,468 |
| New York State DEC | 2014-2015 | 1 | 36 | 2418 |
| CARB Special Studies | 2001-2002 | 1 | 34 | 2098 |
| Ethylene Oxide Special Studies | 2018-2021 | 63 | 1 | 2071 |
| Baltimore Inner Harbor Monitoring Study | 2014-2015 | 6 | 1 | 1734 |
| Colorado DPHE | 2018 | 6 | 47 | 1729 |
| School Air Toxics Ambient Monitoring Program | 2011-2012 | 6 | 80 | 800 |
| Long Island Sound Tropospheric Ozone Study | 2018 | 2 | 34 | 640 |
| Totals | 1990-2021 | 5,756 | 385 | 101,542,651 |

New Data Studies for Phase 2022

- ▶ New York State DEC - South Albany air monitoring
- ▶ Region 5 XAct Monitoring - Ohio
- ▶ EPA Denka Chloroprene Measurements
- ▶ Ethylene Oxide Special Studies in Utah and West Virginia
- ▶ NOAA Surface Observations Program
- ▶ South Coast AQMD sponsored studies: ethylene oxide, Jordan Downs, West Dominguez

Archive Data - What gets in?

- ▶ AQS HAP data from the AMP501 “Extract Raw Data” function is pulled in.
- ▶ For new non-AQS data (e.g., NYSDEC special studies), ERG:
 - ▶ Talks with the Data Owner and/or
 - ▶ Reviews supporting materials (e.g. - MDLs, sampling/analysis methods/coordinates)
- ▶ For recurring non-AQS data (e.g., Allegheny County, PA):
 - ▶ ERG checks in with the Data Owner for new/updated data



Archive Pre-Processing/QA

- ▶ Pollutant name updates
 - ▶ e.g., parameter code = 17141. Renamed from “naphthalene (Tsp) STP” to “naphthalene (total tsp and vapor)”
- ▶ Identify ½ MDLs for non-detects
 - ▶ Over 922,000 concentration records suspected as being ½ MDL. Converted to “0” and flagged accordingly
- ▶ Negative concentrations
 - ▶ Nearly 530,000 concentration records were negative. Converted to “0” and flagged as “NEG”



QA: Data Invalidation

- ▶ Invalidated data
 - ▶ All hexavalent chromium and acrolein data prior to 2005 or ≥ 2005 data with inappropriate methods
 - ▶ All PAH data prior to 2007 or ≥ 2007 data with inappropriate methods
 - ▶ Wholesale datasets restored as “invalidated” for posterity and completeness
 - ▶ “High” concentrations/MDLs reviewed and updated
 - ▶ e.g., blank values entered in AQS rather than concentration

QA: Other Corrections

- ▶ Duplicate data records
- ▶ Alternate MDL values (in AQS) incorrectly populated in the Uncertainty field
- ▶ Revised/updated concentrations
- ▶ Revised/updated native units
- ▶ Populate/QA sampling frequency codes
- ▶ Populate/QA method detection limits
- ▶ Inconsistency of Data Qualifier codes
- ▶ Pollutant overlap (e.g., xylenes)
- ▶ Standardize all concentrations to $\mu\text{g}/\text{m}^3$
- ▶ Convert to local conditions (using temp. and press.)

Database Structure

| Data Table | # Records | # Data Fields | Primary Key(s) |
|---------------------------------------|-------------|---------------|--|
| AMA Input File | 101,542,651 | 44 | Site Code, POC, Sample Date, Start Time, Parameter Code |
| Site Information | 5,756 | 78 | Site Code |
| Monitor Information | 511,556 | 21 | Site Code, POC, Parameter Code, Year |
| Pollutant Information | 386 | 24 | Parameter Code |
| Sampling Method Information | 4,501 | 13 | Parameter Code, Method Code, Unit Code, Sample Duration Code |
| Date and Season Information | 11,688 | 10 | Sample Date |
| Qualifier Code Information | 181 | 4 | Qualifier Code |
| Sample Duration Information | 26 | 5 | Sample Duration Code |
| Unit Code Information | 19 | 4 | Unit Code |
| Collection Frequency Code Information | 30 | 4 | Sampling Frequency Code |
| Data Source Code Information | 110 | 10 | Data Source Code |

Database Tables

| Data Table | Features |
|---------------------------------------|---|
| Site Information | Site locations; Site Name(s); Census tract/block IDs; closest weather station(s); Program designations; CBSA Name |
| Monitor Information | Program affiliation; priority ranking of each dataset |
| Pollutant Information | Physical characteristics and designations (e.g., NATTS) |
| Sampling Method Information | Collection and analysis information; federal MDL |
| Date and Season Information | Quarter and season information |
| Qualifier Code Information | Quality Assurance and Null Data Information |
| Sample Duration Information | Length of sample information |
| Unit Code Information | Unit Description |
| Collection Frequency Code Information | Sampling Frequency information |
| Data Source Code Information | Source of data and date received; year range; count of sites, parameters, and records |

