Overview of Environmental Data Sources for Environmental Public Health Tracking

Environmental data provides information on hazards. An environmental hazard has the potential for harmful effects, but its presence is not sufficient to produce an adverse effect in a population. Most environmental data collected by federal and state environmental agencies is mandated by legislation, such as the Clean Air Act or Clean Water Act and is collected and used for regulatory purposes. Data analysis is often limited to comparison of each data point to an environmental standard or guideline and action is triggered if the data point exceeds the standard. Standards are not available for all environmental contaminants, and those that do exist are extrapolated from toxicological studies and are developed to be protective of public health, not predictive of health outcomes.

I. LEGISLATION FOR ENVIRONMENTAL DATA COLLECTION

Environmental legislation passed by Congress is the reason we have much of the environmental data that exists today. Such legislation includes the:

- Clean Air Act
- Clean Water Act
- Safe Drinking Water Act
- Resource Conservation & Recovery Act
- Comprehensive Environmental Response, Compensation, and Liability Act
- and the Federal Insecticide, Fungicide, & Rodenticide Act

The Clean Air Act (CAA) was passed by Congress in 1970 and was later amended in 1990 to regulate emissions from area, stationary and mobile sources to protect air quality. One feature of the 1990 amendments is the addition of a permit program which sets limits on the amount of a pollutant sources can release into the air. The permit includes information about which pollutants are being released, how much may be released, and the steps the source's owner or operator is taking to reduce pollution, including plans to monitor the pollution.

The law also established the National Ambient Air Quality Standards (NAAQS) for pollutants considered harmful to public health and the environment. EPA identified six criteria air pollutants [ozone, carbon monoxide, particulate matter, nitrogen dioxide, sulfur dioxide, and lead], and used health-based guidelines as the basis for setting permissible levels for each pollutant. These "criteria pollutants" are surrogates for monitoring more complex mixture of pollutants that exist in ambient air. Additionally, EPA regulates 189 hazardous air pollutants (HAPs) believed to cause cancer, reproductive disorders, and other illnesses.

The Federal Water Pollution Control Act (FWPCA) was passed by Congress in 1972. In 1977 it was amended to become the **Clean Water Act (CWA).** The CWA established pollution control programs for regulating discharges into US waterways through a variety of regulatory and non-regulatory tools. The CWA established a permit program for industries that discharge

pollutants into waterbodies by establishing water quality standards for contaminants in surface water. The standards are intended to protect water quality and a water body's designated use, and protect public health and the environment.

Congress passed the **Safe Drinking Water Act (SDWA)** in 1974 to protect public health by regulating the nation's public drinking water supply. The focus is to protect surface and ground water designated as drinking water sources from natural and man-made contaminants. The goal is to protect the nation's water supply from source to tap. The SDWA is divided into three focus areas: drinking water regulations to protect human health; regulation of underground injection of wastes; and ground water protection programs. EPA enforces national standards, the National Primary Drinking Water Regulations, to protect against health risks. EPA sets a goal based on scientific health risk assessment, and then sets a legal limit for the contaminant in drinking water called the Maximum Contaminant Limit (MCL).

The **Resource Conservation and Recovery Act (RCRA)** was enacted in 1976 and gave EPA the authority to control hazardous waste from "cradle to grave", that is, the generation, transportation, treatment, storage, and disposal of hazardous waste. RCRA has three goals: to protect human health and the environment; reduce waste and conserve energy and natural resources; and reduce or eliminate the generation of hazardous waste. RCRA also has provisions for the management of nonhazardous solid waste and regulation of underground storage tanks. RCRA focuses only on active and future facilities and does not address abandoned or historical sites.

The **Comprehensive Environmental Response, Compensation, and Liability Act (CERCLA)**, more commonly known as Superfund, was enacted in 1980. The law created a tax on the chemical and petroleum industries that went to a trust fund for the cleanup of abandoned and uncontrolled hazardous waste sites. The law also gave Federal authority to directly respond to accidents, spills and emergency releases of hazardous substances that may endanger the environment or public health. Under Superfund, the Environmental Protection Agency (EPA) has the power to order responsible parties to clean up the site or EPA will clean up the site when responsible parties refuse or are unable to obey clean-up orders. If EPA cleans up the site, EPA can recover its costs from the responsible parties who are considered liable.

In 1986, CERCLA was amended by the Superfund Amendments and Reauthorization Act (SARA). SARA stressed the importance of permanent actions and innovative treatment technologies for cleaning up hazardous waste sites; new enforcement and settlement tools; increased state participation; increased focus on human health problems posed by hazardous waste sites; encouraged more citizen participation on clean-up; and increased the size of the Superfund trust. Title III of SARA authorized the Emergency Planning and Community Right-to-Know Act (EPCRA) which was designed to help communities protect public health, safety and environment from chemical hazards. Title III established the Toxics Release Inventory (TRI) which provides information to communities about chemical hazards released in their locale. Businesses are required to report the location and quantities of chemicals stored on-site and annually report on the release or transfer of chemicals. TRI requires reporting on approximately 650 chemicals.

The Federal Insecticide, Fungicide, & Rodenticide Act (FIFRA) was enacted in 1996. The focus of FIFRA is the control of the distribution, sale, and use of pesticides. All pesticides must be registered with EPA prior to manufacturing, distribution or sale. Registration assures that pesticides will be properly labeled, and if used in accordance with label specifications, will not cause unreasonable harm to the environment. FIRFA also requires that all pesticide users obtain certification for pesticide application and register when purchasing pesticides.

