# A Methodology for Locating and Managing Dynamic Potential Source Water Contaminant Data

Project #4581

# A Methodology for Locating and Managing Dynamic Potential Source Water Contaminant Data

## About the Water Research Foundation

The Water Research Foundation (WRF) is a member-supported, international, 501(c)3 nonprofit organization that sponsors research that enables water utilities, public health agencies, and other professionals to provide safe and affordable drinking water to consumers.

WRF's mission is to advance the science of water to improve the quality of life. To achieve this mission, WRF sponsors studies on all aspects of drinking water, including resources, treatment, and distribution. Nearly 1,000 water utilities, consulting firms, and manufacturers in North America and abroad contribute subscription payments to support WRF's work. Additional funding comes from collaborative partnerships with other national and international organizations and the U.S. federal government, allowing for resources to be leveraged, expertise to be shared, and broad-based knowledge to be developed and disseminated.

From its headquarters in Denver, Colorado, WRF's staff directs and supports the efforts of more than 800 volunteers who serve on the board of trustees and various committees. These volunteers represent many facets of the water industry, and contribute their expertise to select and monitor research studies that benefit the entire drinking water community.

Research results are disseminated through a number of channels, including reports, the Website, Webcasts, workshops, and periodicals.

WRF serves as a cooperative program providing subscribers the opportunity to pool their resources and build upon each other's expertise. By applying WRF research findings, subscribers can save substantial costs and stay on the leading edge of drinking water science and technology. Since its inception, WRF has supplied the water community with more than $460 million in applied research value.

More information about WRF and how to become a subscriber is available at www.WaterRF.org.

# A Methodology for Locating and Managing Dynamic Potential Source Water Contaminant Data

Prepared by:
**Jeffrey S. Rosen, Ayhan Ergul, Margaret Kearns, Josh Coefer, Tim Bartrand, and Kristyn Stevens**
Corona Environmental Consulting, 1001 Hingham St., Rockland, MA 02370

Jointly sponsored by:
**Water Research Foundation**
6666 West Quincy Avenue, Denver, CO 80235

**Aqua America, Inc.**
762 West Lancaster Avenue, Bryn Mawr, PA 19010

**West Virginia-American Water Company**
4002 Ohio River Road, Huntington, WV 25702

**Greater Cincinnati Water Works**
4747 Spring Grove Avenue, Cincinnati, OH 45232

and

**AWWA Government Affairs Office**
1300 I Street NW, Suite 701W, Washington, DC 20005

Published by:

Water Research Foundation℠

## DISCLAIMER

This study was jointly funded by the Water Research Foundation (WRF), Aqua America, Inc., West Virginia-American Water Company, Greater Cincinnati Water Works, and AWWA Government Affairs Office (co-sponsors). WRF and co-sponsors assume no responsibility for the content of the research study reported in this publication or for the opinions or statements of fact expressed in the report. The mention of trade names for commercial products does not represent or imply the approval or endorsement of WRF or co-sponsors. This report is presented solely for informational purposes.



ISBN 978-1-60573-271-8

Printed in the U.S.A.

# CONTENTS

# TABLES

# FIGURES

# FOREWORD

The Water Research Foundation (WRF) is a nonprofit corporation dedicated to the development and implementation of scientifically sound research designed to help drinking water utilities respond to regulatory requirements and address high-priority concerns. WRF's research agenda is developed through a process of consultation with WRF subscribers and other drinking water professionals. WRF's Board of Trustees and other professional volunteers help prioritize and select research projects for funding based upon current and future industry needs, applicability, and past work. WRF sponsors research projects through the Focus Area, Emerging Opportunities, and Tailored Collaboration programs, as well as various joint research efforts with organizations such as the U.S. Environmental Protection Agency and the U.S. Bureau of Reclamation.

This publication is a result of a research project fully funded or funded in part by WRF subscribers. WRF's subscription program provides a cost-effective and collaborative method for funding research in the public interest. The research investment that underpins this report will intrinsically increase in value as the findings are applied in communities throughout the world. WRF research projects are managed closely from their inception to the final report by the staff and a large cadre of volunteers who willingly contribute their time and expertise. WRF provides planning, management, and technical oversight and awards contracts to other institutions such as water utilities, universities, and engineering firms to conduct the research.

A broad spectrum of water supply issues is addressed by WRF's research agenda, including resources, treatment and operations, distribution and storage, water quality and analysis, toxicology, economics, and management. The ultimate purpose of the coordinated effort is to assist water suppliers to provide a reliable supply of safe and affordable drinking water to consumers. The true benefits of WRF's research are realized when the results are implemented at the utility level. WRF's staff and Board of Trustees are pleased to offer this publication as a contribution toward that end.

Charles M. Murray
Chair, Board of Directors
Water Research Foundation

Robert C. Renner, P.E.
Chief Executive Officer
Water Research Foundation

# ACKNOWLEDGMENTS

Corona Environmental Consulting would like to acknowledge the Metropolitan Washington Council of Governments and its member utilities for their thoughtful participation in the scoping and requirements workshops and other input into the development of this methodology. Research assistance was provided by Corona data analysts Nathan Einstein, Jennifer Benjamin, Jon Janezic, Max Fenton and Saul Rosen with additional analysis performed by Tom Hargy, Tom Taggart and Melissa Rosenbladt.

# EXECUTIVE SUMMARY

## OBJECTIVES

The purpose of this project was to develop and demonstrate a methodology for identifying potential sources of contamination upstream of drinking water intakes, particularly from above ground storage tanks (ASTs) and their contents. A key objective was to develop an information systems approach to visualize and analyze potential sources of contamination using a Geographic Information System (GIS). The project was designed to help drinking water utilities follow the American Water Works Association (AWWA) G300 Standard (American Water Works Association 2014) and guidance (Sham et al. 2010) to complete and update source water characterizations, develop action plans, implement source water protection (SWP) practices, and evaluate and revise their SWP programs. This report provides guidance and a methodology to address many of the elements of a source water characterization (Table 1.1): delineation of upstream contributing areas, collection of water quality and quantity data, and identification of contaminant sources and land use patterns. The work also contributes to emergency preparedness, incident response, and health and safety management efforts.

## BACKGROUND

The 1996 amendments to the Safe Drinking Water Act (SDWA) (EPA 1974) required state regulatory agencies to conduct source water assessments (SWAs) for public drinking water sources. To a large extent, the level of detail in the assessments was left to the discretion of the states, and there was no clear requirement for updating the SWAs. Although some states have updated, or are in the process of updating, these reports, many of these original reports have become obsolete over the past decade as conditions on the ground have changed. Ideally, communities use SWAs to develop and implement source water protection (SWP) programs to strengthen or enact barriers between potential drinking water contaminant sources and drinking water intakes. When information becomes outdated, efforts to implement SWP programs are less effective and can lead to gaps in the knowledge needed to make critical and timely decisions during an incident. Information on the location; chemical contents; quantity; and physical, chemical, and toxicological properties of potential contaminants stored in ASTs is critical for detecting a spilled chemical, locating its source, addressing source cleanup, making operational decisions, and crafting communications. This same type of information can be used before a spill to prioritize potential sources of contamination upstream of intakes and develop SWP actions to minimize risks and identify critical data gaps.

Although the SDWA does not have specific requirements for SWP or require drinking water utilities to implement SWP programs, state regulatory agencies, AWWA, and others in the drinking water community recognize the importance of SWP as a component of the multiple barrier approach to producing safe drinking water. AWWA has developed and refined standards for SWP efforts at drinking water utilities (Sham, Gullick, Long, & Kenel, 2010). The SWP activities described in this report were formulated to be consistent with the elements of the AWWA standard and guidance for drinking water utilities and, in some cases, exceed those requirements. For example, the methodology described in this report includes contaminant research guidelines to document data of particular importance to drinking water utility managers for responding to emergencies and prioritizing monitoring, event detection, and SWP activities.

## APPROACH

The project began with scoping and requirements analysis workshops with project partners. The objectives of these workshops were to articulate specific priority issues, define priority sites within the zones of concern upstream of intakes, identify available datasets for five pilot sites, and detail the necessary and desired functionality for the information system. Prototype GIS tools were demonstrated at the requirements workshop to stimulate creative thinking and discussion about the possible utility of the information system approach for planning, management, and emergency response. Reports detailing the outcomes of the workshops were prepared and distributed to both the project advisory committee and a steering committee made up of pilot site project partners. These two groups had the opportunity to provide input on the methodology as it was being developed.

The data collection methodology was developed simultaneously with the creation of an information system for project pilot sites. Feedback from, and interactions with, pilot site facility managers and other members of the drinking water community generated new leads about organizations with useful datasets and strategies to obtain more challenging datasets. The final report includes a description of the methodology and lessons learned. Detailed information about the pilot sites' potential sources of contamination was loaded into the password-protected information system, but is not presented in the report in light of concerns about data and source water security and confidentiality. Project staff and pilot site partners worked with data providers to complete nondisclosure agreements for confidential data access when necessary.

## RESULTS/CONCLUSIONS

A methodology to collect information about potential sources of acute contamination and load it into an online information system was developed and applied to five pilot surface water systems in the eastern United States. This eleven-step method covered the process from delineation of upstream contributing areas and zones of concern, through data mining, contaminant research, aerial image analysis, and SWP planning and implementation. The challenges of maintaining this type of voluminous, multi-source database in a dynamic, online geospatial platform are discussed along with the many benefits of this approach, which ultimately produces a living source water assessment that incorporates data as they become available.

Some of the data collection challenges centered on confidential data, which required the development of special processes with state agencies, membership in the Local Emergency Planning Committee, and/or non-disclosure agreements. Other challenges included identifying sources of information about the identity and quantity of chemicals traveling on roads, rails, and waterways through the zones of concern and researching trade secret and less commonly encountered chemicals.

Approaches to incorporating these data into a prioritization scheme or, ultimately, full risk assessments, are discussed, along with basic approaches for event detection and early warning systems. Though a detailed discussion of these topics was beyond the scope of this project, the data collected following the methodologies described herein and housed in an online, dynamic GIS could be used as a basis for many of the decisions and analyses that must be made to prioritize sites or detect spill events. For instance, the most ubiquitous chemicals, which were also stored in the largest quantities, were petrochemicals. These were abundant and present in large ASTs in each of the five pilot site drainage areas. This suggests that, for many surface water intakes, important components of a SWP plan include emergency response protocols for petrochemical spills, and plans to track the location, size, secondary containment and inspection status of the largest ASTs.

A number of data gaps remained at the conclusion of the project that will continue to be discussed among the drinking water community at large. The most significant gaps were the number of ASTs identified in aerial imagery for which no data about size, contents or condition were available in existing datasets, and a general lack of information about treatability and chemical and toxicological properties for many contaminants, especially those that were considered trade secrets or were less commonly encountered. Some potentially important datasets and databases had limited access for non-government or non-first responder users.

## APPLICATIONS/RECOMMENDATIONS

The data collection and management methodology described in this report will help drinking water utilities enhance and maintain their source water characterizations and emergency preparedness and response capabilities. The development of a GIS with a regular data update schedule will also enable utilities to review and revise their priorities, detection capabilities, and emergency response plans on a regular schedule. More specifically, utilities can use the contaminant locations, chemical properties, and data gaps identified with this methodology to prioritize specific sites or facilities for follow-up data collection, remedial action, or emergency response planning activities.

The method also generates easy to understand information that will support communication with significant upstream facility owners, the emergency response community, customers, and the general public. For instance, contact information obtained from multiple data sources could be used to initiate or maintain communications with managers of significant facilities upstream. Utilities may also wish to conduct spill response exercises at priority sites to better understand the critical data gaps, resource availability, and communication products (e.g., maps, reports, etc.) needed during an incident.

A review of known potential contaminants upstream of an intake, and their physical characteristics, could inform early warning and event detection activities and technology choices, as well as laboratory sampling capabilities that may be required during a contamination event. These same data will be invaluable during an incident when time is short and rapid access to data on time of travel estimates, detection and treatability guidance, fate and transport factors, health effects, and safety precautions is critical. Following this methodology, the chemical and physical properties of a contaminant can be searched in an emergency situation to help identify the location and/or identity of a chemical spill. For example, because chemical vapors are invisible, odor may be one of the most important early detection signals and may be related to spilled chemicals or breakdown products. Access to a shared contaminant database will increase the chance that breakdown products or chemicals previously unknown within the drainage area will be documented and that information about them will be accessible during an incident. Similarly, sudden changes in color or taste can indicate a change in water quality that could be investigated using an appropriately designed contaminant database.

Finally, since the contaminant database compiled for this project accumulates the information for chemicals at many intakes, the assessment of chemical research gaps will help to inform a larger discussion within the drinking water community about which data gaps are the most critical to fill and how to accomplish this work.