Public Release

- ▶ Archive does not output:
 - ▶ Refineries data
 - ▶ Acrolein - unverified (AQS Parameter code = 43505)
 - ▶ Combined pollutants (e.g. 45110 = styrene and *o*-xylene)
 - ▶ Non-concentration records (e.g., deposition units)
 - ▶ Records with no associated latitude/longitude coordinates
 - ▶ Pre-1990 data

Output by Year(s)

| Year | # Output Records | # Local Condition Records | % Local Conditions |
|--------------|--------------------|---------------------------|--------------------|
| 2021 | 10,367,005 | 10,201,239 | 98.40% |
| 2020 | 4,762,083 | 4,754,150 | 99.83% |
| 2019 | 4,216,648 | 4,123,555 | 97.79% |
| 2018 | 3,970,572 | 3,917,545 | 98.66% |
| 2017 | 4,177,230 | 4,156,003 | 99.49% |
| 2016 | 5,406,628 | 5,396,242 | 99.81% |
| 2015 | 5,156,639 | 5,135,419 | 99.59% |
| 2014 | 5,420,225 | 5,393,294 | 99.50% |
| 2013 | 4,919,293 | 4,899,404 | 99.60% |
| 2012 | 4,561,179 | 4,539,884 | 99.53% |
| 2011 | 4,292,308 | 4,261,797 | 99.29% |
| 2010 | 4,117,815 | 4,081,379 | 99.12% |
| 2009 | 3,955,054 | 3,903,923 | 98.71% |
| 2008 | 3,715,557 | 3,649,678 | 98.23% |
| 2007 | 3,718,783 | 3,590,049 | 96.54% |
| 1990-2006 | 27,495,992 | 23,582,176 | 85.77% |
| Total | 100,253,011 | 95,585,737 | 95.34% |

Data Files Posted

Air Monitoring HAPs Data By State

Files are also available from 1990 to 2021 by state, the District of Columbia, Puerto Rico, and the Virgin Islands in zipped Microsoft Access databases. Please note that file sizes range from 852 KB to 615 MB.

| State | State | State | State |
|---|---------------------------------------|--|---------------------------------------|
| Alabama (9.85 MB) | Alaska (21.9 MB) | Arizona (18.2 MB) | Arkansas (2.24 MB) |
| California (345 MB) | Colorado (36.3 MB) | Connecticut (33.3 MB) | Delaware (7.39MB) |
| District of Columbia (19.4 MB) | Florida (24.0 MB) | Georgia (53.5 MB) | Hawaii (23.1 MB) |

Ambient Monitoring HAPs Data By Year

The 2021 Archive for HAPs data from 1990 to 2021 are presented by year in zipped Microsoft Access databases and .txt files. Please note that file sizes range from 6.53 MB to 465 MB.

| Year | Year | Year | Year |
|----------------------------------|----------------------------------|-----------------------------------|----------------------------------|
| 2021 (465 MB) | 2020 (228 MB) | 2019 (228 MB) | 2018 (215 MB) |
| 2017 (229 MB) | 2016 (286 MB) | 2015 (277 MB) | 2014 (287 MB) |
| 2013 (264 MB) | 2012 (245 MB) | 2011 (224 MB) | 2010 (212 MB) |
| 2009 (205 MB) | 2008 (187 MB) | 2007 (193 MB) | 2006 (183MB) |

<https://www.epa.gov/amtic/amtic-ambient-monitoring-archive-haps>

Additional Data

All 2021 Archive Monitoring Data and Annual Data Summaries

[All .Rda data files by year](#) (738 MB) contains all data in the 2021 Archive presented by year in .Rda files (generated by the R programming language).

[Annual Statistics \(XLSX\)](#) (185 MB) provides annual statistics by pollutant, site, sampling duration, and year and supporting information assisting in trends analysis.

Supporting Files for the 2021 Archive

- [Field Descriptions \(PDF\)](#) (1 page, 68.3 KB) describes the fields in the 2021 Archive files.
- [Data Dictionary \(ZIP\)](#) (11.9 MB) contains ten data dictionaries needed for describing and standardizing the raw data and are recommended for providing additional context to the concentration records. Data dictionaries are initially retrieved from EPA's AQS website with additional data elements added.
- [Supporting appendices \(ZIP\)](#) (86.4 MB) contains six supporting appendices including information regarding overlapping records, invalidated records, sampling frequency code corrections, questionable values, negative concentrations, and program rankings.
- [Lookup Tables \(XLSX\)](#) (241 KB) contains additional cross-referenced information in the annual statistics excel file, field descriptions in the annual statistics excel file, R code used to calculate the annual statistics, and other descriptive information.