II. CHARACTERISTICS OF ENVIRONMENTAL DATA

A number of challenges exist in using environmental data to describe trends in environmental contamination or characterize environmental conditions. Not all of the environmental standards adopted under congressional legislation were created with human health criteria in mind. With the possible exception of SDWA and the criteria pollutants outlined under CAA, many environmental regulations were not based upon considerations of specific human health outcomes, but rather upon principles of curbing environmental degradation. Therefore, data is generally collected and reported in a manner consistent with such principles.

Many environmental databases and reporting practices only record events when there has been a violation. Often times, unless a specific chemical exceeds a standard level or concentration, sampling results may go unreported. These non-violation detected levels are rarely reported, if at all and therefore such monitoring programs may not represent a source of reliable, quantifiable data. Furthermore, monitoring equipment may be calibrated to detect only levels that exceed the standard, and not sensitive enough to accurately read levels below the standard. Therefore, some environmental data can be difficult to collect, interpret, or extrapolate toward human health endpoints.

It is important to distinguish between two primary classes of reported environmental data—emissions data and monitoring data. ***Emissions*** data are generally reported through the facility permitting process. Emissions data summarize the quantity of pollutants released into the environment during a year. They are normally estimated from amounts of material consumed or product produced within a given facility. Some estimates are for individual sources, such as factories, and some estimates are county totals for classes of sources, such as vehicles. Emissions data may be reported as monthly, quarterly, or annual totals, and do not necessarily provide an accurate quantification of specific ambient chemical levels at any given point in time.

Monitoring data, on the other hand, is sampling by standard sampling methodologies, and provides quantification of ambient conditions of chemicals in the environment at a given point in time. For example, ambient concentrations of pollutants in outdoor air are measured at more than 4000 monitoring stations owned and operated by state environmental agencies. They forward hourly or daily measurements of pollutant concentration to EPA, and EPA computes a yearly summary for each monitoring station (e.g., maximum value, average value, number of measurements, etc.)

Most emissions inventory databases track air emissions and water effluent based upon annually submitted reports. These reports are based upon facility estimates of their own emissions, rather

than quantified measurements (i.e., monitoring). While such data can be useful for many types of assessments, it can be limited in its accuracy or applicability to short-term time trend analyses.

Some environmental standards were adopted with consideration of human health conditions in mind. These include the maximum contaminant levels (MCL's) in SDWA legislation, and criteria pollutant standards in the CAA. However, these standards were calculated using scientific risk assessment methodologies. Risk assessments use conservative values that are orders of magnitude higher than chemical concentration that may be predictive of a health outcome. Therefore, just because chemical levels are observed which may exceed a health-based standard, it does not necessarily mean that one would expect any adverse health outcome to result.

III. SURVEY of EXISTING ENVIRONMENTAL DATABASES

The following survey provides a brief description of the most frequently accessed Federal environmental databases. Data from the EPA website may be found in multiple locations with each portal offering different ways to query the data and output. Envirofacts is a data warehouse for accessing multiple datasets from a single location (see description below). States agencies are mandated by environmental legislation to monitor for regulatory compliance and compile inventories of air, water, and hazardous wastes. These agencies generally house repositories of additional data that may be more extensive and specific to individual states than the data reported to EPA. EPA houses data that states are mandated to report; therefore it is important to investigate any additional specialized databases that may exist for individual state programs.

AirData/AIRS Database (EPA) Air Pollution Data

EPA's AirData/AIRS database contains both emissions and monitoring data. The data are drawn from the National Emissions Inventory (NEI) database on point, non-point and mobiles sources and the Air Quality System (AQS) which houses data on criteria and hazardous air pollutants. The NEI data includes information on sources that emit criteria pollutants and their precursors and hazardous air pollutants (HAPs). The NEI is estimated data computed from five primary data sources: 1) emissions inventories conducted by state and local agencies; 2) databases related to EPA's Maximum Achievable Control Technology programs to reduce HAP; 3) TRI data; 4) emission estimates computed using mobile source methodology; and 5) area source emission estimates computed using emission factors and activity data. The database includes estimates of annual pollutant emissions from point, non-point and mobile sources for the 50 states, Washington DC, Puerto Rico and the Virgin Islands. The national inventories are conducted every three years; this data is added to the inventory database. Annual estimates of air pollutants emitted from point, non-point and mobile sources are extracted for the AirData database.

The Air Quality System (AQS) is a database containing criteria pollutant monitoring and hazardous air pollutant data. AirData provides reports and maps based on a subset of the criteria pollutant data housed in AQS. AQS is updated weekly, although most states report monitoring data monthly. AQS computes annual summary values (e.g., average, high concentration) based on hourly and daily measurements submitted for each monitor. Only summary data are available

through AQS. Individual hourly and daily measurements (raw data) are available by submitting an AQS data request to EPA.

Utility of AirData: Some monitors capture ongoing samples at regular intervals over time. Data can be paired with meteorological data to model fate and transport or exposure models within a region. The annual summaries are available by monitor, geography, pollutant and year.

Limitations of AirData: Many monitors operate for only short periods of time, or have intermittent periods with no data collection which yields gaps in data. In several States, violation data is not flowing into AirData. AirData represents the best available data based on state reporting. AQS is updated daily by states and local organizations that own and submit data. Efforts are made at EPA to ensure the accuracy of the data but missing values do occur due to incomplete reporting or values have been changed due to quality assurance measures. Maps and reports created in AirData are extracted from EPA's primary pollution databases and thus reflect the most current data available however; some values may be out of date. EPA does try to synchronize AirData with its other primary databases through regular monthly updates but the user should check the database status to find out the latest update.