## RESEARCH PARTNERS

- American Water Works Association
- West Virginia American Water Company
- Aqua America, Inc.
- Greater Cincinnati Water Works

## PARTICIPATING UTILITIES

- Metropolitan Washington Council of Governments
- Interstate Commission on the Potomac River Basin
- Fairfax Water
- Loudoun Water
- Washington Aqueduct
- City of Rockville
- Washington Suburban Sanitary Commission
- DC Water
- City of Leesburg
- Prince William County Service Authority

# CHAPTER 1
# INTRODUCTION

## OVERVIEW

Amendments to the Safe Drinking Water Act (SDWA) required state regulatory agencies to develop source water assessment programs (SWAPs). To a large extent, the level of detail of source water assessments (SWAs) was left to the discretion of the states and there was no clear requirement for updating the SWAs, though some states have taken on this challenge in recent years. Ideally, communities and utilities use the SWAs to develop and implement source water protection (SWP) programs to strengthen or enact barriers between potential drinking water contaminant sources and drinking water intakes. SWAs completed in the early 2000s are now out of date and a number of recent spills and drinking water contamination events have highlighted the need to update these assessments, revisit them regularly and improve SWP planning and emergency spill response.

Although the SDWA does not have specific requirements for SWP, state regulatory agencies, the American Water Works Association (AWWA) and others in the drinking water community recognize the importance of SWP as a component of the multiple barrier approach to producing safe drinking water. The AWWA developed and refined standards and guidance for source water protection (Sham, Gullick, Long, & Kenel, 2010) (American Water Works Association, 2014) that include:

- A SWP vision statement
- Source water characterization
- A statement of program goals and objectives
- An action plan
- Implementation of SWP practices
- SWP program evaluation and revision.

## PURPOSE

The purpose of Water Research Foundation project 4581 was to develop and demonstrate a methodology for identifying above ground storage tanks (ASTs) and their contents as well as other potential sources of contamination upstream of drinking water intakes. The project was designed to help drinking water utilities follow AWWA guidance to complete and update source water characterizations, develop action plans, implement SWP practices and evaluate and revise their SWP programs. This report provides guidance and a methodology to address many of the elements of AWWA's source water characterization, as described in the G300 Standard Guidance (Sham, Gullick, Long, & Kenel, 2010). These elements include delineation, water quality and quantity data, contaminant sources and land use (Table 1.1). The work also contributes to emergency preparedness, response and health and safety management.

**Table 1.1**
**Elements of source water characterization**

| | |
|---|---|
| **Delineation** | Utility delineates source waters and areas of concern (for both surface and ground waters) |
| **Water quality and quantity data** | Water quality data from the point of withdrawal from the source and from other locations within the watershed are assembled and used to identify water quality issues and to formulate monitoring plans |
| **Contaminant sources and land use** | Points of concern should be identified, along with documentary information supporting their selection as points of concern. Plans should be in place to monitor future activities and development that can affect the watershed. |
| **Compliance with regulatory requirements** | The utility should identify local, state, federal and other regulations that apply to the utility and source water protection area and make those regulations accessible. |
| **Security planning and implementation** | The utility should have plans that address the response of personnel in the event of a security incident. |
| **Emergency preparedness and response** | The utility shall document emergency response plans (with specific instructions to personnel) for use in an emergency. |
| **Health safety management** | The utility shall document the health and safety procedures that are designed to safeguard employees and visitors engaged in operational activities pertaining to watershed management. |
| **Stakeholders** | The utility shall identify source water area stakeholders, their roles and existing initiatives in which they may be engaged. |

*Source:* Adapted from *AWWA G300-14 Source Water Protection* by permission. © 2014 American Water Works Association.

## PROCESS

The project was conducted for five pilot sites and began with a preliminary data collection effort that was followed by scoping sessions for each of the pilot sites. During the scoping sessions, partner utilities discussed topics of concern for their source waters (e.g., specific facilities or contaminants) as well as potential data sources (Appendix A). A prototype information system was developed over the next few months based on the scoping session and other discussions with partner utilities. The prototype was used to facilitate a full requirements analysis workshop with all of the project partners to fully describe the requirements and "desirements" for the information system. These requirements were documented in the requirements report (Appendix D) and the most critical functions were implemented over the course of the project.

The methodology for data collection, quality assurance and data management were developed and tested on five pilot sites with three partner utilities: West Virginia American Water, Aqua Pennsylvania and Greater Cincinnati Water Works. For four of the five pilot sites, the full data collection and aerial image analysis processes were completed to provide the utilities with actionable information about potential threats to their source waters. The work performed for the Greater Cincinnati Water Works pilot site assessment had a different goal, which was to compare the utility's existing source water assessment database and methodology to the results of the state and federal data collection methodology developed for this project and identify the strengths and weaknesses of the two methods. As such, aerial image analysis was completed for just two facilities in this utility's zone of concern.

## APPLICATION

Data collected following the methodology described in this report can be used by drinking water utilities to support a number of SWP activities:

- Prioritize specific sites for follow-up data collection, remedial action or emergency response planning activities;
- Implement and manage contaminant warning and event detection technologies;
- Collaborate with others in the drinking water community to prioritize and fill critical gaps in the knowledge of contaminant physical, chemical or toxicological properties;
- Manage data updates;
- Regularly review and update source data, risk assessments, SWP priorities and response plans;
- Conduct spill exercises; and
- Respond to emergencies.

# CHAPTER 2
# METHODOLOGY

## PURPOSE AND DESIRED OUTCOME

The purpose of WRF 4581 was to develop and demonstrate a methodology for identifying potential sources of contamination. The project had a dual focus on improved data stewardship and development of an information systems approach that included tools for visualization and analysis. The desired outcomes were to facilitate the organization, maintenance and delivery of critical information about the locations and physical, chemical and toxicological properties of chemicals stored in the utilities' critical zones and to produce a detailed methodology for locating, extracting, quality controlling and maintaining this information using a GIS-based system. The project also developed methods for visually analyzing aerial imagery to identify potential sources of contamination not indicated in available GIS datasets. The report includes an assessment of data sources that were difficult to access and critical data gaps that the drinking water community may desire to fill for the protection of public health and safety. The approach was applied to five pilot sites to ground-truth the methodology and provide useful input during the development of the GIS. Data collected for the pilot site analyses are not presented here due to security concerns but were entered into a secure online geospatial platform for review and use by the partner utilities. In addition to pilot site results for partner utilities, the project's final products will help the utilities improve and update their source water assessments and develop long term sustainable plans for defending water intakes from chemical contamination from sources in the critical zones of intakes.

## DATA COLLECTION

Many current source water protection plans for surface water withdrawals were developed during the late 1990's and early 2000's. Since that time, the digital data management capabilities of both the federal government and state agencies have advanced and there are many more spatial and regulatory data products available, many accessible through the internet and some with near real-time access. These advancements in data management and the increased use of GIS technology have made it possible for drinking water utilities to update their SWP plans with previously unavailable information, updated regulatory databases and live links to data sources that did not exist a decade ago. This increased availability of digital data makes decisions about which data are relevant for SWP efforts even more important.

The methodology outlined below (Table 2.1) articulates a data collection process to identify the locations of potential sources of acute contamination, identify the type(s) of chemicals stored and used on site, and research the most useful chemical attributes for drinking water management. This information can be used to make sound decisions during an emergency and to prioritize and implement source water protection activities.

**Table 2.1**
**Data Collection Steps**

| | |
|---|---|
| 1 | Delineate the Zone of Concern |
| 2 | Collect Base Spatial and Imagery Data |
| 3 | Mine Existing Data Sources for Locations of Potential Source of Contamination |
| 4 | Identify Chemical Locations and Quantities |
| 5 | Analyze Imagery to Identify Potential Sources of Contamination |
| 6 | Contact Partners and Facility Owners or Managers Directly |
| 7 | Research Contaminant Properties |
| 8 | Generate Summary Reports |
| 9 | Identify Data Gaps and Decide Whether and How to Fill Them |
| 10 | Prioritize Potential Sources of Contamination |
| 11 | Work with High Priority Facility Managers and Property Owners to Reduce Risks, Facilitate Communication and Plan Responses |

### Step 1 – Delineate the Zone of Concern

The first step in completing a source water assessment is to determine the zone of concern affecting drinking water quality. Legislation in the late 20$^{th}$ century required states to identify zones of concern and many states took a tiered approach, defining a range of zones from those nearest to the drinking water intake out to the entire drainage area upstream, and sometimes even downstream, of the intake. These regulatory zones of concern can generally be obtained from the state agency that developed them, although in some cases they were developed by consulting companies or other agencies for specific source water assessments and may not be retrievable in easy to use file formats. The zones are often described as a certain buffer distance from the edge of streams or reservoirs coupled with a specified distance or travel time upstream (e.g., a buffer of 1,000 feet along the banks of rivers upstream for a five-hour travel distance).

In some cases, the regulatory zones of concern may not fully reflect the area of interest to utility managers. For instance, many of the zones developed in the early 2000s used a hydrography dataset that did not include many of the small streams that appear in more recent, higher-resolution hydrography data. Thus, facilities near many smaller tributaries within the zone of concern may have been overlooked in past assessments. In addition, utilities may wish to be more conservative in the area they analyze than the legislatively defined buffer distances or times of travel. Finally, SWP staff are often aware of sites outside of the regulatory zone of concern that pose an elevated threat because of the activities or materials stored on site or the management of storm water or waste materials.

Several factors to consider in determining an area of concern include regulatory requirements, distance or travel time from the intake, drainage patterns, effect of stormwater infrastructure on travel times or drainage patterns and level of effort required for analysis. GIS technology is useful for analyzing these multiple factors to determine an area of concern. For many sites, a tiered approach to defining the area of concern is warranted based on the level of analytical effort required. Much of the data related to permits for facilities that may be of concern are available at the federal or state scale and thus can easily be summarized in a GIS over larger areas of interest. More detailed analyses that require additional effort include aerial image review for ASTs and other potential sources of concern and individual requests for sensitive data. These more detailed data requests and analyses should be focused on areas that are closer to the intake or those that pose elevated risks to source water quality due to the activities occurring or materials stored on site. A tiered approach may also include discontinuous areas of concern where sites located farther from the intake are known to pose more serious public health or water treatment impacts in the event of a spill or leak.

For the purposes of this method, the Zone of Concern is used to delineate the area in which a visual analysis of aerial imagery is conducted and additional data requests are made for more specific information about the location, identity and quantity of potential contaminants. This effort is in addition to compiling existing federal and state GIS data, which are collected for the whole upstream drainage area and may be summarized for any upstream area of interest (e.g., sub-watershed, zone of concern, neighborhood, etc.).

### Step 2 – Collect Base Spatial and Imagery Data

Base spatial datasets, such as state boundaries, counties, roads and hydrography, are essential features of GIS that allow users to quickly find points of interest. In addition to these traditional base layers, high resolution aerial imagery can be used to identify potential sources of contamination that are not otherwise documented. Images with a resolution better than 1 foot per pixel (e.g., 3 or 6 inches per pixel) provided the best results for visually identifying ASTs and other potential sources of contamination within the Zone of Concern. Many imagery datasets labeled “high resolution” are one foot per pixel or greater, which is not ideal for differentiating storage tanks in the 200-1,000-gallon range. Many base layers can be obtained from government or private sources, provided as on-line map services or are available for download from data warehouses. Useful sources include ESRI online data and the USGS National Map. Google, Bing and MapQuest allow users to view imagery for any area in the country. The quality and dates of images can vary widely from place to place and among these major providers. For this reason, having access to a variety of high resolution imagery taken within the last few years is helpful.

### Step 3 – Mine Existing Data Sources for Locations of Potential Sources of Contamination

Once the zone of concern is identified, the search for other relevant regulatory and contaminant data begins. Common sources of data are federal and state permitting agencies, GIS clearinghouses and data warehouses, local universities and searchable content available through the Environmental Systems Research Institute (ESRI) or other private sector sources (e.g., Google, Bing). Speaking with the managers of these data warehouses and others knowledgeable about local resources also often results in additional suggestions for sources of relevant data. Some data will be available in traditional downloadable GIS file formats but other data may be available as a real-time web services that provide the most up to date information directly from the source. Web services can be incorporated into a GIS but may limit the user’s ability to summarize and manipulate information compared to more traditional data formats. In addition, web services connect to servers in real-time and may experience connectivity issues if the server is down or overloaded. In many cases, data available through a web service may also be available for download or as a static GIS dataset. The decision to use live or cached data should take into consideration the number and performance reliability of the full set of desired web services.

The search for data should focus on information available for permitted facilities of concern, impacted resources and previously identified potential sources of contamination (e.g., from past source water assessments) (Table 2.2). For some data, there may be several different sources available for the same type of information. In these cases, using federal data can reduce the effort needed for data collection across multiple public water supplies. However, local data are often more accurate and it may be worthwhile to compare similar datasets to determine which has more appropriate data quality. This is particularly true of parcel data, where local tax parcel data may provide the most up to date contact information for property owners.