Supporting Documentation for the 2021 Archive

- [Technical Report](#) (54 pages, 1.01 MB) provides background information, describes data sources, and documents the improvements, modifications, and additional data incorporated into the development of the 2021 Archive.
- [Annual Statistics Documentation \(PDF\)](#) (10 pages, 437 KB) documents the steps used to calculate the annual statistics along with descriptors of the corresponding R code, the quality assurance methodology and results, and answers to frequent questions.

Number of Sites by HAP¹ (1990-2021)

| Pollutant/Pollutant Group | # Sites | Pollutant/Pollutant Group | # Sites |
|---------------------------|---------|---------------------------|---------|
| Lead Compounds | 1,856 | Tetrachloroethylene | 1,106 |
| Benzene | 1,298 | Selenium Compounds | 1,081 |
| Arsenic Compounds | 1,291 | Cadmium Compounds | 1,072 |
| Chromium Compounds | 1,263 | Trichloroethylene | 1,062 |
| Toluene | 1,259 | Carbon Tetrachloride | 1,023 |
| Ethylbenzene | 1,256 | 1,3-Butadiene | 1,018 |
| Manganese Compounds | 1,246 | Methylene Chloride | 1,017 |
| Nickel Compounds | 1,244 | Chloroform | 980 |
| Xylenes (mixed isomers) | 1,241 | Methyl Chloroform | 951 |
| Styrene | 1,160 | Vinyl Chloride | 931 |

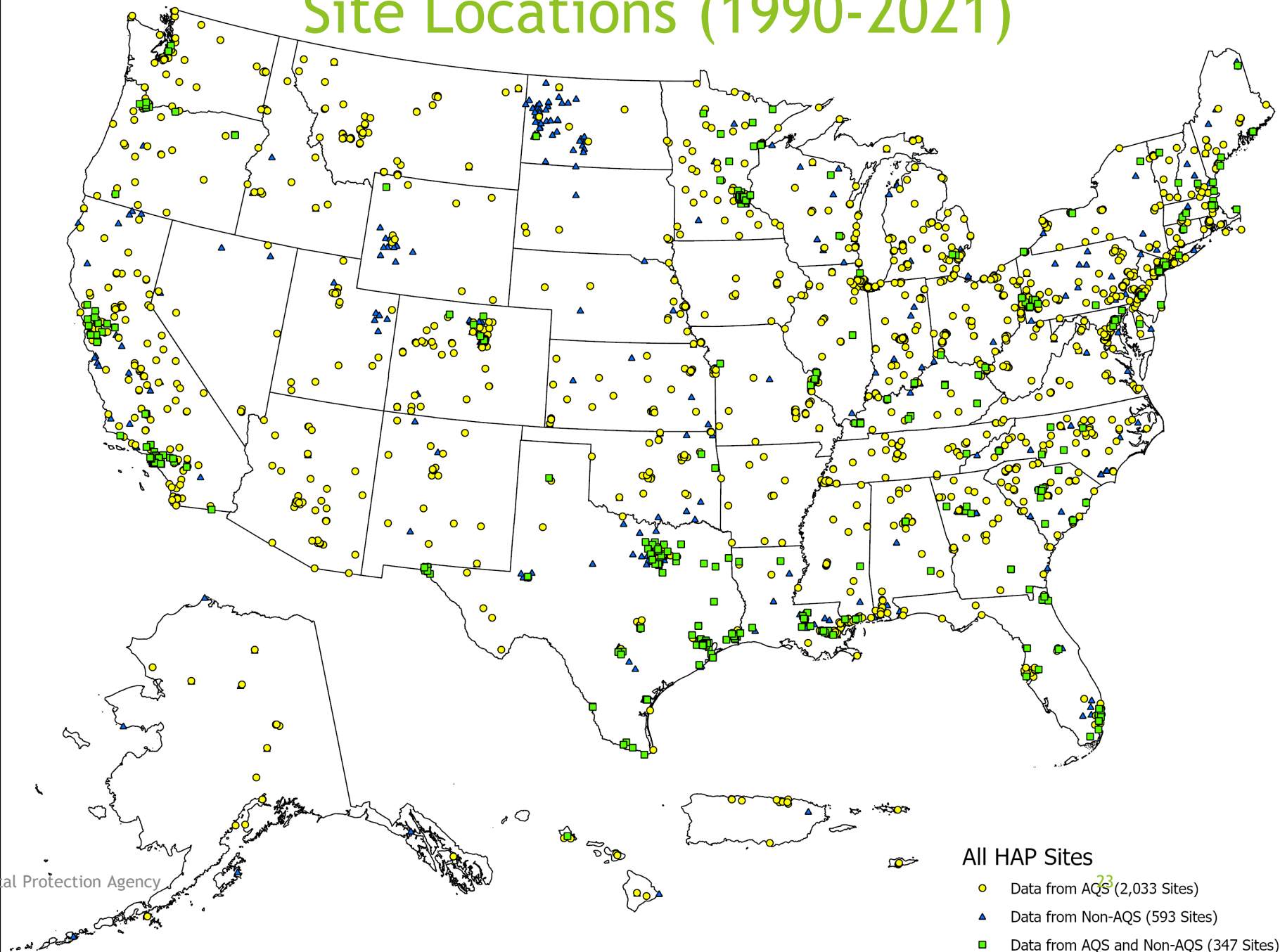
¹ Non-Refinery locations. These 20 HAP category pollutants account for over 60% of the total Archive records