NEI is based on data from state and local agencies and thus not a database of official state emissions data. Because the NEI is a composite of emission estimates, the data vary in quality, level of detail, and geographic coverage. Additionally, the quality and accuracy of emissions estimates varies based on pollutant and source. Estimation techniques and quality of the estimates will vary between source categories and major, area and mobile sources. NEI data is updated semi-annually when new quality assured data becomes available. New methods for estimating emissions are used to improve consistency and accuracy. When new methods are applied, revisions to the data are only made for the most recent data. Therefore, EPA cautions that AirData should not be used for trend analysis without confirming that similar estimation methods were used for the years being compared.

EPA cautions users not to infer a qualitative ranking order of geographic areas based on AirData reports. Air pollution measured near a monitor may not be representative of prevailing conditions in the surrounding area. Pollutants emitted from a particular source may have little impact on the immediate surrounding area and the estimated amount does not indicate if the source is in compliance with regulations.

Air Quality Index (AQI)-AirNow (EPA)

The Air Quality Index (formerly known as the Pollution Standard Index or PSI) is a scale that provides information about local air quality and whether pollution levels pose a health concern. The AQI is computed for five major air pollutants: ground-level ozone, particulate matter, carbon monoxide, sulfur dioxide, and nitrogen dioxide for approximately 275 major metropolitan areas. The Environmental Protection Agency (EPA) has established national air quality standards to protect against harmful health effects for each of these pollutants.

Networks of monitors measure the concentrations of the major pollutants. These measurements are converted to an Air Quality Index value for each pollutant. The AQI is calculated by converting the measured criteria pollutant concentration into an index value. The daily AQI is the highest index value among the major pollutants and placed in a category based on the index value. AQI scale ranges from 0 to 500. The higher the AQI value, the greater the level of air pollution and the greater the health concern. EPA's AirNow website provides a variety of air quality maps and forecasts. Current AQI conditions are available as well as daily air quality forecasts provided by state and local agencies. Nationwide and regional ozone maps covering 44 states are updated hourly.

Utility of AQI Data: Provides good general guidance for overall air quality for at-risk populations. If daily values are available, AQI is useful for tracking general air quality trends.

Limitations of AQI Data: Data are not fully verified and validated through QA procedures because of the need to make maps as "real-time" as possible. Daily values (raw data) are not available from EPA. The data collected for AQI reporting are not submitted to the AQS.

Toxic Release Inventory System (TRIS) / TRI Explorer (EPA)

Toxics Release Inventory is a database containing information on toxic releases and waste management activities for individual facilities. Certain industry groups and federal facilities are required to report annual releases for approximately 600 chemical compounds. Facilities meeting the activity threshold listed below are required to calculate annual toxic releases, in pounds per year, based on monitoring data or estimates. TRI data describes the amount of chemical(s) released and whether they are emitted into the air, water, underground well injection, transported off-site, etc. A facility must file a TRI report with the State/EPA if the facility:

- Has 10 or more full-time employees, and
- Manufactures or processes over 25,000 pounds of the approximately 600 designated chemicals or 28 chemical categories specified in the regulations, or uses more than 10,000 pounds of any designated chemical or category, and
- Engages in certain manufacturing operations in the industry groups specified in the U.S. Government Standard Industrial Classification Codes (SIC) 20 through 39, or
- Is a federal facility

Utility of TRI Data: TRI data is useful for identifying potential concerns due to chemical releases and waste management activities. Facilities or geographical areas can be compared as well as identifying potential hot spots using TRI data. Additionally, pollution control and waste reduction programs can be tracked. TRI data can also provide a starting point to evaluate potential exposures.

Limitations of TRI Data: Data is reported on an annual basis only. The data do not reflect exposures to chemicals, only chemical releases and waste management activities. TRI does not cover all sources of releases or all toxic chemicals or industries. Some facilities report releases based on monitoring data, others report estimates based upon operational records. TRI values do not represent ambient concentrations/levels at any given point in time. Reported releases and waste management activities do not provide information to determine if the public has been

exposed. Release estimates are not sufficient to determine exposure or calculate adverse health effects to humans or the environment.

Safe Drinking Water Information System (SDWIS) (EPA)

Drinking water information is stored in SDWIS, the EPA's Safe Drinking Water Information System. Data can also be accessed through EnviroFacts. SDWIS contains information about public water systems and violations of EPA's drinking water regulations. The database allows users to search for their water supplier, violations and enforcement actions.

Utility of SDWIS Data: SDWIS provides detailed information about a water systems location, population and county served, system status, primary water source, any violations and enforcement actions. For violations, the dates, type of violation, contaminant and contaminant level is provided.

Limitations of SDWIS Data: EPA acknowledges past inaccuracies and underreporting of some data in the Safe Drinking Water Information System. EPA is working with the states to improve the quality of the data. For example, a deficiency in the quality of reporting is that violations are often reported simply as detect vs. non-detect, as opposed to reporting a specific value that can be compared to the MCL standard.

National Drinking Water Contaminant Occurrence Database (NCOD) (EPA)

NCOD is a national database containing occurrence data for regulated and unregulated contaminants in public water systems. The database contains sampling data of drinking water collected from public water systems that EPA is currently using or has used for analysis, rulemaking and rule evaluation. Data is available by contaminant, geography, sample date and other water system attributes.