**Table 2.2**
**Commonly available datasets indicating potential sources of contamination**

| Dataset | Typical Data Source |
|---|---|
| Sites noted in previous source water assessments | Water utility, state DEP or Dept. of Health |
| Above ground and underground storage tanks | State DEP or fire marshal |
| NPDES facilities, especially those with individual permits | EPA and/or state DEP |
| RCRA sites | EPA |
| TRI sites | EPA |
| OIL sites | EPA |
| CERCLIS sites | EPA |
| Oil and gas operations | EPA and/or state DEP |
| Mining operations | State DEP |
| Waste management sites | State DEP |
| Pipelines | Energy Information Administration |
| Power plants | Energy Information Administration or state DEP |
| Refineries | Energy Information Administration |
| Airports | Federal Aviation Administration |
| Hazardous cargo transport routes | Federal Motor Safety Carrier Administration |
| Railroad stream crossings | Federal Railroad Administration |
| Road stream crossings | Federal Highway Administration |
| EPCRA Tier II reporting facilities | State or local Emergency Management Agency |
| Slurry ponds | State DEP |
| Parcel data | Town or county tax assessor |

After obtaining a dataset, considerable effort is required to investigate data quality and document the dataset (metadata). Within the zone of concern, it is worthwhile to quality assure individual data points to determine whether the location and type assigned to the data point are reasonable. In many cases, the locations of specific points are derived from a geocoding process that relies on proprietary algorithms and datasets to produce the latitude and longitude associated with a given address. This process is prone to errors that compound more standard data entry errors. It is also advisable to exercise caution when using address data because many datasets include mailing addresses in addition to or instead of street addresses or provide off-site corporate addresses rather than facility-specific locations. In many cases, permit data include a field or metadata document that describes the process by which the point was georeferenced. Metadata documentation, though not always available, can also be useful for identifying limitations of the dataset or data quality as well as any codes or shorthand used within the dataset.

The permit type can generally be verified by reviewing aerial imagery and oblique images to determine if the facility matches the permitted facility type (e.g., use Google Street View to verify that a commercial lot is an auto repair facility). Oblique (i.e., side angle) views such as Google Street View or Bing Bird's Eye can provide useful checks by revealing road or street signs and/or different angles that provide information about the height of a structure (e.g., differentiate a storage tank from a gazebo or silo).

Finally, it may be useful to limit the analysis to active or recently expired permits because many state and federal databases retain all permit actions and decisions made over many years and thus may include sites that were never completed or have ceased operation. However, some facilities with expired permits may still be operational.

### Step 4 – Identify Chemical Locations and Quantities

The locations of many facilities and land uses of potential concern are frequently publicly available in downloadable or otherwise easily accessible GIS formats. Information about specific chemicals used or stored at a particular location, however, is often more tightly controlled due to security, public health and proprietary concerns. The best sources for these data are state above- and underground storage tank databases, downstream notifications (sent by tank owners to water utilities, when required by state law), discharge monitoring reports, Tier II Emergency and Hazardous Chemical Inventory reports, EPA's Toxics Release Inventory database (TRI), individual permits and direct communication with staff located at facilities of concern (Table 2.3). In addition, general information about commonly carried hazardous materials on mobile sources, including truck, train and barge, are sometimes available.

**Table 2.3**
**Data sources for chemical locations and quantities**

| Data Source | Data Type | Critical Information |
|---|---|---|
| State DEP or Fire Marshall | AST and UST databases | Chemical name, storage capacity, installation date, inspection date and notes, secondary containment (not all of these data types are available in every state AST database) |
| PWS | Downstream Notifications | Chemical name and capacity (only available if required by state law) |
| EPA | Discharge Monitoring Reports | Chemical name, mass of past discharges, receiving water body name, permit limits |
| Local Emergency Planning Committee (LEPC) | Tier II Hazardous Chemical Inventory Reports | Chemical name, maximum and average daily mass stored on site, general location of ASTs (sometimes) |
| LEPC or Rail Company | Hazardous Commodity Flows (train) | Names of 25-50 most common hazardous chemicals carried on a specific rail line |
| LEPC | Hazardous Commodity Flows (road) | Written report on flow of hazardous goods by truck through LEPC's jurisdiction (not always available) |

(continued)

**Table 2.3 (Continued)**

| | | |
|---|---|---|
| USACE | Barge Tonnage | Barge commodity type, monthly or annual tonnage and direction of travel |
| PWS | Site Visit or Phone Interview | Chemical name, storage capacity, AST location and condition, secondary containment |

### Step 5 – Analyze Imagery to Identify Potential Sources of Contamination

Once data that will help identify potential sources of contamination and high-resolution imagery are obtained, detailed analysis of potential sources of contamination using aerial images and GIS can proceed. The goal of the image analysis process is to visually identify ASTs and other potential sources of contamination within the zone of concern that are not already indicated by existing data. Figures 2.1-2.5 provide examples of types of potential sources of contamination that can be identified with this method. In addition to potential sources of acute contamination (e.g., ASTs, publicly owned treatment works (POTWs), road crossings, etc.), sites more closely related to non-point source pollution can also be identified during the imagery analysis process, if desired (e.g., golf courses, parking lots, gas stations, etc.).

The process begins by loading the federal, state and local data into the GIS and displaying datasets indicative of potential sources of contamination over high-resolution aerial imagery with the grid displayed on top (Figures 2.6-2.7). By overlaying a numbered grid on top of the imagery and point data, a trained analyst can progress in analyzing the images and use the grid to track progress. Focusing in on a single 1 $km^2$ grid cell, an analyst reviews the known potential sources of contamination to understand the types of facilities present (Figure 2.8). Next, the analyst zooms in to the highest level and begins to scan the area for unmarked tanks and other potential sources of contamination (e.g., unmarked road crossings, waste piles, concentrated animal feeding operations (CAFOs), etc.) (Figure 2.9). If multiple high-resolution images are available, the analyst can review each image, providing a better sense of the site in multiple years, seasons and angles.

When an undocumented potential source of contamination is identified, the analyst creates a new GIS point within a GIS mapping application, identifies the type of facility and records any related information that is known, such as permit numbers, facility name, type of contaminant stored or used on site, etc. Data about the point can be edited later as new information is uncovered.

For many sites, the contents of ASTs and the identity of other materials used on-site may not be available in any public record or database available to a water utility. In these cases, a phone call or visit to the facility may yield the best results for data collection as well as building communication channels and coordinating emergency response efforts. A site visit may also be helpful for complex industrial or commercial properties with multiple permits and processes to understand the overall operations and potential associated risks. There may also be local or regional agencies that collect water quality information or other useful data that may not be generally publicly available but could be provided on request for public health and emergency response planning purposes. This step was beyond the scope of the pilot site work completed as part of this project.


*Source*: Imagery © 2016 Google, Map data © 2016 Google

**Figure 2.1 Potential source of contamination: confined animal feeding operation**


*Source*: Imagery © 2016 Google, Map data © 2016 Google

**Figure 2.2 Potential source of contamination: fish hatchery**


*Source*: Imagery © 2016 Google, Map data © 2016 Google

**Figure 2.3 Potential source of contamination: fly ash pond**

*Source*: Imagery © 2016 Google, Map data © 2016 Google

**Figure 2.4 Potential source of contamination: oil or gas well.** (Note the nearby above ground storage tank with earthen berm for secondary containment).


*Source*: Imagery © 2016 Google, Map data © 2016 Google

**Figure 2.5 Potential source of contamination: composting facility**

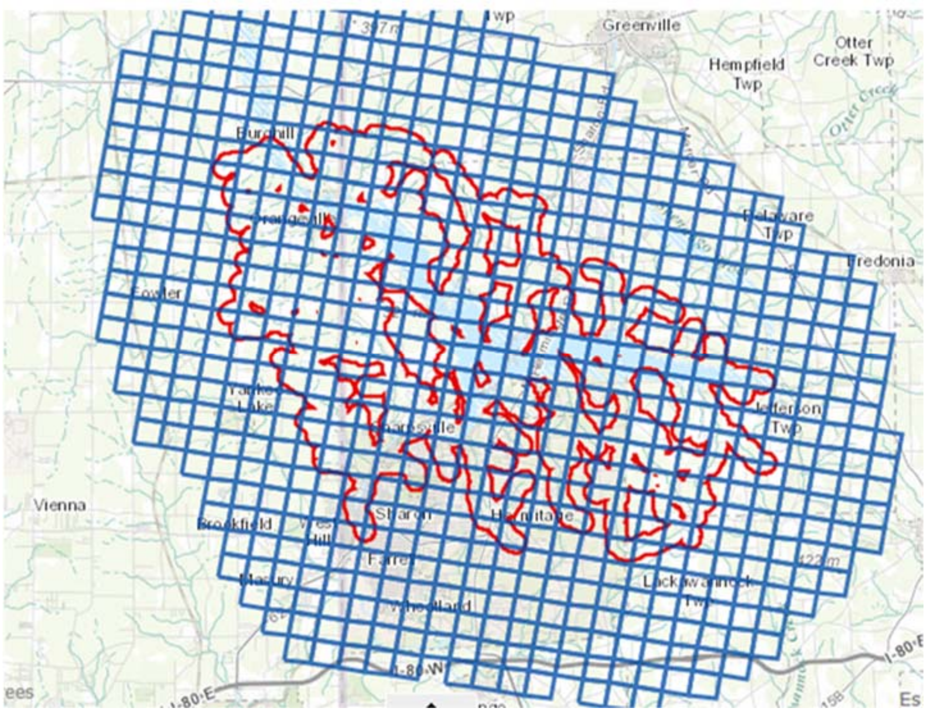


**Figure 2.6 Image Analysis - zone of concern (red lines) with grid cell overlay (blue lines)**

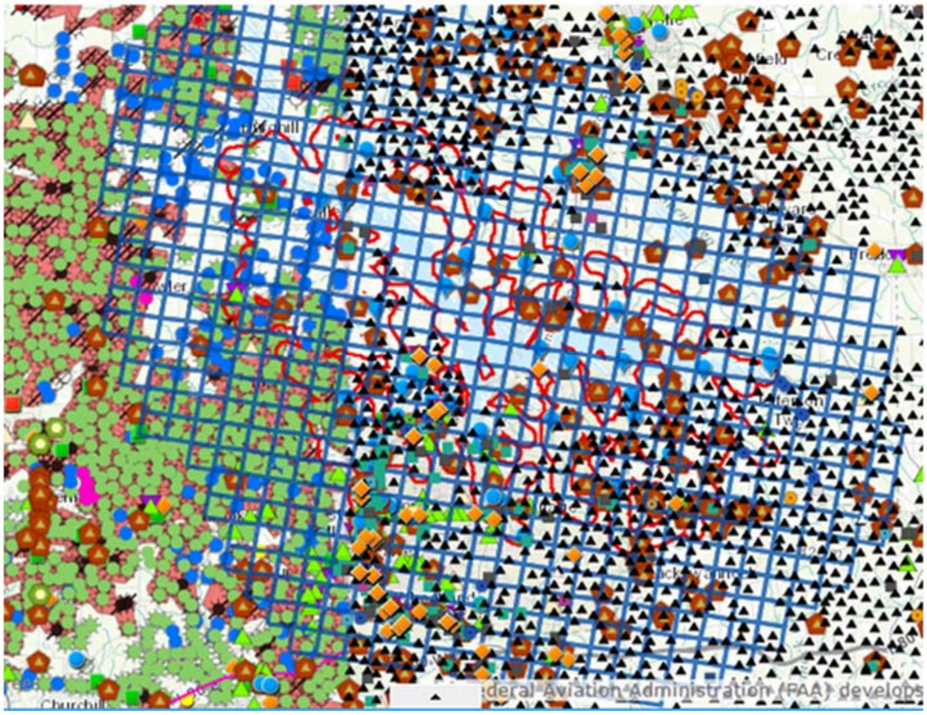


**Figure 2.7 Image analysis - zone of concern, grid cell overlay and datasets indicating potential sources of contamination.** Note that this Zone of Concern crosses state boundaries. Federal datasets have complete coverage for this area but state different state datasets are available in each state.

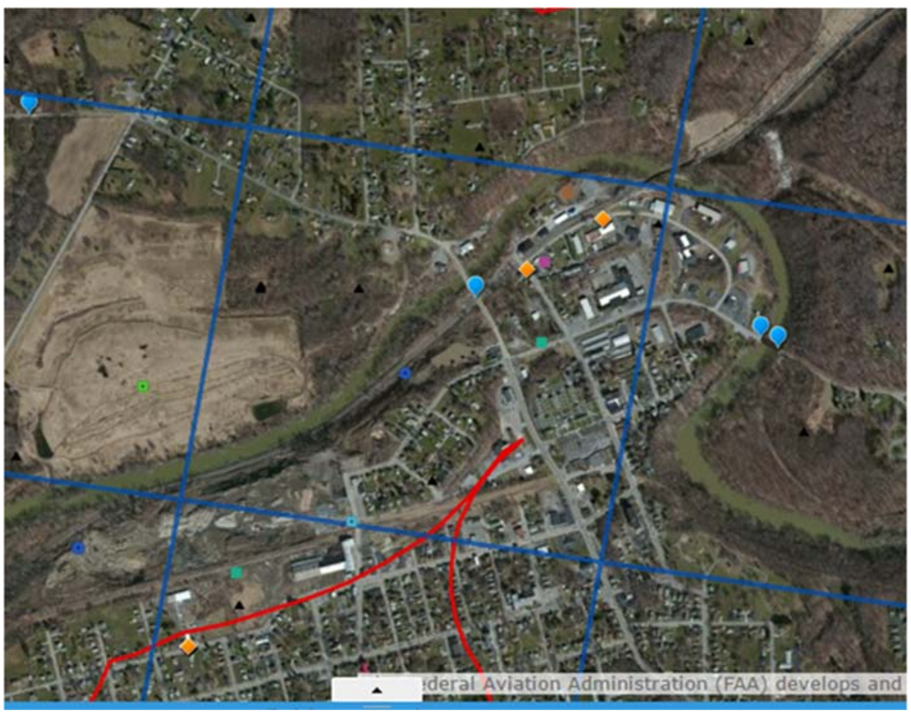


**Figure 2.8 Single grid cell with zone of concern (red line), aerial imagery, Federal and State data**

**Figure 2.9 Grid cell with data, zoomed in to analysis scale**

### Step 6 – Contact Partners and Facility Owners or Managers Directly

For many sites, the contents of ASTs and the identity of other materials used on-site may not be available in any public record or database available to a water utility. In these cases, a phone call or visit to the facility may yield the best results for data collection as well as building communication channels and coordinating emergency response efforts. A site visit may also be helpful for complex industrial or commercial properties with multiple permits and processes to understand the overall operations and potential associated risks. There may also be local or regional agencies that collect water quality information or other useful data that may not be generally publicly available but could be provided on request for public health and emergency response planning purposes. This step was beyond the scope of the pilot site work completed as part of this project.