Number of Sites by HAP¹ (2021)

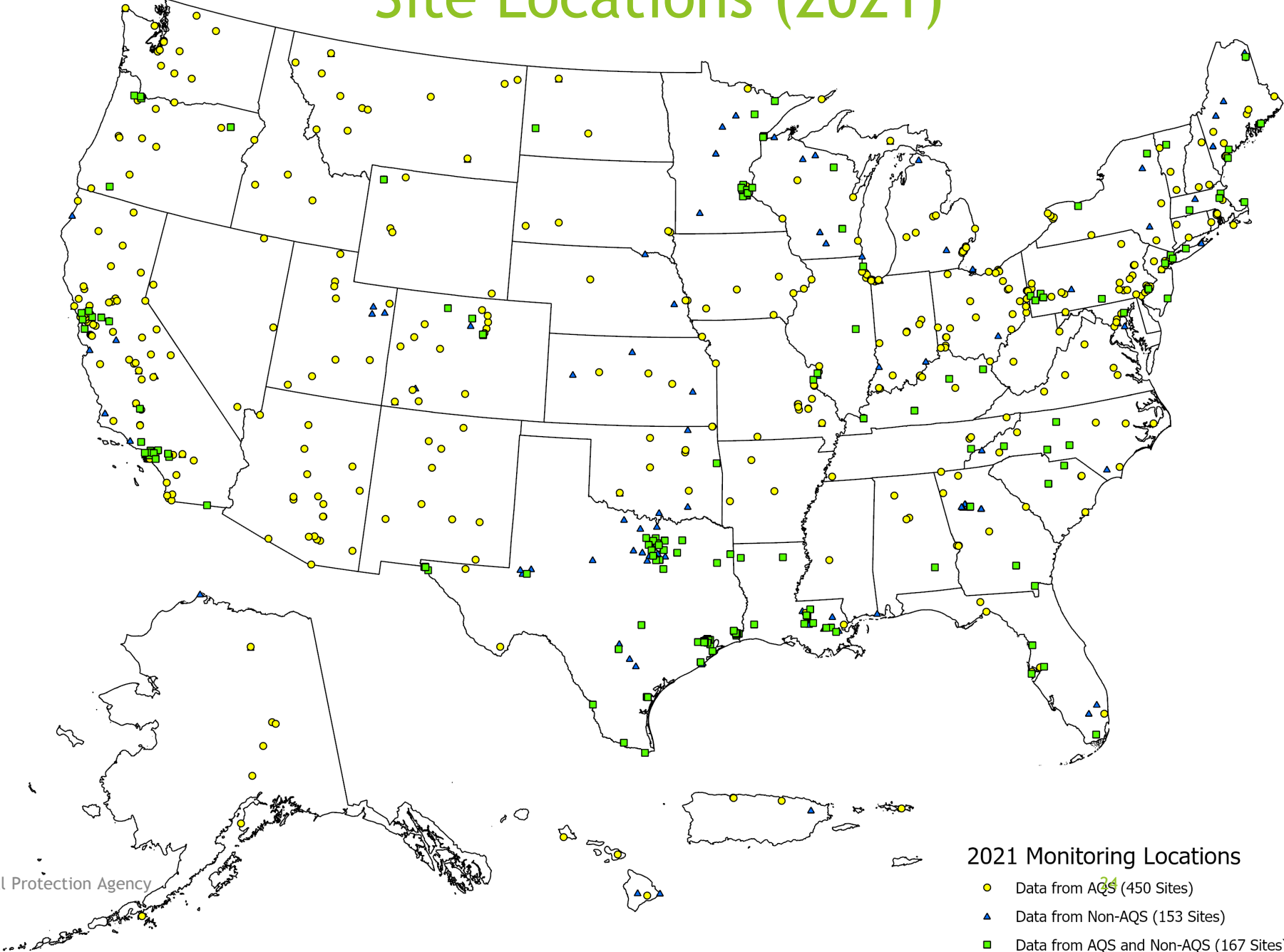
| Pollutant/Pollutant Group | # Sites | Pollutant/Pollutant Group | # Sites |
|---------------------------|---------|---------------------------|---------|
| Lead Compounds | 452 | Toluene | 295 |
| Manganese Compounds | 374 | Xylenes (mixed isomers) | 295 |
| Arsenic Compounds | 368 | 1,3-Butadiene | 287 |
| Nickel Compounds | 365 | Styrene | 268 |
| Chromium Compounds | 357 | Tetrachloroethylene | 242 |
| Selenium Compounds | 330 | Carbon Tetrachloride | 239 |
| Benzene | 313 | Methylene Chloride | 237 |
| Chlorine | 303 | Methyl Chloroform | 235 |
| Phosphorus Compounds | 303 | Chloroform | 235 |
| Ethylbenzene | 300 | Trichloroethylene | 234 |