Regulated occurrence data are sample data from public water systems for contaminants with a health-based standard. The unregulated occurrence data are sample data from public water systems for contaminants not having a health-based standard under the Safe Drinking Water Act. Under the Unregulated Contaminant Monitoring Rule (UCMR), EPA collects data to evaluate and prioritize contaminants on the Contaminant Candidate List. These contaminants are considered for possible new drinking water standards. Two rounds of data were collected for these unregulated contaminants. The Round 1 dataset contains monitoring data for 62 unregulated contaminants collected from 1988-1992, from 40 states. The Round 2 dataset contains monitoring data for 48 unregulated contaminants collected from 35 states from 1993-1997. A summary report (June 2001) describes the contaminants, data quality assessments and contaminant occurrence analysis. Additional data is available for 2001-2003.

Ambient occurrence data is also available from monitoring of surface and ground water sources. This data is available from the Legacy Data Center and STORET. Both systems contain raw biological, chemical and physical data on surface and ground water collected by federal, state, and local agencies, volunteer groups, academics and other.

Utility of NCOD: NOCD provides occurrence data for both regulated and unregulated contaminants in public water systems. The data have been extensively checked for data quality and analyzed for national representativeness. The data can be organized and analyzed by contaminant, geography, sample date and other water system attributes. Additionally, EPA reports are available which provide information for understanding the data and their quality, and analyses that have been performed using this data.

This site has recently been rebuilt. Previously, you could obtain only "raw," unchecked data. Now you can access the latest UMCR data as well as static datasets that have been used in published regulatory analyses. The static datasets have been extensively quality-checked, and their corresponding reports provide metadata of the data.

Limitations of NCOD Data: The UCMR was designed to assess regulation of contaminants on a national level; therefore caution should be used in interpreting the data at a smaller geographical scale.

Permit Compliance System (PCS) (EPA)

The Permit Compliance System (PCS) provides information on facility operations that have been issued permits to discharge wastewater into ambient water bodies, such as rivers, under the National Pollutant Discharge Elimination System (NPDES). States report facility effluents to EPA's NPDES program. Reports can be generated bout facility information, when a permit was issued/expires, how much the facility is permitted to discharge, monitoring data on pollutants discharged, quantity of effluent, receiving waters, date and nature of violations, and more. States utilize reported quantities to devise comprehensive water quality management plans and assess whether specific water bodies are threatened by nearby facility operations.

Utility of PCS Data: PCS data provides detailed information about a facility; violations; effluent limits and measurements; permit history; hearings; pretreatment assessments; compliance schedules and violations; and inspections. Data can be queried by facility name, geographic location, standard industrial classification, and chemicals. Additionally, customized queries can be created. Facilities can be mapped to track discharges into waterways that serve as drinking water sources.

Limitations of PCS Data: The data from customized queries is relatively cumbersome to download, and exceedingly difficult to get into a database. The process is made even more cumbersome by output that contains many blanks, repeated data entries, and incomplete fields with no apparent explanation. EPA cautions that the completeness of data for larger facilities is much higher than for smaller facilities for which required data elements varies. It may be necessary to request additional records from State agencies on small facilities. Data regarding inspections from California is not being entered into PCS. Research is being done to find the cause of the problem. Violations are calculated in the PCS system by comparing monthly discharge monitoring reports with permit limits. If up to date permits limits and current discharge monitoring reports are not available, miscalculation of violations may occur or missing or erroneously calculated.

Legacy Data Center (LDC) (EPA

The LDC is a static, historical water quality database containing data dating back to the early part of the 20th century and collected up to the end of 1998. It contains raw biological, chemical, and physical data on surface and ground water for all 50 States. Each sampling result provides locational information (including latitude and longitude), when the sample was taken, the media sampled (water, sediment or fish tissue) and the name of the organization that sponsored the monitoring.

Utility of LDC data: LDC data can provide some information on background water quality conditions.

Limitations of LDC data: The LDC data are static and of undocumented quality. Data is not updated and will not change over time. All new data is being entered into STORET. EPA does not endorse the quality of this data.

Storage and Retrieval System for Water and Biological Monitoring Data (STORET) (EPA)

STORET is an operational data system that contains data collected beginning in 1999, along with older data that has been properly documented from the Legacy Data Center. It contains raw biological, chemical, and physical data on surface and ground water for all 50 States. The data is grouped into five categories: Organization, Projects, Sites (Stations), Samples and Results. Each sampling result provides locational information (including latitude and longitude), when the sample was taken, the media sampled (water, sediment or fish tissue), and the name of the organization that sponsored the monitoring. STORET also contains information on why the data were gathered; sampling and analytical methods; analytical laboratory; quality control checks; and the personnel responsible for the data.

Utility of STORET Data: This web site provides query access to over 200 million monitoring records originally reported to EPA via the Legacy STORET System. The STORET data warehouse has extensive metadata documenting partner organizations methods, processes and standard practices. Data can be used to describe historical and current ambient water quality trends over time. Data can be retrieved based on geographic location, projects, stations or specific biological, chemical and physical parameters.

Limitations of STORET Data: All data from the United States Geological Survey was removed from the system.