### Step 7 – Research Contaminant Properties

Contaminants known to be present within a zone of concern represent potential threats to water quality. Building a georeferenced database of chemical properties and locations within the drainage area will help utilities prioritize the sites and contaminants that pose the greatest concern and manage chemical information in the event of an emergency. For some facilities, it may not be possible to positively identify specific contaminants used on site, but a general category of possible contaminants may be available. For example, railroad commodity reports provide general information about the types of chemicals hauled on specific lines but the exact contents of each car are not available except in an emergency situation. Similarly, the exact chemicals used during hydraulic fracturing are often undisclosed, but various organizations have compiled lists of chemicals used by this industry in general. Thus, it may be useful to include some contaminants in the contaminant database even if their specific locations are not known.

The data collection process for populating the contaminants database entails (1) collection and documentation of contaminant property data from data sources (Table 2.4), (2) search of supplemental data sources for missing data to fill high priority data gaps and (3) quality assurance and quality control. A specific list of properties (Table 2.4) for collection for each contaminant was chosen in consideration of the data required in a spill/event response and to facilitate comparison of contaminants in susceptibility analyses. A standard operating procedure (SOP) for contaminant data collection ensures consistency among analysts collecting chemical data. The SOP for contaminant data collection developed for this project is provided in Appendix E.

**Table 2.4**
**Chemical properties and data sources**

| **Chemical Property** | **Common Data Source** |
|---|---|
| | *Identification and Use* |
| CAS | Chemical abstract service |
| Name | Chemical abstract service |
| Alternative Names | TOXNET (HSDB), CAMEO |
| | |
| Purpose | TOXNET(HSDB) |
| Manufacturers | TOXNET(HSDB) |
| | |
| | *Regulations* |
| Maximum Contaminant Level (MCL) | EPA 2012 Edition of the Drinking Water Standards and Health Advisories |
| Secondary Drinking Water Regulations (SDWR) | EPA 2012 Edition of the Drinking Water Standards and Health Advisories |
| Unregulated Contaminant Monitoring Rules (UCMR 1, 2, 3) | EPA Unregulated Contaminant Monitoring Rule |
| Contaminant Candidate Lists (CCL 1, 2, 3, 4) | EPA Contaminant Candidate List (CCL) and Regulatory Determination |
| Six Year Review | EPA Six-Year Review of Drinking Water Standards |
| Chemical Facility Anti-Terrorism Standards (CFATS) | Department of Homeland Security 6 CFR Part 27 Appendix to Chemical Facility Anti-Terrorism Standards; Final Rule 2007 |
| List of Lists | Consolidated List of Chemicals Subject to EPCRA, CERCLA, and Section 112(r) of the Clean Air Act |

(continued)

**Table 2.4 (Continued)**

| | |
|---|---|
| *Reactivity, Detection and Treatability* | |
| Decay constant | ad hoc |
| Flammability | CAMEO |
| Reactivity | CAMEO |
| Reaction products | ad hoc |
| Method detection limit (MDL) | ad hoc |
| Treatability | US EPA Drinking water treatability database |
| | |
| *Physical and Chemical Properties* | |
| Color | TOXNET (HSDB), CAMEO |
| Odor | TOXNET (HSDB), CAMEO |
| Taste | TOXNET (HSDB), CAMEO |
| Specific gravity | TOXNET (HSDB), CAMEO |
| Solubility | TOXNET (HSDB), CAMEO |
| Vapor pressure | TOXNET (HSDB), TOXNET (HAZMAP), CAMEO |
| Henry's Law Constant, $K_{ow}$ and other partitioning data | TOXNET(HSDB) |
| | |
| *Health Effects and Medical Treatment* | |
| Health effects | TOXNET (HSDB), CAMEO |
| Protective equipment | TOXNET (HSDB), CAMEO |
| Medical treatment | TOXNET (HSDB), CAMEO |
| LD50 | TOXNET(ChemIDPlus), TOXNET(HSDB) |
| No Observed Adverse Effect Limit (NOAEL) | IRIS, Agency for toxic substances and disease registry (ATSDR) |
| Reference dose (RfD) | IRIS, Agency for toxic substances and disease registry (ATSDR) |

The SOP discusses data collection for pure contaminants and recommended methods for data collection and characterization of mixtures. Mixtures do not generally have their own unique Chemical Abstract Service (CAS) numbers but may have unique properties that are unlike their individual components. In general, information about each component of a mixture should be included in the contaminant database as well as any data specific to the mixture itself. Proprietary contaminants and mixture proportions should also be documented in the contaminant database. Including proprietary substances in a database necessitates additional security measures for avoiding release of sensitive data.

After all primary and secondary data sources are mined some data gaps will remain. Because the process of reviewing the primary literature to fill these data gaps is time-consuming and must be done by analysts with a strong understanding of the subject matter, the drinking water community should determine which gaps are most critical and work together to complete the research needed to fill these gaps. Developing data for parameters such as treatability, in particular, requires significant effort. A systematic literature search must be conducted, removal data must be extracted from studies and verified (quality controlled), data must be qualified (e.g., bench scale data should be distinguished from data collected in full scale studies), and some level of meta-analysis must be conducted to abstract treatability estimates from the reviewed studies. Alternatively, treatability studies conducted at a specific plant could be used instead of data drawn from published studies.

**Step 8 – Generate Summary Reports**

After the potential sources of contamination are identified and the contaminant research has been completed, summary reporting can be generated to begin the process of prioritizing follow-up efforts, emergency planning and response. The following types of reports may be useful:

1. Potential sources of contamination with type of facility, name and address
2. Known AST chemical contents and capacities
3. Toxicity, detection and treatability information by chemical
4. Emergency contacts
5. Facility permits and regulatory reports
6. Facility compliance history

**Step 9 - Identify Data Gaps and Decide Whether and How to Fill Them**

One consideration in using summary data to generate source water protection priorities is the prevalence of data gaps for key pieces of information, such as toxicity, detection and treatability. The precautionary principle suggests that these data gaps should result in a high priority for follow-up efforts to obtain this information based on the possibility that a contaminant may be highly toxic, difficult to detect and/or untreatable. However, the effort needed to fill these data gaps is likely to exceed the expertise and resources of a drinking water utility and may require partnerships with federal or state agencies or academic institutions.

The summary reports will also reveal data gaps that can be filled within the expertise and available resources of a drinking water utility. For instance, data gaps in AST contents, the locations of specific contaminants and emergency contact information should be carefully reviewed to determine whether they are high priority for further research or outreach. The data gaps analysis for the pilot sites is discussed in Chapter 3.

**Step 10 – Prioritize Potential Sources of Contamination**

Source water assessments completed in the early 2000s relied on a variety of methodologies to determine the susceptibility of an intake to contamination. These ranged from single parameter assessments (e.g., time of travel), to assessments of the zone of concern’s setting (e.g., soils and slope), to methods that addressed the magnitude and likelihood of impact from individual sources. These efforts generally relied on qualitative approaches due to the lack of resources and data to develop more focused risk assessments.

The data collection effort described in steps 1-9 above generally exceeds that of most past source water assessments in that it integrates more detailed information about specific chemicals and their physical and toxicological properties with geospatial information. Even with this expanded dataset, however, the data needed for a full risk ranking are beyond those collected in this project.

The Lessons Learned section of this report describes the data required for a full, quantitative risk assessment, the elements of those data that were collected in this effort, and approaches that can be taken with the information collected (short of full risk assessments) to rank points of interest with respect to their potential impacts on drinking water operations. A qualitative index based on the results developed in answer to these questions may be created and weighted as desired to help develop source water protection action plans based on the needs and interests of the local utility.

### Step 11 – Work with High Priority Facility Managers and Property Owners to Reduce Risks, Facilitate Communication and Plan Responses

The final step in SWP is implementing priority actions to prepare for spills and coordinating with high risk facilities. This step is not part of the current project but is also a necessary component of effective SWP and emergency response programs. Site visits are an invaluable tool that can provide both accurate data about activities and materials at a particular site as well as the opportunity to strengthen relationships and share resources and experiences. The methodology described above can provide water utility staff with important information to help plan site visits, such as which permits a facility may have, what is known or not known about the type and quantity of chemicals stored on site or transported to and from the site, the age of ASTs, the site's compliance history, etc. A review of this information will help generate a list of questions to ask during a site visit and identify important data gaps. Discussion during the site visit should aim to answer these questions as well as provide a foundation for continuing communication between the appropriate staff at the drinking water utility and facility in question. For example, water utility staff may wish to share emergency contact information, spill response plans or information about how the particular chemicals stored at the facility could affect the delivery of safe water in the event of a spill. Facility managers may be willing to provided MSDSs, Spill Response Plans, Spill Prevention and Counter Control (SPCC) plans, notification contact lists, or other resources with drinking water staff. To facilitate regular communication, utility staff may wish to suggest an annual check-in to update emergency contact information and share any new emergency response information or changes in activities at the facility.

For facilities identified as high priority potential sources of contamination, it may be worthwhile to recommend joining the Local Emergency Planning Committee to ensure regular communication and joint planning for emergency situations and/or to organize tabletop or on-site spill response exercises as an effective tool for identifying available resources, improving communication processes, planning priorities and updating response strategies.

## LESSONS LEARNED

The method outlined above describes a general data collection process to identify potential sources of contamination and link known location of specific contaminants to a database of their physical, chemical and toxicological properties. During the application of this method to the five pilot sites assessed for this study, a number of useful datasets and helpful hints emerged. Many of these tips are relevant to particular datasets or data sources that were available during the timeframe of this study (2014). Given the pace at which GIS data and display technologies are evolving, some of these observations and methods may become obsolete as technology and data management practices evolve.

### Identifying the Locations of Specific Contaminants

A relatively small number of datasets were identified that provided the specific identity and location of contaminants. The most useful of these are discussed in more detail below.

In some states, one of the best sources of information was the state's AST database. However, one of the major impediments to obtaining this information was the regulatory structure for collecting data on the contents of ASTs. At the federal level, only hazardous chemicals are regulated. At the state level, some states have comprehensive programs to regulate ASTs and collect information about their contents, inspection history, etc., while other states have no regulatory authority at all over ASTs. For this reason, state AST databases vary widely in the types of ASTs that are included as well as the information that is collected and made public. Each state with a regulatory requirement to register or inspect ASTs has its own definition of which ASTs must comply, often based on a minimum size threshold and tank content. State agencies may also limit the type of information that is made available publicly (e.g., AST contents or capacities) for confidentiality and/or security purposes. In some cases, drinking water utilities may obtain additional information, such as the identity of hazardous liquids, on request. Often, these data must then be merged with available AST locations to be useful in a GIS.

In some states, owners of ASTs meeting certain requirements must notify downstream water users, including drinking water utilities, of the identity and quantity of hazardous chemicals stored on site. Regulations regarding which tank owners must send notifications and how often vary from state to state and some states do not have any requirement for direct notification of drinking water utilities.

Facilities with NPDES permitted discharges submit discharge monitoring reports to EPA or the designated state agency. EPA compiles this information into its online Discharge Monitoring Report Pollutant Loading tool. This tool provides pounds and toxic weighted pounds of pollutants discharged as well as the name of the receiving water body. The information is not provided directly as geospatial data but may be associated with the location of the permitted facility using the Facility Registry System (FRS) ID number for the facility or requested from EPA with latitude and longitude fields. EPA also makes facility locations available through several different GIS data export options.