¹ These 20 HAP category pollutants account for nearly 63% of the 2021 records

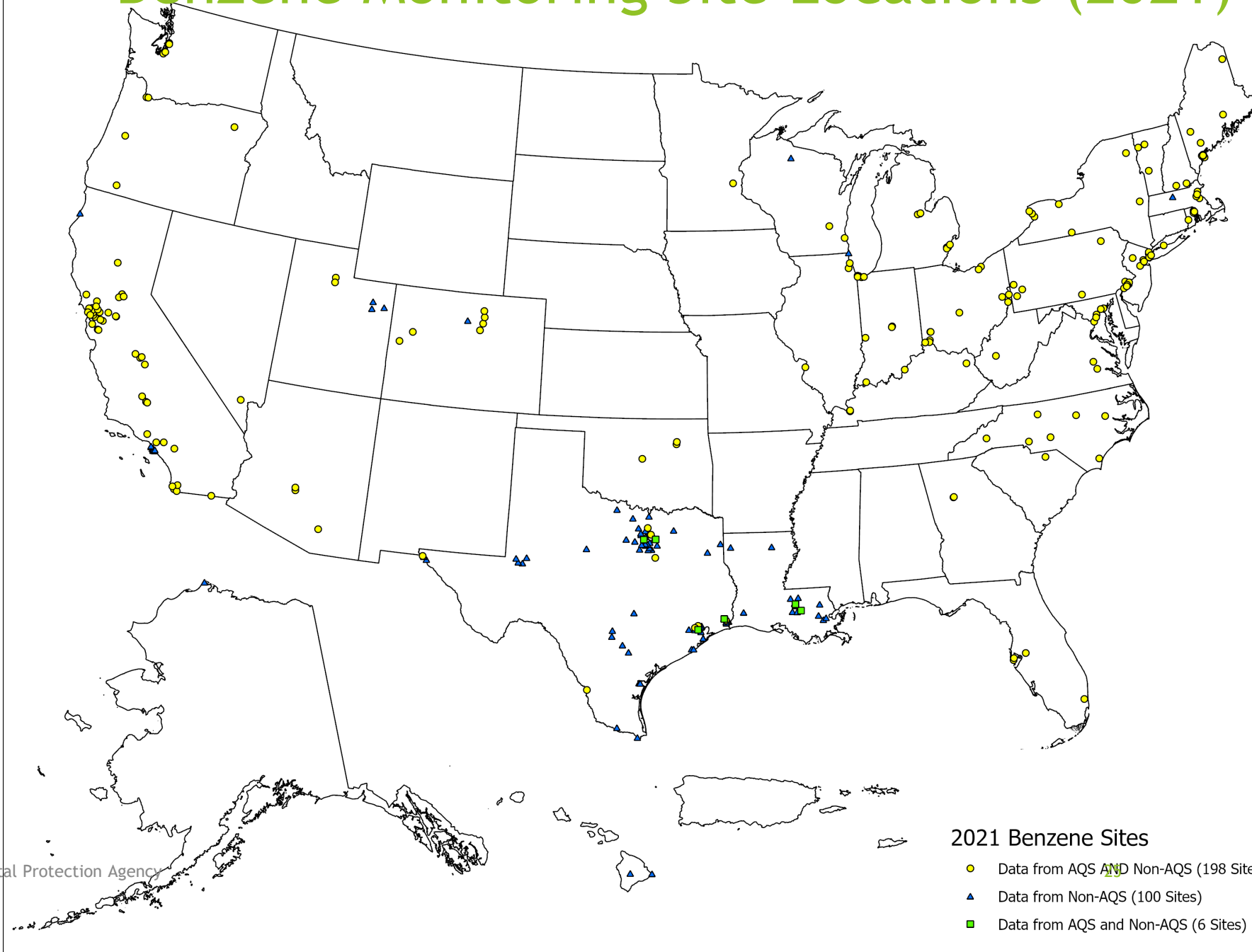
Site Locations (1990-2021)