National Water Information System (NWIS) (USGS)

NWIS provides access to water-resources data collected at approximately 1.5 million sites in all 50 states, the District of Columbia and Puerto Rico. NWIS stores data on site information, time-series (e.g., flow, stage), peak flow, ground water and water quality. Data can be retrieved by geographic area and by category: surface water, ground water and water quality. NWIS contains about 4.3 million water quality samples and 64 million water quality sampling results. The majority of the sites are wells; the remainder are water quality and stream-flow sites. Real-

time data is available for 5,000 sites. Extensive quality assurance/quality control documentation is available pertaining to laboratory protocols and changes, codes, sampling methods and analytical methods.

Utility of NWIS Data: Real-time data is available for evaluating current water quality conditions. Archived data is useful for time trend analyses of water quality conditions.

Limitations of NWIS Data: Real-time data is provisional and posted without any quality control review. Inaccuracies could be present due to instrument malfunctions and physical changes at the measurement site. Hence, the data may be revised after further review. Data users should also be cautioned of the provisional nature of real-time data before using it for public safety decisions. There are known coding problems with the data. Missing codes (e.g., agency, parameter) are not mandatory and therefore these fields may not be populated, especially in older data. Additionally, some options for sampling selection criteria are not populated (e.g., drainage area) as these field are not mandatory. Therefore a search for certain criteria will miss many records due to the fields not being populated.

The Facility Registry System (FRS) (EPA)

The Facility Registry System identifies facilities, sites or places subject to environmental regulations or of environmental interest. FRS creates a facility identification record that incorporates information from national systems, state facility records, and data collected from EPA's Central Data Exchange. FRS provides comprehensive environmental information (e.g., air, water, waste) about facilities, sites or places in a single report. The Facility Registration System has over 1.5 million unique facility records linking over two million programs of interest such as toxic releases, hazardous waste management, permit compliance and air quality. Queries can be done by facility, geography, industrial classification, selected program category (e.g., air, water, hazardous waste), or selected national system (e.g., TRI, AIRS, SDWIS).

Utility of FRS Data: A "one-stop shopping" resource for identifying all regulated facilities, sites or places. FRS integrates all environmental information about a facility in one report.

Limitations of FRS Data: The data in the reports cannot easily be downloaded and imported into a database. Data quality is equal to known problems inherent in source databases.

Envirofacts

Envirofacts is EPA's single point of access to environmental data. Reports and data are available in a variety of ways from the Envirofacts homepage. Beginning users can access the data using the Quick Start feature which are preformatted queries in which data can be retrieved by entering a zip code, city and state or county and state. More in-depth information can be retrieved using advanced query capabilities based on topics and other user-defined criteria. Envirofacts also incorporates mapping applications using EPA's EnviroMapper, Window to my Environment and Open Link. Maps can be generated of a specific geographic area to determine environmental conditions for that area, use GIS functions to examine spatial data, display multiple spatial layers, and query single points.

Utility of Envirofacts: Comprehensive database that is cross-referenced with other environmental databases for a wide selection of parameters of interest.

Limitations of Envirofacts: Data quality issues are the same as for individual datasets.

Comprehensive Environmental Response, Compensation, and Liability Information System (CERCLIS) (EPA)

CERCLIS is the official repository for site and non-site specific Superfund data in support of the Comprehensive Environmental Response, Compensation, and Liability Act (CERCLA). It contains information on hazardous waste site assessment and remediation from 1983 to the present. This includes location, status, contaminants, and actions taken.

Utility of CERCLIS Data: Identification and location of Superfund sites. Data element edits prevent invalid data from being entered into CERCLIS. Recent data are verified via various reports so that at the end of the month there is a high degree of confidence that the data are transferred correctly. Overall, the data are probably 95% accurately transferred.

Limitations of CERCLIS Data: Data contained in these reports are intended solely for use by employees of the U.S. EPA for management of the Superfund program. EPA reserves the right to change the data at any time without public notice. The National Priority List (NPL) sites do have latitude and longitude data accurate to within 30 meters and, in some cases, in the sub-meter range. The latitude and longitude for other sites is the centroid of the ZIP code area for the ZIP code of the site's address.

Resource Conservation and Recovery Act Information (RCRAInfo) (EPA)

RCRAInfo is a national program management and inventory system containing data about hazardous waste and hazardous waste handlers. It replaces the data recording and reporting functions the Resource Conservation and Recovery Information System (RCRIS) and the Biennial Reporting System (BRS). All generators, transporters, treaters, storers, and disposers of hazardous waste are required to report information about their activities. RCRAInfo queries can provide identification and location data for specific hazardous waste handlers. Additionally, queries can be used to find a wide range of information on treatment, storage, and disposal facilities regarding permit/closure status, compliance with Federal and State regulations, and cleanup activities.

Utility of RCRA Data: Identification and location of hazardous waste handlers and disposal sites. Trend analysis can be performed on hazardous waster generation, management and disposal within a specific region or from a specific facility.

Limitations of RCRA Data: The database includes both active facilities and those that are no longer managing hazardous waste and/or are permanently closed. When interpreting the data, it is likely that a recent inspection, violation, or enforcement action is a good indicator that the facility is actively managing waste. Efforts are under way to better identify active and inactive facilities within the datasets. EPA is evaluating the completeness of data reporting in regard to

the identification of significant noncompliance (SNC) and other violations under the RCRA program.

Enforcement and Compliance History Online (ECHO) (EPA)

ECHO is a database that allows one to search for facilities in a particular community for the purpose of determining whether EPA or state/local governments have conducted compliance inspections, whether violations were detected, and if enforcement actions and penalties were taken in response to environmental law violations. The information provided in the database relates to facilities regulated as CAA stationary sources, as CWA permitted dischargers, and/or as hazardous waste generators/handlers pursuant to RCRA.