The Emergency Planning and Community Right-to-Know Act of 1986 (EPCRA, a.k.a. SARA Title III) required EPA to create a chemical release inventory, among other planning and emergency response requirements for chemical hazards. The Act established State Emergency Response Commissions (SERC) and Local Emergency Planning Committees (LEPC). LEPCs were charged with developing Emergency Response Plans for federally designated Extremely Hazardous substances present above the threshold planning quantity. Both the SERC and LEPC receive copies of MSDS sheets and Tier II Hazardous Chemical Inventory Reports from each facility that stores designated hazardous chemicals above the threshold planning quantity. Tier II reports are submitted annually in the spring and may be available as soon as they are received by the LEPC. In some cases, LEPCs incorporate the information into a database or GIS, and these data may take longer for the LEPC to process. Tier II data may be requested from either the SERC or LEPC. A sample data request letter to LEPCs for Tier II and rail, road and pipeline commodity data is provided in Appendix G.

During the pilot studies, responses to requests to LEPCs for Tier II hazardous chemical storage information generally fell into one of three categories: 1) no response, 2) requests must be made for specific facilities and cannot be fulfilled for a geographic area (such as a zone of concern or county), or 3) information was provided. The names of specific chemicals claimed as trade secrets are withheld from public disclosure in this process. Drinking water utilities may become members of the local LEPC to gain more immediate access to this information as it is updated as well as opportunities to help guide local chemical emergency planning and response activities. Data from the LEPCs were provided in a variety of formats, ranging from pdfs of paper applications to spreadsheets compiled by the LEPC to CAMEOfm files, a mini-relational database developed by the US EPA. A significant amount of data processing was required to combined these data into a single useful GIS data layer. Another challenge in obtaining the Tier II reports was that the contact information for LEPCs, especially for the smaller, volunteer organizations, was frequently out of date and several calls to local EMAs and/or the SERCs were required to identify the correct point of contact.

EPCRA also requires facilities to report toxic releases to the EPA. These data are available through the EPA's online Toxics Release Inventory (TRI) database, which provides the chemical name, quantity released and the medium to which it was released (e.g., air, water, etc.). For releases to water, the name of the stream or water body is also provided. Transfers of chemicals to off-site locations are also documented.

In some cases, information about the location of specific chemicals must be requested through a public information request (Freedom of Information Act, FOIA) and/or file review at the office of a state agency, for which a nominal fee may be charged. It is advisable to contact the agency that manages the data directly prior to submitting a FIOA request or conducting a file review to ascertain that the data exist and determine how the request should be made. FOIA requests should be as specific as possible and it may take several iterations and communications with the agency to obtain the correct information in a useful format. For several of the pilot studies, special processes were developed in cooperation with state agencies to make non-public data available to water utilities through various combinations of FOIA requests, files reviews and non-disclosure agreements.

Some useful data may only be available at greater than nominal cost. The most common example of this situation was parcel data with contact information for landowners. These data are highly useful in an emergency situation, as they provide a quick contact for property owners when an emergency situation is reported. These data are generally collected and maintained by town or county tax assessors and their availability varies widely from freely available on a GIS website to available for a large fee in a variety of formats. Each utility will need to evaluate the costs and benefits of obtaining this type of data for their source water protection and emergency response programs.

Information about chemicals carried on mobile sources was difficult to find and manage but was available from a few sources. The US Army Corps of Engineers (USACE) maintains annual and monthly databases of tons of commodities, by type, carried by water past US inland ports and along US waterways. The US Army Corps of Engineers (USACE) Navigation Data Center waterway (barge) traffic database was most recently updated in 2014, to cover calendar year 2012, and it is not evident when it will again be updated. USACE also tracks vessel movement through locks on 38 major rivers in real time (15 minute updates) and provides reports with the vessel name, ID number, direction, date and time passing the locks, and number of barges in tow, but not the identity of materials on board. Both of these USACE data sources are available online.

To develop the Emergency Response Plans, many LEPCs collect information about the flow of hazardous materials on rail lines and roads. Rail companies are required to track the predominant hazardous materials carried over specific sections of their track and must make this information available to emergency response agencies. The information is often provided as a table of the top 25-50 materials hauled over the previous 12-month period. These data may be referred to as Commodity Flow, Traffic Flow or Density Study Reports. Because railroad companies will generally only distribute commodity information to emergency response agencies, local LEPC or Emergency Management Agencies (EMAs) must be enlisted to request those data.

In addition, local agencies such as the LEPCs may conduct commodity flow studies that include surveys of hazardous materials carried over major road routes, as determined by observations of hazardous materials placards on vehicles. These data are often only available in report format and may require data entry and processing time to incorporate into a GIS.

In general, information about specific contaminants traveling on road, rail and river is not provided in GIS format and only includes aggregated information about the most common commodities carried. The contents of a specific truck, rail car or barge are generally known to the carrier company but are only provided through formal response protocols during emergency situations. Often, these protocols do not directly include notifying downstream drinking water utilities.

## Incorporating Data into A GIS

GIS technology serves as the core of the information system approach described in this methodology. GIS is a mature technology that has been in use for over 30 years and it continues to evolve and improve. GIS is used by many government and private sector organizations to visualize and analyze complex data from multiple sources. Two major improvements have emerged in the past 5 years that are particularly relevant to SWP activities. These are the ability to publish map services over the Internet and the use of cloud based data storage and emerging Software as service capabilities. These technologies are improving the ease of sharing spatial data across organizations and fueling a tremendous increase in the use of GIS by non-GIS professionals and casual users.

The GIS that has been developed for this project is based on open source technology components, has a very intuitive user interface and can be delivered as Software as a Service on a subscription basis. GIS software is available through several private firms (the most commonly used software is offered by ESRI) as commercial, off the shelf products.

To use these software tools, a greater level of skill and expertise is required to develop a GIS tool that incorporates the information as described in this document than to use the information in a tool that has already been created. Thus, many of the data collection, documentation and management challenges may be best addressed by staff with specialized GIS training and experience. Less experienced staff can generally be trained relatively easily to access the information they need once a GIS tool is set up.

### *Data Storage*

Any effort to collect a large amount of data from multiple sources and keep it up to date requires a coordinated and sustained data management effort. Data managers should first consider the volume of data that will be obtained and any hardware or software changes that are needed to handle it. The format of the data will affect the data volume to some degree. For instance, connections to live web services will not require large amounts of storage space, while image files will generally be much larger than their point, line and polygon counterparts. Web services may have limited analysis and symbology options. To overcome these limitations, original data files may be requested from the data owner or special scripts may be written to obtain the data from the web service. Consideration should also be given to whether there is a need to normalize data from multiple sources into a single dataset (e.g., Tier II data received as pdfs, spreadsheets and CAMEOfm relational databases).

***Data Dictionary***

The next consideration is the importance of developing a data dictionary. Essential elements of a data dictionary are known as metadata (i.e., data describing the data) and include the data source, format, date, field names, update schedule and contact information for the data owner. Many datasets contain field names and attributes that are recorded in code or shorthand to facilitate data entry or to meet other internal database requirements of the source agency. These codes can hinder the user's understanding of the information and can cause incorrect conclusions to be drawn; thus, it is important to provide common language translations for obtuse field names, attribute codes and acronyms in the data dictionary. Ideally, these codes should be integrated into the data in the GIS so that users do not need to refer back to a separate data dictionary for interpretation. Any confidentiality requirements for each dataset should also be carefully tracked and managed so that only users with the proper permissions have access to each dataset. Finally, any data processing or other steps that could alter the data from its original condition should be recorded.

***Data Quality***

Managing, integrating and normalizing datasets from multiple organizations that relate to similar information can be a challenge because the information available and database structures are not consistent. For example, road and street centerline data may be available from federal agencies, multiple state departments and private data companies. Each of these may have different levels of accuracy and attributes that add value to the spatial data.

Some data may arrive only partially useful. It may have useful fields but lack geospatial locations or include geospatial data but lack critical attributes found in another file or database and thus require data processing. The most common types of data processing involve merging files together based on a common identifying field or geocoding addresses to create coordinate points from which to map the data in a GIS. Both types of manipulations are prone to introducing errors into the data and results should be carefully reviewed before any source water assessment analysis is completed. For example, merging files requires a common field that identifies unique points to be mapped, such as a facility or site ID. These fields vary from agency to agency and often even within the same agency.

Geocoding can also introduce errors into the dataset due to the inherent inaccuracy of address data. Results from different geocoding processors can produce different results and the spatial accuracy may not match the necessary scale for analysis (i.e., the mapped address coordinates may be so far away from a facility visible in an aerial image that the point is not visible when zoomed in to the scale of analysis). Geocoders will also fail to locate some addresses, causing the data to be unrepresented in the GIS. Thus, these unmappable data cannot be included in any geospatial analysis.

### *Symbology*

Choosing a scheme for symbolizing the data can also be challenging. Some datasets may be broken down into a set of individual layers by the original data owner. Alternatively, a single data layer may be better represented as several layers if multiple pieces of information are important. Ultimately, the decision of how to symbolize the data will depend on the fields of interest to the user and the need to visually communicate those data.

Mobile sources of contamination pose a special challenge in GIS. Data sources indicating the most commonly carried chemicals on road, rail or river are often associated with a specific road, rail line or river segment. These data can be symbolized using linear segments for roads, rail lines and river reaches; however, a decision must be made about how far to extend the association. For example, if a particular chemical is known to be carried on barges on a particular river, is it likely to travel along the whole length of the river or just the lower reaches closest to its confluence with a larger river? Occasionally, observational data for mobile sources of contamination are available. These data may indicate that a particular type of chemical travelled past a specific road location, port, etc. Options for representing this information include associating the information with the nearest linear reach (e.g., road, rail or river segment) or creating points that indicate the actual location of the observations. Users should consider the ultimate users of the data to determine how best to symbolize these challenging potential sources of contamination. For example, is the information likely to be overlooked if the observation point is located outside the drainage area while the road, rail or river passes through it? Conversely, are false assumptions likely to be made if a particular chemical is associated with a road, rail or river segment that it is not commonly transported on?

### *Keeping Data Updated*

Updates to existing datasets present several challenges. First, an update schedule must be established based on a combination of the update schedule at the source and the needs of the users. Next, the data must be downloaded or requested from the source and processed the same way as the original dataset. Finally, differences between the old and new data should be evaluated. These may include edits to existing data, new records added or old records deleted. In some cases, the structure of the database or naming conventions for field attributes may have changed, requiring additional processing to determine which underlying data were affected. Finally, data managers must evaluate whether any changes have occurred that are significant and should be reviewed by the source water protection specialist.

Once all of the data are loaded into a GIS, systems must be put in place to manage new data, data edits and updates. For instance, it should be possible to record the results of file reviews, site visits or phone calls to facilities of interest in the GIS by making notes or edits to the existing data as information is collected. This capability requires the GIS to be able to track the date and user who made the change along with the edited data. Similar requirements exist for quality assurance checks, where the analyst, date and result of the quality check must be recorded as data are reviewed.

Data gaps may be filled by integrating additional datasets, updating existing datasets, or creating new data, such as through the aerial image analysis described above. Automated image analysis is also expected to become an important tool for filling data gaps by identifying storage tanks from satellite or aerial imagery. Significant challenges exist such as variability in color, size and shape of storage tanks; visual obstruction due to cloud or man-made cover; and the cost of acquiring and processing large amounts of imagery. Computer vision, image analysis and object recognition fields have advanced rapidly over the past decade and are expected to continue to do so. Similarly, costs are coming down and it is inevitable and possibly imminent for automated image analysis to start playing an important role in identifying storage tanks.

Other computational methods also exist and may hold promise: Light Detection and Ranging (LIDAR), remote sensing and analysis of derivative image products. LIDAR data can be used to reconstruct a 3-D model of physical structures, which could allow the identification of tank structures based on form. Remote sensing includes the use of non-visual spectrum sensors to extract features. While the resolution and specificity of known sensors in use today are unlikely to result in the identification of storage tanks (based on tank construction material, paint or liquid mass), remote sensing as a field is progressing rapidly. Derivative image products are imagery that is derived from other products, frequently with the integration of multiple channels and/or additional data sources using sophisticated algorithms. For example, the so called "Roads" imagery used as base maps by map providers is starting to include 3-D structures often derived from LIDAR data. The resulting derived imagery is much easier to analyze in an automated fashion. However, the availability of this type of data is largely limited to select urban areas at this time. These technologies may play a larger role in identifying undocumented sources of potential contamination in the near future.

### Addressing the Prevalence of Petroleum Products

Based on the pilot site studies and other related work by CEC, it is apparent that petroleum product storage in zones of concern is ubiquitous. Because most intakes could be impacted by petroleum products, one method for characterizing their chemical properties is provided in detail here.

Petroleum products are typically mixtures, and many data sources lack specificity in their descriptions of the chemical compositions. Because many of the source datasets do not use CAS numbers to identify chemicals, petroleum products are labeled following nomenclature rules that are often specific to a particular agency and dataset. Petroleum is crude black oil which is distilled into categories of refined oils, such as naphtha, gasoline, kerosene and diesel. Merely categorizing oil as a "petroleum product" is not a specific classification, as all oils, fuels and lubricants are refined from petroleum. Because the chemical properties of these materials vary greatly over this spectrum it is necessary to classify them further for the purposes of drinking water monitoring, prioritization and response. Refined oils are distilled from petroleum according to the lengths of hydrocarbon chains in the material. For example, hydrocarbon chains in naptha, a light oil with a low boiling point, are shorter than those the heavier gasoline. The difference in chain-length leads to differences in many of the chemical properties of these refined oils. These oils can thus be classified based on hydrocarbon chain-length. The relative major categories of refined petroleum products in order of increasing hydrocarbon chain-length are as follows: petroleum gas, naptha, gasoline, kerosene and diesel. Chemical and physical properties of these materials vary little within a category, but can vary substantially between categories. The petroleum data specificity gap can be partially addressed by developing chemical profiles for these major categories. One approach for characterizing petroleum products is to identify the typical components of a particular petroleum mixture (e.g., fuel oil) and the ranges of the mass fractions for each of the components.