Site Locations (2021)



Benzene Monitoring Site Locations (2021)



Air Enforcement Targeting Activities



Common Dataset for Report On Environment (ROE)

EPA's Report on the Environment (ROE)



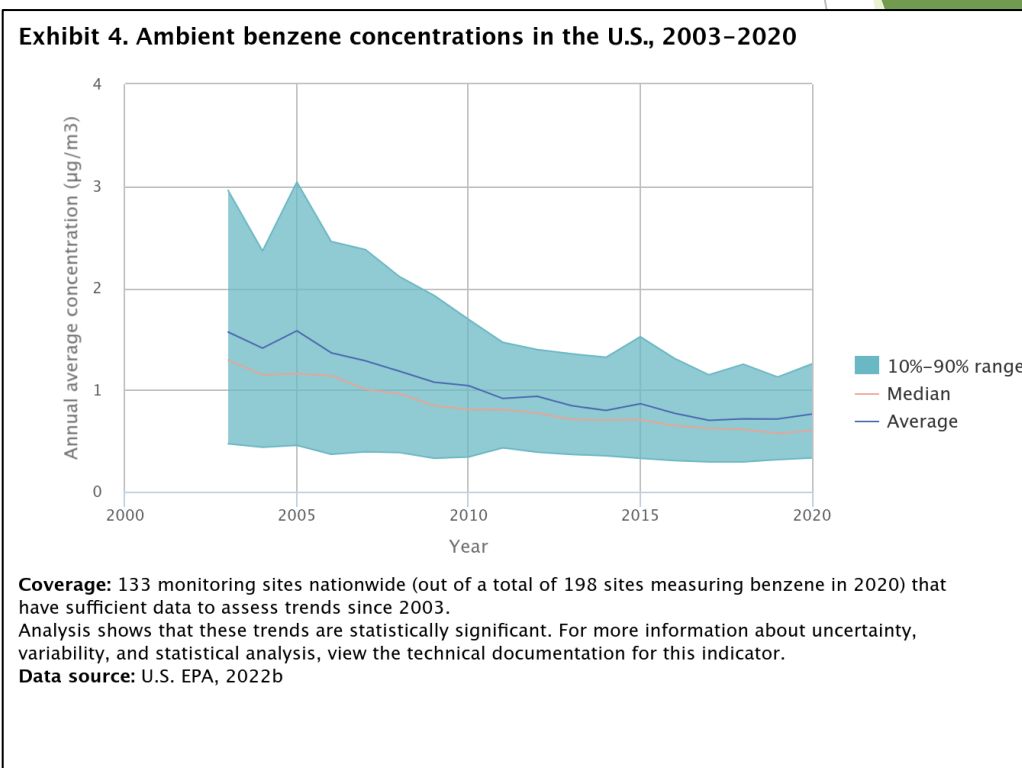
What is an ROE Indicator?

EPA's Report on the Environment (ROE) indicators are simple measures that track the state of the environment and human health over time. [Learn about ROE indicators](#)