Utility of ECHO Data: Database consolidates search parameters for air, water, and hazardous materials into one comprehensive search engine. The data may be useful for identifying and locating potential environmental hazards.

Limitations of ECHO Data: Data does not encompass minor CAA dischargers. Smaller facilities may not be inspected every two years, thus they may appear as "uninspected" in the database. Inspections for smaller Clean Air Act and Clean Water Act facilities may only be tracked in a State's database, not the Federal database. It may be necessary to request additional records directly from State officials for complete data for smaller facilities. In general, ECHO displays the past two years (eight quarters) of data from the date that the information was extracted from the program systems. ECHO reports on formal EPA and state enforcement actions, it does not include "informal" actions. Actions taken prior to the two years of data available in ECHO will not be included in ECHO query results. ECHO represents information as reported by the States into the Federal database. Some information is required and some is put in voluntarily by the States, therefore it is necessary to understand the required and non-required data elements in order to determine if the reports are complete. EPA provides documentation on what information is required to be entered into the system. The relevant State environmental agency may have additional information. See limitations of AirData/AIRS Database, PCS/NPDES Database, and RCRAInfo for additional information

Environmental Radiation Ambient Monitoring System (ERAMS) (EPA)

ERAMS is a national network of monitoring stations that regularly collect samples of air, precipitation, drinking water, and pasteurized milk for a variety of radionuclides and radiation types. The ERAMS network has more than 200 stations distributed across all 50 states. ERAMS samples radiation in all media on a regular schedule, with the capability to operate in an "emergency" mode in the event of the threat of a significant radiation release. The ERAMS Air Program consists of 50 sampling locations continuously sampling particulates for analysis of radioactivity. Analysis of air particulates occurs twice weekly. The ERAMS Drinking Water Program samples quarterly at 76 sites which are primarily major population centers. The Pasteurized Milk Program consists of 42 sampling locations representing a significant portion of the milk consumed in major population centers. The ERAMS Precipitation Program is made up of 37 sampling stations which routinely submit precipitation samples for radioactive analysis as rainfall, snow or sleet occurs.

Utility of ERAMS Data: ERAMS data can be use to estimate levels of radioactivity in the environment, including background radiation levels. ERAMS data can be used to track environmental releases resulting from nuclear emergencies and provide information for protecting public health.

Limitations of ERAMS Data: Data is not available for all media in all states. The sampling scheme is based on population centers which makes it difficult to describe conditions in small areas. Sampling stations may have changed location within a city or town over time and results are reported in the database from the same location. If the station is moved to a new city/town, it is entered into the database as a new location. Additionally, states may request that stations be established, relocated or terminated which results in changes in location or total number of monitoring stations within a state. The exact station location of historical data (i.e., latitude and longitude) is unknown as this information was not tracked. Latitude and longitude were assigned to the data using the city centroid for each monitoring station. Current station location is being tracked using latitude and longitude. The same station may monitor for multiple media; this is distinguished by a station-specific numbering system. Data from milk monitoring is reported monthly from 1973 to 1999. After March 1999, the sampling frequency was changed to quarterly. Data prior to 1974 is reported in summary form only. Individual sample results are available for data after 1978.

National Climate Data Center (NOAA)

The National Climatic Data Center (NCDC) is the largest active archive of weather data dating back to 1895. Data and maps are available on climate and weather, satellite and radar. More recent weather data and forecasts (i.e. days/weeks) can be found through the National Weather Service, also housed on NOAA's website. Unedited local data is also available through the NCDC.

Utility of Climatic Data: Archived data is useful for modeling fate and transport of chemicals in air and water.

Limitations of Climatic Data: Recent data has not undergone quality control processing.

National Agriculture Statistics Service (NASS) Agricultural Chemical Use Database and Agricultural Statistics Database

In coordination with the National Science Foundation (NSF) the U.S. Department of Agriculture (USDA) maintains the National Agriculture Statistics Service (NASS). Agricultural chemical usage data can be searched by commodity, year, state and active ingredient. Dynamic U.S. maps of can be generated for each active ingredient, commodity, year, and state. As the site is updated, it is expected to provide mean and total statistics over years, states, and commodities. County level data is available for crops, livestock and number of farms.

Utility of NASS data: Identification and location of pesticide usage to identify populations with the potential for exposure.

Limitations: This is survey data, not data on individual crops and pesticide usage at the farm level. Crops surveyed vary depending on year and state. Some crops were surveyed infrequently in both year and state. For example, Almond-All was surveyed in California in 1999 only. Not all states are surveyed each year. NASS statistics are not representative of total use of any agricultural chemical in the U.S. or at the state level. Statistics are provided for selected states and crops. The Census of Agriculture is only performed every five years.

IV. OVERALL UTILITY OF ENVIRONMENTAL DATA

Environmental monitoring and emissions data are valuable for characterizing environmental conditions. Many statistical procedures, database/spreadsheet manipulations, GIS tools, and simulation/forecasting models can be applied to assist with their interpretation. Through these methods, it is possible to approximate concentrations and transport of environmental contaminants throughout a locale or region.

V. OVERALL LIMITATIONS OF ENVIRONMENTAL DATA

Environmental data provides information about the quantity of a contaminant in the environment. Most environmental data is collected by federal and state environmental agencies as mandated by legislation. Environmental data is collected for regulatory or compliance purposes, that is, monitoring occurs to determine if the concentration of a compound is above or below a standard or if emission standards need to be reduced. The data are generally used for political or technological decision making. The standards that do exist are predominately based on toxicological studies that extrapolate the study results to standards that are protective of public health, not predictive of health outcomes. Environmental data are generally available as individual data points and is not usually in a format that can be used for tracking without modification of the database.