Once the constituents of a petroleum mixture are known, property data for distinct constituents can be queried (e.g., from the database assembled in this project) and used to develop spill responses.

**Prioritization**

A full quantitative risk assessment (an estimate of the likelihood of specific adverse public health impacts) for a contaminant stored upstream of an intake requires data and exposure modeling that are beyond the scope of the current effort. Some of the data required to conduct a risk assessment are:

- The identity of a contaminant
- The volume of the contaminant stored
- The integrity and last inspection date of the tank or structure housing the contaminant and of the secondary containment, if present
- Distance from the containment to receiving waters
- Nature of the flow path between the containment and potential receiving waters
- The fate and transport properties for the contaminant in the waters receiving a potential spill
- Stream flow in the river network between the spill location and the intake
- Removal of the contaminant during treatment (including removal via spill response measures such as increased powdered activated carbon (PAC) dosing)
- Toxicity data for the contaminant for susceptible and non-susceptible populations

Some, but not all, of the data listed above have been collected and are housed in the contaminants database prepared in the course of this research project. In the absence of those data and detailed fate and transport models, a comprehensive, quantitative risk assessment for all the contaminants identified in this effort is not possible. However, sufficient data have been collected to allow users to conduct more comprehensive qualitative risk assessments or risk rankings. Further, if a utility were particularly concerned about a specific point of interest, additional data on the contaminant housed at the point of interest could be added to the available data and used in a quantitative risk assessment.

Following the methodologies described in this report, users will be able to compile a database that includes the locations, identities and volumes or masses of potential contaminants coupled with toxicity, detection, health effects and treatability information. A time of travel and contaminant concentration model, such as the one described in Appendix B or a more site-specific model, can be used to incorporate travel times and expected peak concentrations at the intake into a prioritization plan. Collectively, these data can be used to answer a common set of questions that utility managers may ask to quantify the magnitude and likelihood of a contamination event from a potential source of contamination.

***Application of the Methods and Data for Spill Preparation and Spill Response***

Utilities are likely to use the methods and data described in this report for both spill risk management and spill response. The most important spill-related questions they may ask and the methods and data to answer them are presented below. If desired, answers to these questions may be combined in a qualitative index weighted to make use of the most important and complete data for developing priorities for utility Source Water Protection programs.

***1. What is the likelihood that a potential source of contamination will become a real source of contamination?*** Sources that currently discharge pollutants or are known to have discharged pollutants in the past have already become real sources of contamination. Information about these sites can be found in existing databases such as EPA's database of discharge monitoring (DMR) reports, NPDES compliance data, the TRI database and AST inspections. The National Response Center also maintains downloadable historical information about the locations and types of contaminants spilled into surface waters. For sites identified as potential sources of contamination due to their large AST volumes, several pieces of information may be useful for estimating the likelihood of a spill event: AST installation dates, inspection reports and secondary containment data from the state's AST database; file reviews, site visits or phone calls; and the number of ASTs containing a particular contaminant.

Other researchers have related spill likelihood to the historic frequency of spills of a particular contaminant on a given waterway (Cao, Lia, Joksimovica, Yuana, & Banting, 2013), average accident rates for vehicles transporting contaminants (Foster & McDonald, 2000) or watershed characteristics and contaminant source locations (for non-point loading of *Cryptosporidium* during rain events (Foster & McDonald, 2000)). Alternatively, in lieu of hard data on the likelihood of spills, some researchers have used expert judgment to develop scenarios such as diesel fuel spills of different volumes and at different locations to explore spill response and risk management (Martin, LeBoeuf, Danieal, Dobbins, & Abkowitz, 2004; Camp, LeBoeuf, & Abkowitz, 2010).

***2. If there is a spill at a specific potential contaminant source, what will be the peak concentration of the contaminant at the intake?*** The travel time and concentration model developed by the USGS and described in Appendix B uses the distance between the point of concern and intake, the maximum mass of the spill (which for ASTs can be derived from their storage capacities), the decay rate of the contaminant, the drainage areas of the intake and spill site and the stream flow rate to determine the time of travel and concentration at the intake for the leading edge, peak concentration and trailing edge of the contaminant plume. This national model assumes that the contaminant mixes thoroughly throughout the water column. More site-specific time of travel models based on dye tests or river morphology may also be used to calculate time of travel and concentrations at the intake, if available. Several data collected or calculated during pilot studies for this project and related to prediction of peak spill contaminant concentration and time of arrival at the intake are highlighted below.

*a. What is the maximum storage capacity for the contaminant?* To calculate the peak concentration at the intake, knowledge of the maximum volume that may be spilled is needed. These data are available in a variety of data sources. State AST databases and downstream notifications generally include AST capacities in gallons. Tier II reports provide maximum and average daily masses stored on site. EPA's TRI database and DMRs report the mass of pollutant released. Site visits, phone calls and/or file reviews are also good sources of this type of detailed information.

*b. How far is the contaminant storage site from the intake?* To calculate travel time and thus peak concentration, it is necessary to know the distance along the stream from the point of interest to the intake. The time of travel model under development during this project calculates this distance using geospatial tools that rely on the coordinates of both points and the hydrologic network between them. Overland travel time is currently omitted from the overall time of travel calculation. For contaminants spilled directly into streams or into storm water collection connected directly to streams, that omission is not critical. For sites separated from streams by buffers, swales, detention basins and other features, that omission is significant. Adding overland travel time will require a detailed analysis of the terrain between a potential contaminant source and a stream. It is likely that highly-accurate overland travel time estimation is only merited for sites of high concern to utilities, while general estimates can be made based on existing topographic data.

*c. What is the fate of a contaminant in receiving water?* The fate of a contaminant as it is transported downstream integrates a variety of physical and chemical processes and reactions that are difficult to accurately quantify. A partial list of pertinent processes includes biotic and abiotic geochemical reactions, sediment interactions, desorption (at the gas-water interface), mixing, decomposition and reaction with other materials in the water matrix. Although a full fate and transport model is beyond the scope of the current effort, chemical attributes gathered in the course of this project such as the decay constant, reactivity, vapor pressure and Henry's Law constant can offer insights into the general reactivity and volatility of the contaminant.

Specific gravity is a particularly important indicator of the susceptibility of intakes. Contaminants that preferentially float or sink will result in different levels of concern for utility managers depending on the vertical location of the intake within the water column. The specific gravity and solubility of a contaminant can be used to understand where within the vertical water column the peak concentration may occur and can inform to a decision to close an intake.

**3. *How will available treatment processes affect the concentration of the contaminant at the tap?*** Treatability (the degree to which unit operations remove contaminants) can vary widely among operations due to differences in source water characteristics, operation design and differences among operators. Further, removal is expected to be different for bench scale studies, pilot scale studies and full scale operations. Treatability data collected by the EPA and housed in their treatability database (EPA, 2015) are summarized as ranges of removal for different contaminants by various processes. All of the data housed in the EPA treatability database are accompanied by detailed data describing the studies the data were drawn from. Review of those underlying data is a critical element of understanding removal of a particular contaminant and estimating removal at a particular operation. In general, the treatability data do not address site-specific features such as mixing/equalization/hydraulics or operator choices such as changes in coagulant dose or adjustment of pH in response to an event.

***4. What are the potential human health impacts of the contaminant?*** For contaminants with MCLs or other drinking water advisories, the determination of human health impacts has already been made and these levels can be added to the contaminant database for reference. For other contaminants, the toxicity of a potential contaminant can be characterized using the $LD_{50}$, no observed adverse effect level (NOAEL) and reference dose (RfD). All exposure routes and population segments should be considered when evaluating health impacts of a potential spill. In all cases, oral ingestion of contaminants is a concern. For contaminants that are particularly volatile or associated with human health impacts when inhaled, inhalation exposure (and its associated endpoints and toxicity data) should also be considered. In some cases, dermal exposure should also be considered. Although data on susceptible populations are scarce, attempts should be made to determine the impacts of a contaminant spill on groups with increased susceptibility to the contaminant.

***5. What are the significant data gaps that inhibit the ability to answer these questions?*** Chapter 3 includes a discussion of known data gaps in the contaminant database after completion of the pilot site research and several other source water vulnerability assessments. In addition, managers may wish to review the number of ASTs with unknown contents upstream of their intakes and determine whether any of them warrant additional follow up phone calls or sites visits.

Finally, prioritization should be repeated at regular intervals as new data become available from source agencies. The ability to access up to date information will be most useful if updates to the information are monitored as they occur and prioritization is revisited when changes in the underlying data are deemed to be significant. Similarly, it may be useful to repeat the aerial image analyses every few years as new images become available to identify changes that are otherwise undocumented.

***Time of Travel and Contaminant Concentration***

A time of travel and contaminant concentration model can be used to incorporate travel times and expected peak concentrations at the intake into a prioritization plan. One method that can be used to calculate a spill's time of travel to the downstream drinking water intake relies on the work of Jobson and others at the USGS (Jobson, 1996; Ries III, 2007). These equations predict travel times to the leading edge of a plume, peak concentration and 10% of peak concentration (i.e., trailing edge) at the probable and maximum probable velocities. Other site specific models based on dye tests or river morphology may be available for some intakes.

The Jobson method makes several assumptions about river mixing and flow characteristics that simplify the modeling of pollutant and tracer transport and allow for the use of regression equations to predict travel time and concentration. These include the assumptions that vertical and lateral dispersion fully take place upstream of the measurement location, so that longitudinal dispersion of the pollutant is the dominant mixing process being measured or predicted. Generally, vertical mixing is said to occur rather rapidly in streams while lateral mixing is thought to be complete within the first several kilometers downstream of the entry point (Jobson, 1996). It is also assumed that additional contributory water sources, such as tributaries, outfalls from treatment plants, and even the merging streams, do not significantly alter the discharge and velocity of streamflow as to disrupt the process of longitudinal mixing. In determining the discharge (instantaneous or otherwise) at the intake and at the spill location, as well as at the midpoint of the stream traversal, an assumption is made about the relationship between the upstream contributing area and the discharge at the points of interest. Jobson (1996) uses a scaling equation of the form $Q_u = (A_u/A_g)^b Q_g$, with the exponent *b* set to a value of 1.

The first step in the calculation of the travel time following the Jobson and Ries method (2007) is the identification of a *reference gage*. For each implementation, each zone of concern or area of interest, the CEC tool identifies a nearby USGS gage as an appropriate source of current streamflow conditions. The appropriate donor gage is determined by compiling and screening a list of nearby¸ active USGS gages whose contributing areas share similar attributes to the contributing area of the project's drinking water intake. Attributes taken into consideration during the screening may include: total distance from the drinking water intake, total contributing area, ecoregion classification (level III), basin compactness, dominant geology (Reed & Bush, 2005; Hunt, 1979), and land use and land cover types within the gage's contributing area. Additional attributes currently under study by CEC for inclusion in the selection process include: average temperature and precipitation (PRISM dataset), latitude and longitude of the watershed's centroid, average baseflow index of the watershed, number of dams present on the main channel, Strahler stream order, percentage of different hydrologic soil groups within the watershed and watershed slope. With a reference gage identified, the time of travel calculation proceeds using the combined results of querying several web services hosted by the EPA and USGS. The EPA's WATERS ('WATERS') web services (USEPA, 2014) provide "REST like-access to the WATERS database services." These database services provide access to the latest enhanced edition of the National Hydrography Dataset (NHDPlus v2.1), allowing users to perform complex analyses based on querying the underlying data stored within the NHD. The USGS's REST ('UREST') service (USGS, 2015) provides access to almost all available data (real time, historical statistics, site information, etc.) for any gage maintained by the USGS. Both web services provide significant hydrologic analysis capabilities of which only a small subset is used in this methodology.

CEC's time of travel tool allows for on demand calculation of travel times anywhere within the drinking water intake's upstream contributing area. Additionally, travel times will be pre-calculated for each point of interest across a range of non-parametric percentiles calculated from the reference gage's historical streamflow measurements. The travel times from each point of interest for these percentiles will be stored within the GIS tool and available regardless of the current status of the WATERS service and the UREST service, whose use in the time of travel calculation is outlined below. This approach allows the user to compare on demand travel time calculation results with baseline travel times representative of various historical flows when both web services are running, while providing some estimate of travel time if one or both web services are unresponsive since the baseline estimates are stored internally. Accordingly, if a user knows the current streamflow near the intake and how this flow compares to historical values, they may be able to roughly estimate the travel time of the pollutant.