1 2 3

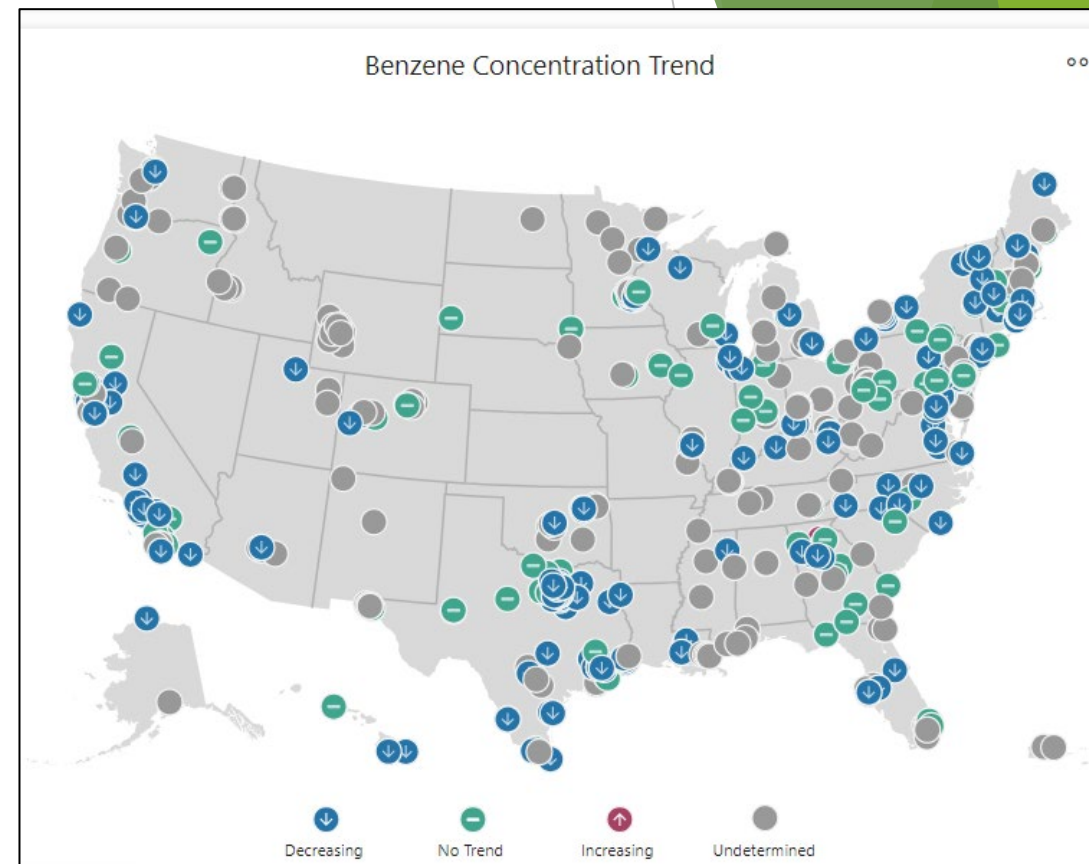
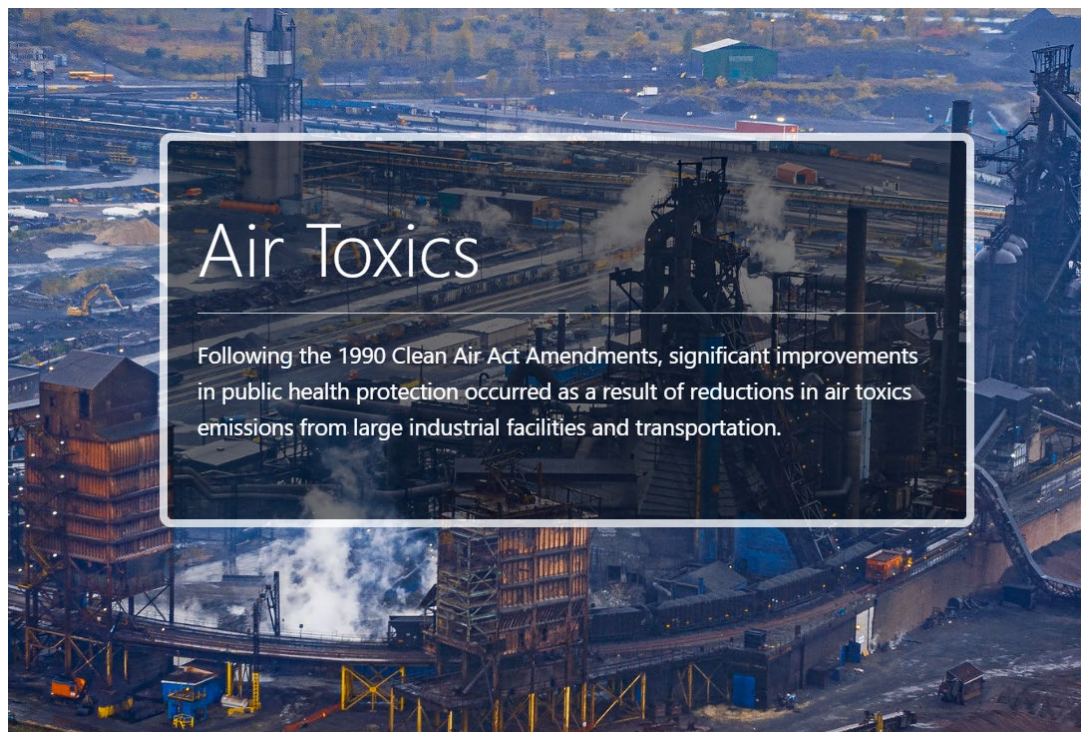
What is the Report on the Environment?