One of the biggest limitations to using environmental data is that is cannot be used to measure individual exposure; at best it can be used to characterize population exposure. Models exist to estimate exposure but these have their limitations in that they are based on assumption and uncertainties that are often difficult to account for in statistical analyses. Availability, coverage and quality are also important issues in determining the utility of environmental data. Frequency of data collection presents a problem in determining the concentration of a pollutant in the environment, especially low levels, because samples are collected at time intervals that are not comparable (e.g., hourly, daily, every six days, seasonally, yearly). Additionally, analytical methods are not sensitive enough to detect chemicals at low levels resulting in a dataset full of nondetected values for which estimated values are inserted (e.g., the detection limit or half the detection limit) in order to complete the dataset for analysis. Coverage of the data is not consistent across media and geographic boundaries. Monitors are located based on population density and proximity to identified hazards or emissions. In some cases, values from a single monitor may be available to characterize exposure for an entire state.

An additional limitation is that the data represents a “point in time”, that is, the value is representative of what is present in the media at the time of sampling and not reflective of its

variability in the environment. Relatedly, in many cases there is not enough data to look at trends of a pollutant because the type and number of pollutants monitored varies within and across geographical boundaries.

Monitoring data provide a valuable resource in terms of understanding levels of chemical compounds in our environment; however, the data do not provide a measure of exposure. Most environmental monitoring efforts derive numeric data gathered at a point in time and space. The problems associated with the analysis and interpretation of such data will always have temporal and spatial numerical variances to accommodate. For example, air pollution levels measured in the vicinity of a particular monitoring site may not be representative of the prevailing air quality of a county or urban area. Likewise, pollutants emitted from a particular localized source may have little impact on the immediate geographic area due to various meteorological patterns and trajectories.

Furthermore, the nature of many environmental monitoring programs is such that monitoring may occur sporadically; monitoring schedules or compound being monitored change, or monitoring stations are closed. As a result, one often encounters data sets where monitoring may have occurred for some duration, dropped off, and then resumed once again.

Likewise, especially with emissions data (i.e. NPDES), it is not unusual to encounter data sets containing blank fields, identical repeat entries, or entries where major parameters such as name or location of facility are left incomplete. It is not always apparent or explained as to why these inconsistencies exist in the data, but is likely attributed to technical difficulties in transmittal of permit data. Additional quality control measures to ensure smooth transmittal of data from states to EPA would be beneficial for facilitating public health tracking efforts and removing the uncertainty associated with inconsistent reporting.

VI. SUMMARY CONCLUSION

Environmental data is generally collected from different sources, sometimes through different collection methods, data formats, and with varying degrees of data quality control/assurance. These variations could limit the usefulness of collected data if they are not highlighted and assessed (Shao, 2003). For example, under the CAA, states can decide whether or not to monitor ambient concentrations of hazardous air pollutants (HAPs), and the level of effort varies considerably (Kyle, 2001).

The historical data collected from statutory monitoring sites has thus far provided useful background information. However, limitations on resources may restrict monitoring programs to taking 'snapshots' of air quality, for example, 24 hr values, which may prove unrepresentative. Conversely, longer-term data, such as weekly or monthly averages, may conceal important peaks in pollutants which may be more significant for health (Dunn, 1996). It may be beneficial if environmental health researchers were able to provide input related to how environmental monitoring was conducted. This may provide a means of recording meaningful data for both environmental compliance and environmental health tracking interests.

To some extent, the large volumes of environmental quality data, often of variable quality, have failed to provide decision-makers and the public with the answers to basic questions especially

concerned with human health and environmental impacts from multiple chemical stressors in the atmosphere (Peterson, 1999). It shall be increasingly important to develop the indicators and other tools which may bridge the gap between environmental and health data.

It is important to devise and implement a coordinated approach for the interpretation and reporting of environmental health data. To a large extent, this work has been approached in a rather piecemeal way, with epidemiologists, toxicologists or medical geographers, for example, operating independently and within disciplinary boundaries. The issue of public health-environment relations is one which crosses the boundaries of many disciplines and each of these disciplines brings its on set of approaches and methodologies. In establishing reliable indicators for interpretation of environmental and health data, it is important to integrate the knowledge, perspective, and results of all associated disciplines (Dunn, 1996).

References

Dunn, C and Kingham, S. 1996. Establishing links between air quality and health: searching for the impossible. Soc Sci Med. 42: 831-841.

Kyle, AD; Wright, CC; Caldwell, JC; Buffler, PA; and Woodruff, TJ. 2001. Evaluating the health significance of hazardous air pollutants using monitoring data. Public Health Reports 116: 32-44.

Peterson, PJ and Williams, WP. 1999. New indicator approaches for effective urban air quality management. Environ Sci & Pollut Res 6: 225-232.