Currently recognized limitations in the tool consist of challenges in identifying appropriate stream donor gages, concerns about the availability of web services providing data used in the calculation and inability for the methodology to correctly handle estuaries, reservoirs, and damned reaches. The tool does not currently track the travel time for the overland flow path the spill takes to the stream; this will be addressed in future updates. Related to this are improvements that could be made to the prediction of the overland spill path through the use of higher resolution digital elevation models (DEMs); current overland flow path predictions are based on flow direction grids based on nationally available 30m digital elevation models (DEMs). Similarly, the delineation of upstream contributing areas relies on a combination of 30m DEMs and the medium resolution National Hydrography Dataset (NHD), which is accessed through the web services discussed below. Referencing the high resolution NHD dataset and higher resolution DEMs will refine upstream contributing area delineations and therefore improve streamflow predictions; this may be possible in the future depending on improvements to the web services accessed by the tool. Additional details about the time of travel methodology are provided in Appendix B.

### Reporting

Another important feature of a GIS is its ability to generate useful reports and data exports. For zones of concern that cross state boundaries, different state datasets for similar types of information must be integrated and differences between the information available from different states must be clearly understood. For example, when summarizing the number of ASTs in a zone of concern that crosses state boundaries, a report function must identify both states' datasets and the user should be aware of differences in the regulatory requirements for different threshold tank sizes and/or lists of chemicals for which reporting is required. Some reports will be useful for source water planning and prioritization (e.g., ASTs in the zone of concern, number of facilities of each type of potential source of contamination, etc.) and some may be specific to emergency response needs (e.g., list of facility contacts, chemicals and permits). In some cases, multiple datasets from different sources representing similar features may exist (e.g., brownfields and hazardous waste cleanup sites). Reporting functionalities should include the ability to display both datasets and allow water quality managers to compare information point by point.

# CHAPTER 3
# CONCLUSIONS

## POTENTIAL SOURCES OF CONTAMINATION AND CONTAMINANT DATA

### Most Useful Datasets

A wide variety of datasets indicating the locations of potential sources of contamination were discovered, generally from state and federal agencies, but a relatively short list of sources provided specific information about stored chemical identities or quantities (Table 2.3). These datasets are most useful for prioritization and emergency response. This list included EPA's TRI and DMR databases, state AST and UST databases, road and rail commodity flow data from LEPCs or EMAs, Tier II reports from LEPCs or SERCs, USACE barge commodity data, PWS downstream notifications and interviews or site visits. The complete list of federal, state and private datasets that were obtained to indicate potential sources of acute contamination can be found in Appendix A. The most helpful databases for populating the majority of the chemical data attributes included TOXNET, CAMEO Chemicals, CAS 'Common Chemistry', IRIS, MSDS and the Agency for Toxic Substances and Disease Registry. More information about each of these data sources is provided in Appendix E.

Several of the most useful datasets require the cooperation of LEPC's to obtain. Under the Emergency Planning and Community Right to Know Act, LEPC's are charged with developing an emergency response plan, updating it annually and providing information about chemicals in the community to citizens. These emergency response plans must include the following elements:

- "Identification of facilities and transportation routes of extremely hazardous substances
- Description of emergency response procedures, on and off site
- Designation of a community coordinator and facility emergency coordinator(s) to implement the plan
- Outline of emergency notification procedures
- Description of how to determine the probable affected area and population by releases
- Description of local emergency equipment and facilities and the persons responsible for them
- Outline of evacuation plans
- A training program for emergency responders (including schedules)
- Methods and schedules for exercising emergency response plans" (US EPA, 2015)

In most states, grant funding is available to support the activities of the LEPCs. Clearly, many of these activities are of interest to drinking water utilities, and utilities may join their local LEPC to facilitate data acquisition, strengthen local emergency response relationships and help guide efforts to prepare for local chemical emergencies. The process for joining and LEPC varies from place to place and is often as simple as calling the LEPC contact or filling out an application. For broader application, some of the data available from LEPCs may also be available by request from SERCs, although the request may be funneled back to the LEPC or the request may need to be made only for already known facilities at specified addresses.

## Challenging Datasets

The most common data challenges encountered were:

- A lack of data access due to security concerns
- A lack of response from data owners to requests for information
- Data that required significant processing to associate useful information with geospatial locations
- Datasets that were inconsistently formatted from one update to the next
- Lack of published information for some contaminant attributes

Specific challenges associated with some of the most useful datasets are described below. The most difficult data to obtain and manage were the specific chemicals and capacities stored in ASTs. For example, approximately half of CEC's requests for Tier II information were not fulfilled, either due to a lack of response from the LEPC or SERC or because the LEPC required that requests be submitted for individual facilities rather than geographic areas such as counties or zones of concern. Utilities that were members of their LEPCs or maintained close relationships were generally able to facilitate access to the complete set of useful data. Chemicals claimed as trade secrets on Tier II reports posed a unique problem because drinking water utility managers, while interested in the locations, quantities and chemical and toxicological properties of these chemicals, may not wish to assume responsibility for safeguarding trade secret information. This category of contaminants is also least likely to have thorough research available on physical, chemical and toxicological properties.

Mobile sources of contamination also posed a data collection challenge. Relatively few sources of data were identified to characterize these types of sources (i.e., truck, rail, barge, etc.). The most useful data sources were the US ACE's Ports dataset and hazardous commodity flow reports available to Emergency Management Agencies by request from rail companies. In one case, a local Emergency Management Agency had conducted a hazardous commodity flow study that included information about substances traveling by road through the county, but this type of study is not standard practice. Information about mobile sources was both difficult to find and also posed security challenges to access. With the exception of the ACE's Ports dataset, obtaining information about mobile sources required significant cooperation from the local Emergency Management Agency.

AST databases in particular required additional processing steps to incorporate useful information. In some states, no AST records were kept (e.g., IN) while in others only a subset of the data was publicly available. For several states, information about the specific chemicals stored in ASTs was provided as a separate table, which was then joined to geospatial locations of ASTs. These tables did not always have the same number of ASTs and in some cases did not use the same tank ID numbers, making this process more difficult. The state AST datasets also posed some of the greatest data access challenges. For example, information about the location and contents of ASTs with extremely hazardous chemicals and large storage capacities was not provided by the Pennsylvania DEP in response to the initial request for information, although a file review process to provide this information was agreed upon after discussions between the utility and the DEP. For West Virginia, the locations of ASTs were publicly available online but the contents and capacities of ASTs were obtained by special request. The contents of ASTs that were claimed as trade secrets were not provided.

Some datasets (e.g., Tier II reports, downstream notifications, industrial pre-treaters, some ASTs) were provided as tables with addresses but not with coordinate locations, requiring geocoding using several different processors. Invariably, some addresses were unable to be geocoded either due to errors in the source's address data or limitations of the geocoding applications. This type of challenge resulted in some data that could not be mapped in the GIS, potentially causing data to be overlooked during spatial analyses. These data can be reviewed in tabular format by water quality or GIS managers to determine whether additional information is known that would allow them to be geospatially located.

Datasets from state permitting agencies also posed challenges for data updates. Update frequencies varied widely, from a snapshot that would never be updated to live web services that instantly incorporated data updates from the source. Most state programs did not have a regular update schedule for their data and in some cases different versions of the same dataset were available on different web pages or from different staff members. Often the data available on a state agency's website was a subset of a more detailed internal database and the specific query used to create the publicly available dataset determined the inclusion or exclusion of useful fields as well as the number of records. These queries in some cases did not include the same fields and formats in updated versions of the same dataset. For example, many of the useful fields in the WV AST database (e.g., tank installation date, secondary containment, date of last inspection, etc.) were not available in the public data viewer but were provided when specifically requested. Different versions of the PA AST dataset requested several months apart lacked common identifying codes for sites, making the assessment of changes in the data more complex. Interestingly, the datasets specific to the identity and storage volume of hazardous chemicals, such as Tier II reports, were generally maintained on an annual basis due in part to annual regulatory reporting requirements.

Data generated from the analysis of aerial imagery posed another set of unique challenges. High-resolution and oblique imagery were in some cases difficult to incorporate into a GIS due to proprietary restrictions on their use. In some cases, alternative methods for investigating specific locations with these technologies were developed external to the GIS platform. The image analyses typically resulted in the identification of hundreds to more than 2,000 previously undocumented ASTs as well as other points of interest, such as gas stations, waste piles, etc. The contents of a relatively small percentage of these visible ASTs were known through federal or state regulatory databases or other sources. In part, the large number of visible ASTs with unknown contents stemmed from state regulatory definitions that determined which ASTs must be registered or inspected, if any. For example, in Indiana, owners of ASTs containing flammable or combustible liquids must apply to the Department of Homeland Security, which maintains a database of inspections for these tanks, and also comply with inspection and emergency planning regulations. However, ASTs containing other chemicals are not required to be registered with the state. Thus, most ASTs do not appear in the state's AST database. Although information about ASTs with listed hazardous substances above threshold quantities was generally obtainable in all states, information was not available for ASTs that did not meet the minimum capacities or did not contain hazardous substances as defined in state regulations. For many ASTs, a phone call or site visit may be the only method capable of determining the tank's contents.

Another challenge with ASTs identified from aerial imagery was the need to subjectively determine whether a nearby state AST database point referred to an AST seen in the imagery. Due to the level of error in the location of state data points, either due to errors associated with the equipment used to locate the tanks or the geocoding process used to locate an address, it was rarely possible to definitively determine which AST the state data point described. Therefore, a certain amount of professional judgment was needed to determine whether an AST seen in an image and information about the AST from a nearby state data point referred to the same point. This lack of certainty could be important during an emergency situation if an incorrect assumption was made about which data corresponded to which AST, particularly at sites with many ASTs in close proximity. These uncertainties could be handled several ways:

1) A user could make a subjective decision about which ASTs visible in the aerial imagery correspond to specific ASTs in another dataset. The level of certainty in this decision could be recorded as an attribute of the AST data point.
2) All AST datasets could be displayed without transferring information from one dataset to another. For instance, the locations of ASTs visible in aerial images but lacking chemical or other tank data could be displayed alongside AST data points from state data sources that include chemical information. This approach allows the user to quickly make his/her own subjective decision about which ASTs correspond to each other and understand the risks associated with this decision in an emergency situation. This is the approach used for the current project.
3) Site visits could be conducted to definitively determine which AST contains which chemical.

Some of the contaminant data could not consistently be found using the main data sources such as TOXNET and CAMEO. These data include RfD, method detection limit (MDL), treatability, Henry's law constant and no observed adverse effect levels. Often, these data for a specific chemical were only documented once, in a private publication. It was not possible to access this information without a membership or special credentials. Two major sources that would allow access to some valuable information not found in the other databases are the publications of the American Chemical Society (American Chemical Society, 2015) and the Royal Society of Chemistry (Royal Society of Chemistry, 2015). These sites appear to be the only source of the Henry's Law Constants for several of the more complex acids in the database. In addition, many RfD and NOAEL values for complex organic molecules and refined petroleum products are only found in publications on these sites. Method detection limits for a few complex, rare chemicals could not be found. However, the results of several experiments dealing with these chemicals were published on one of these sites. Likewise, more specific and in-depth processes for the removal of many of the contaminants identified in the zones of concern for the pilot sites were detailed in these publications. Information from these two websites is quite current, and many online databases do not include the data that these publications provide. For data gaps that are significant enough to delve into the primary literature, the American Chemical Society and Royal Society of Chemistry may be important sources of less commonly available information.

### Currently Inaccessible Datasets

Some datasets had limited access that precluded CEC from incorporating the information into the GIS platform (Table 3.1). Some of the information contained in these datasets was presumably already identified through other datasets (e.g., EIA's pipeline database, TOXNET, CAMEO, etc.) or data requests (e.g., Tier II reports). However, there are likely to be new sites and new chemical information contained in the inaccessible datasets that are not currently accounted for in CEC's GIS and chemical database. Facilities with trade secret chemicals or chemicals identified as terrorist security threats were particularly underrepresented in the current methodology due to these access restrictions. Further discussion among the drinking water community is needed to determine whether these sources contain useful information and whether there is a viable path to incorporating them into this methodology.

**Data Acquisition Highlights**

Overall, the state data acquisition process produced a wealth of information about potential sources of contamination that will aid users in source water protection assessment and planning as well as emergency response. However, it also highlighted several challenges of working with state data, especially across state boundaries:

- Data become out of date quickly, and there can be a tendency not to update on a regular basis. Out-of-date or incorrect phone numbers and addresses, for example, appeared to be quite common in both federal and state datasets.
- Metadata, if available, do not provide enough detail to make sense of data attributes. Metadata are particularly important because the ability to make use of a dataset can hinge on whether or not the domains for specific attributes/fields are clearly understood. For example, a priority code of 1-3 is meaningless without the definition of which level is highest priority and how that decision was made.
- In some cases, datasets that would seem to provide the most useful information are not as comprehensive as might be expected. For example, Ohio's AST GIS dataset appears to miss the majority of ASTs in the state and even those that are documented are missing data in many of the most useful fields.
- The type of information collected by state agencies is ultimately governed by state and federal regulation. For ASTs in particular, the wide variability in state regulation governing which ASTs are regulated produces wide variability in the number of ASTs that are documented. This type of difference among state regulations can make source water planning and prioritization processes more challenging, particularly for sites with multi-state drainages.