EPA's Report on the Environment (ROE) shows how the condition of the U.S. environment and human health is changing over time. The



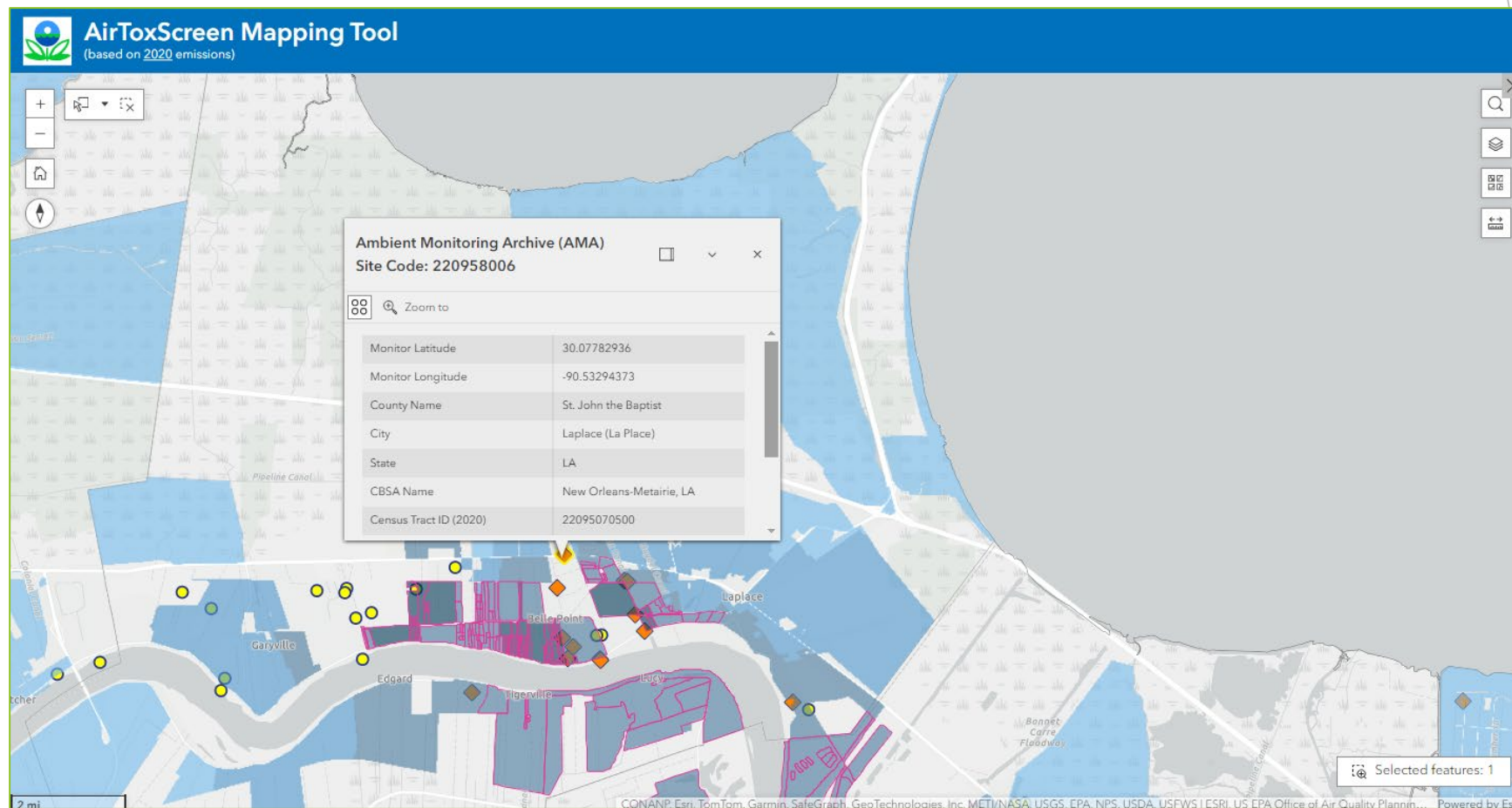
<https://www.epa.gov/report-environment>

Common Dataset for EPA Trends



<https://gispub.epa.gov/air/trendsreport/2023/#toxics>

AirToxScreen Mapping Tool



<https://www.epa.gov/AirToxScreen/airtoxscreen-mapping-tool>

Potential Activities

- ▶ Sector Profiles
 - ▶ Develop concentration profiles based on sectors
- ▶ Risk and Technology Review (RTR)
 - ▶ Compare modeled concentrations to monitoring locations
- ▶ Emission Inventory Validation
 - ▶ Comparison of HAP pollutants observed and what's reported
- ▶ HAP Reduction Strategies
 - ▶ Comparison of HAP concentrations before and after rule implementation

Potential Activities

- ▶ Continue to Include New Data from:
 - ▶ Community-Scale Air Toxics Monitoring
 - ▶ EJ/Grant Funded Projects
 - ▶ Special studies from State/Local/Tribal agencies

▶ **WE WANT YOUR DATA!**



Thank You!

- ▶ Regi Oommen (Regi.Oommen@erg.com)
- ▶ Jeanette Reyes (reyes.jeanette@epa.gov)
- ▶ Doris Chen (chen.xi@epa.gov)
- ▶ Nealson Watkins (watkins.nealson@epa.gov)