Shao, Q; Li, Y; Campbell, E; DeBoer, ES; Laginestra, E; Statzenko, A. 2003. Statistical visualization for data exploration: a case study on Sydney Olympic Park. Chemosphere 52: 1601-1614

Appendix
Table of Federal Environmental Databases

Database	Agency	Description of Data	Coverage	Website
AirData/AIRS	EPA	Criteria Pollutants; Hazardous Air Pollutants (HAPs); Photochemical Assessment Monitoring (PAMS) and Meterological data	1971 – Current Criteria pollutants monitoring data updated monthly; HAPs monitoring data updated quarterly; Emissions data updated at least annually	AirData http://www.epa.gov/aqspubl1/select.html
Detailed Air Quality System	EPA	Criteria pollutants and PAMS VOCs	1994 – Current Data from 2001 to present may be updated	Technology Transfer System Air Quality System http://www.epa.gov/ttn/airs/airsaqs/detaildata/downloadaqsdata.htm
Air Quality Index (AQI)-AirNow	EPA	Air quality index based on criteria pollutants	Current conditions and forecasts Updated daily	AIRNow http://www.epa.gov/airnow/
Toxic Release Inventory System (TRIS)	EPA	Reported chemical releases to air, water, ground, or transported off site	1988 – 2002 Updated on rolling basis; data released annually	TRI Explorer http://www.epa.gov/triexplorer/ Envirofacts Data Warehouse Toxic Release Inventory Query Form http://www.epa.gov/enviro/html/tris/tris_query.html
Safe Drinking Water Information System (SDWIS)	EPA	Chemical contamination levels in drinking water systems	10 Years worth of data are maintained	Office of Ground Water and Drinking Water Safe Drinking Water Information System/Federal Version

			Updated quarterly	http://www.epa.gov/ogwdw/data/getdata.html Envirofacts Data Warehouse SDWIS Query Form http://www.epa.gov/enviro/html/sdwis/sdwis_query.html
National Contaminant Occurrence Database (NCOD)	EPA	Physical, chemical, microbial, and radiological contaminants in drinking water	1988 – 1992, 1993 – 1997, 2001 – 2003 (pending)	Office of Ground Water and Drinking Water National Contaminant Occurrence Database http://www.epa.gov/safewater/data/ncod.html
Permit Compliance System (PCS)	EPA	Permit compliance data for facilities discharging waste water into ambient waters (e.g., rivers, lakes, etc.)	Early 1970's – Present Updated monthly	Envirofacts Data Warehouse PCS Query Form http://www.epa.gov/enviro/html/pcs/pcs_query_java.html
Legacy Data Center (LDC)	EPA	Raw biological, chemical, and physical data on surface and ground water, sediment, and fish tissues	Early 1900 – 1998	STORET Legacy Data Center http://www.epa.gov/storpubl/legacy/gateway.htm
Storage and Retrieval System (STORET)	EPA	Water quality, biological, and physical data	1999 – Present Updated monthly	STORET Data Warehouse http://www.epa.gov/storet/dw_home.html
National Water Information System (NWIS)	USGS	Chemical and physical properties, water levels in various water bodies (real-time data often available)	Mid-1900 – Present Updated monthly	USGS Water Quality Data for the Nation http://waterdata.usgs.gov/nwis/qw
Resource Conservation & Recovery Act Information (RCRAInfo)	EPA	Identification and location data for specific hazardous waste handlers, and information on treatment, storage, and disposal facilities regarding permit/closure status, compliance, and cleanup activities	1980 – Present Updated monthly	Envirofacts Data Warehouse RCRAInfo http://www.epa.gov/enviro/html/rcris/
Biennial Reporting System (BRS)	EPA	Hazardous waste management data by facility	1989-1997 Updated every 2 years	Envirofacts Data Warehouse Biennial Reporting Query Form

				http://www.epa.gov/enviro
Facility Registration System (FRS)	EPA	Identifies facilities, sites, and places subject to environmental regulation	Mid-1900's to most current available Updated nightly	Envirofacts Date Warehouse Facility Registration System http://www.epa.gov/enviro/html/fii/index.html
Comprehensive Environmental Response, Compensation, and Liability Information System (CERCLIS)	EPA	Hazardous waste site assessments and remediation activities for abandoned or historical sites	1983 to most recent available Updated monthly	Superfund Information Systems http://www.epa.gov/superfund/sites/siteinfo.htm Envirofacts Data Warehouse CERCLIS Query Form http://www.epa.gov/enviro/html/cerclis/cerclis_query.html
Enforcement and Compliance History Online (ECHO)	EPA	Search engine for facilities that have undergone inspections and whether violations were detected	Three year history of compliance, enforcement and violation Updated monthly	Enforcement and Compliance Online http://www.epa.gov/echo/index.html Envirofacts Data Warehouse Enforcement and Compliance http://www.epa.gov/enviro/html/echo/index.html
Environmental Radiation Ambient Monitoring System (ERAMS)	EPA	Radiation data for air, water, precipitation, and pasteurized milk	1960 to most recent available Quarterly report	Envirofacts Data Warehouse ERAM Query Form http://oaspub.epa.gov/enviro/erams_query.simple_query
Locational Reference Tables	EPA	Describes location of regulated facilities	Updated Monthly	Envirofacts Data Warehouse Locational Information Query Form http://www.epa.gov/enviro/html/locational/lrt/ez.html
Agricultural Chemical Use Database	USDA/NSF	Agricultural chemical usage by state, year, acreage, and active ingredient	1990-2002 Updated annually	National Agricultural Statistics Service http://www.usda.gov/nass/ Agricultural Chemical Use Database http://www.pestmanagement.info/nass/index.html

Climatic Data Center	NOAA	Historical weather statistics regarding climatology, precipitation, temperature, etc.	1895-current Real-time; archived daily	National Climatic Data Center http://lwf.ncdc.noaa.gov/oa/ncdc.html