**Table 3.1**
**Challenging data sources**

| Source | Dataset | Challenge |
|---|---|---|
| *US Department of Transportation, Pipeline and Hazardous Materials Safety Administration* | National Pipeline Mapping System | This database and GIS viewer contain the locations of gas transmission pipelines, hazardous liquid pipelines, liquefied natural gas plants and breakout tank farms that are regulated by the administration or provided voluntarily. It does not include distribution or gathering pipelines. Some of the data are available through a public viewer, including the locations and ownership of gas and liquid hazardous pipelines. The viewer allows a user to access information for a single county and save the map as a pdf file but not to download any GIS formatted data or access the GIS server directly. The map scale and attributes are limited in the public viewer. The full system is accessible only to pipeline operators and government officials and includes more attributes and datasets. |
| *US EPA* | Water Contaminant Information Tool (WCIT) | This database contains information about contaminants of concern for drinking water security, including chemical, biological and radiological contaminants. Access is limited to government officials, select water utility personnel, and certain other water associations. |
| *US EPA* | State Drinking Water Information System (SDWIS) | The database contains drinking water intake locations and violations of safe drinking water regulations. Most of the data are accessible through the web search tool; however, the exact locations of drinking water intakes are not provided. |
| LEPCs / SERCs | Tier II hazardous chemical storage data | Some LEPCs require that requests be submitted for specific facilities rather than geographic areas. Data for ASTs containing materials claimed as trade secrets is only disclosed to health officials during emergency situations. |

(continued)

**Table 3.1 (Continued)**

| | | |
|---|---|---|
| *University of Texas at Dallas, US EPA, US Department of Homeland Security* | E-Plan Hazmat Emergency Management System | This online tool provides access to Tier II and Risk Management Plan hazardous chemical storage data, chemical profiles, MSDS sheets, emergency response plans, facility and area maps, federal area contingency plans and nearby schools and hospitals. Not all states require Tier II reporting to be completed through E-Plan. The system is accessible by first responders and state and federal users. |
| *US EPA* | Risk Management Plans (RMP) | Facilities using designated extremely hazardous substances above threshold amounts are required to submit Risk Management Plans to EPA. Plans must be revised and resubmitted every 5 years. These data are not available online. Some of the data may be requested by FOIA but some can only be accessed at a federal reading room. |
| Rail Companies and EMAs | Commodity Flow Reports / Density Studies (Rail) | Rail companies track the contents of each rail car. However, this level of detail is only released in an emergency situation. The rail companies are required to provide the most commonly carried hazardous materials for rail lines to First Responders, including Emergency Management Agencies. These reports may be requested directly from the rail company, but will often be redirected back to the local Emergency Management Agency. Some reports have standard confidentiality clauses that limit their use to first responders, a designation that does not include drinking water utilities. |
| DHS | Chemical Security Assessment Tools (CSAT) | Information about facilities that meet the definition of high-risk chemical facilities. User must be Chemical-terrorism Vulnerability Information (CVI) certified to access this system. |

## DETECTION AND CONTAMINANT WARNING SYSTEMS

Robust drinking water utility responses to spills and other acute impairments of source water quality are more likely given robust preparation. The keys to spill preparedness are advance knowledge about the contaminants that might be spilled (including lines of communication with the entities responsible for and maintaining the contaminants), up to date information that can be used to facilitate response team activities, water quality monitoring and comprehensive management and analysis of water quality data.

Water quality monitoring provides the most direct means for assessing risks to customers, for detecting unreported spills and for informing the timing and strategy of spill responses (e.g., increased treatment, switching sources or other responses). Water quality monitoring also presents opportunities for improved operations, delivery of higher quality water to customers and cost savings derived from greater operational efficiency. Utilities face a significant challenge in deciding what to monitor and how to monitor it.

There are a significant number of contaminants that could appear at a given intake, although some are more likely to appear than others. Monitoring might be conducted for:

- Contaminants with extremely high potential adverse impacts (e.g., biological, chemical or radiological warfare agents)
- Contaminants most likely to appear in source water
- General water quality characteristics (e.g., pH, total organic carbon, conductivity, ORP and other readily-measured parameters)
- Toxicity
- Specific analytes (e.g., toluene or microcystin concentration)

Risk and practical considerations should guide the selection of parameters for monitoring. Risk is the product of the likelihood of exposure and the likelihood of an adverse impact. The likelihood of exposure to contaminants stored or otherwise present upstream of a given intake is greater than that of contaminants not known to be present in the catchment. Therefore, logical choices for monitoring are the contaminants stored upstream of an intake in significant volumes and that are known to significantly impact human health. Results of the pilot and other recent source water vulnerability studies indicate that refined hydrocarbons and crude oils are the potential contaminants most commonly found in drinking water plant zones of concern and are good candidates for monitoring. Other contaminants known to occur frequently in many catchments are algal toxins, pesticides, herbicides, metals and *Cryptosporidium* oocysts.

Online instruments for measurement of specific analytes in drinking waters are increasingly used by the water community and the variety of instruments is also increasing. However, there is no all-inclusive instrument that can cost-effectively monitor for the whole suite of potential contaminants in source water. A number of instruments that are currently available are described here briefly. This description is intended to be general and is not intended to provide an assessment of the types of instruments described or as a confirmation of claims of the instrument vendors. Many online instruments have outputs that facilitate communication (e.g., cellular modems or wireless internet capability) and some have on-board diagnostic software, such as simple threshold alarming capabilities. Spectral scanning instruments are becoming increasingly available and can be used for detection of more than one contaminant with a single instrument. Most spectral instruments rely on absorbance measurements, though at least one device currently marketed to water and wastewater operations uses a hybrid multispectral approach intended to provide more specific measurement of target analytes. Online instruments are also being marketed for metals detection, algae detection, detection of volatile compounds and other parameters. Instruments are also available for online monitoring of more familiar drinking water quality parameters such as chlorine, conductivity, pH, TOC, ORP, turbidity and others and it is possible that outputs from those more familiar devices can be integrated into a signal capable of indicating adverse water quality events.

Rather than monitor specific contaminants, utilities might choose to monitor a general water quality characteristic such as toxicity. Monitoring non-specific water quality characteristics allows detection of a wider variety of contaminants, including contaminants that were not known to be present in the catchment. Most frequently, toxicity monitoring is conducted with living organisms such as fish, mussels, bacteria or algae. Depending on the organism used in a toxicity monitor, the contaminants indicating water toxicity might not indicate a human health threat. It is also not clear what response a water utility can take responsibly based on positive toxicity testing. The tests do not relate directly to any contaminant and, therefore, it is difficult to determine from the results what actions should be taken. Further, maintenance of the test organism can be difficult and lead to false positive results. These limitations need to be considered when evaluating toxicity monitors for raw water monitoring.

### How to Monitor

Once parameters and instruments are chosen for raw water monitoring, utilities must determine how monitoring will be conducted. Important considerations for sample location are:

- Whether there is any chemical addition (e.g., pre-oxidation) upstream of a candidate sample location
- The depth of the sample line intake
- Ability to sample when the intake is shut or if the plant does not run continuously (i.e., sampling location relative to the intake pump)
- Access and power supply where the monitor is housed
- Waste stream handling if there is a waste stream from the instrument
- Ease of calibrating and maintaining instruments
- Suspended material load in the sample line
- Sample line pressure

Depth of the sample intake is particularly important if detecting fuel spills is the primary intent of monitoring. Sample line intakes near the surface of the raw water source are more likely to detect fuel spills (fuel specific gravity is usually $< 1$), but samples collected near the water surface might not be representative of water quality for intakes significantly lower than the water surface. Sampling can be done online or via grab samples. In both cases, the frequency of measurement can be adjusted and should be set based on the variability and uncertainty of the parameter of interest and the detectable difference required for the measurement.

### Capturing and Using Water Quality Data

Data management and analyses are required for inferring events from water quality data. Online water quality data require quality assurance and control to account for noise, spikes and periods of sensor down time or calibration. Some instruments have on-board data analysis capabilities, such as alarming on exceedance of thresholds, calculation of moving averages or estimation of rates of change. Ideally, all water quality data are integrated into a comprehensive event detection system. Many vendors have developed “middleware” event detection packages that consume water quality and other data, assess whether the data indicate an event and report the outcome to a Supervisory Control and Data Acquisition (SCADA) system or other downstream data system. Event detection software packages differ widely in their ability to distinguish credible water quality events from changes in water quality that do not pose a concern. Calibration or conditioning of the algorithms and parameters used in the event detection packages is a critical step in limiting the number of false positive alarms generated and in keeping the system from becoming a burden to utility staff who must respond to alarms. Adequate calibration is equally critical to ensure that important events are not missed.

## STRENGTHS AND LIMITATIONS

### Image Analysis

The image analysis methodology successfully identified the locations of many previously unidentified ASTs, ranging from hundreds to more than 2,000 new ASTs identified per zone of concern. A variety of other, less common potential sources of contamination were also marked, including gas stations, bridges, animal feeding operations, slurry ponds, etc. Part of the success of this methodology relies on decisions made by trained image analysts. The advantages of human observation and analysis of areas of concern for water plant intakes derive from the analyst’s ability to make determinations based on professional judgment and past experience, such as:

- Interpret the morphology of visible features: In a satellite image, an Airstream camper and a horizontal AST can appear to be very similar, as can innocuous farm silos and vertical ASTs. Gazebos can also mimic ground tanks, etc.
- Comprehend the context of features: A swimming pool and a waste water process tank may look identical in aerial images.
- Expect points of interest to be present at locations where they are not immediately visible: Highway departments, trucking companies, school bus fleet stations and mid to large sized farms will likely have fuel tanks on site, but they may not be readily apparent in aerial views. Some additional sleuthing may uncover these hidden features.
- Select alternative supplemental imagery tools: Suspected sites may not have visible ASTs from one aerial image source, but an analyst may opt to use other satellite imagery, including Bing Bird’s Eye and Google Street images, to reveal them. Further, the dates that satellite or street images were captured may be available, allowing a determination that a point of interest visible in one image set is in fact no longer present, as determined by its absence in a more recent image.
- Locate points of interest from related permitted features: NPDES and RCRA permit details may indicate a gas station that is not distinguishable from the imagery.

- Use external information to locate points of interest or supplement knowledge of a potential point of interest: Boating web sites were used to confirm whether a marina does or does not have fuel on-site.

Weaknesses of the manual image interpretation method include the reliance on the availability of recent, very high resolution imagery and the likelihood that different analysts may produce slightly different results, especially where the imagery is lower resolution or has greater leaf or cloud cover. Finally, the identification of a potential source of contamination is unlikely to provide detailed information about the chemical nature of the threat, such as the content of an AST.

**Comparison of Greater Cincinnati Water Works (GCWW) and CEC Methods**

The goal for the GCWW pilot site was to compare GCWW's past source water assessment results to CEC's results from the new methodology. Two levels of assessment were conducted. First, federal and state datasets indicative of potential sources of contamination were found and loaded into the GIS platform. These data were systematically compared to the set of potential sources of contamination in GCWW's database. Second, two sites of particular interest to GCWW, totaling less than one square kilometer, were selected for aerial image analysis. This approach differed from the other pilot sites, where image analysis was completed for the entire zone of concern and current results were not systematically compared to past results.

GCWW collected data over the course of many years to assess potential hazards to source water intakes in their Zone of High Concern, one of three areas defined by state regulation for source water assessment purposes. The data sets ranged from a survey of NPDES sites in 2001 to a field survey conducted in 2005. In total, GCWW collected 901 individual points of interest (POI) from a variety of data sets and field observations. In many instances, multiple, spatially-distinct points referenced different attributes of the same facility. Conversely, many POIs represented a point found through a database that, when visually assessed by GCWW, led to identification of a separate feature of interest at the identified site. For example, a marina identified through the Inland Sensitivity Atlas – Marinas data query, was noted to have two portable ASTs on-site that were documented through local knowledge and visual analysis.

A comparison of GCWW and CEC POIs was performed through a combination of visual identification and data source comparisons. CEC utilized 39 data layers available in-house to assess any potential matches to GCWW data (Table 3.2). This approach involved attempting to match GCWW-identified points to any corresponding points for the same facilities from CEC datasets. The methods included visual proximity and address and facility name matching, as well as EPA program ID matching to GCWW unique ID listings. As an example, although CEC did not include a dataset for underground storage tanks, 33 POIs in the Ohio Bureau of Underground Storage Tank Regulations (BUSTR) – Leaking Underground Storage Tank (LUST) dataset used by GCWW were matched to CEC POIs through other datasets (RCRA, NPDES and Ohio AST registry) via visual proximity and facility name matching. In addition to this comparison, CEC also completed aerial imagery analysis for two sites in GCWW's Zone of High Concern.

The differences between the results of the two methodologies were due to several factors: 1) real changes in the number and type of facilities and regulatory datasets over time, 2) different datasets and data categories used, and 3) level of image analysis and different types of points identified during image analyses. In general, the CEC methodology focused on potential sources of acute contamination while GCWW also included an emphasis on chronic, non-point sources of potential contamination. Table 3.2 offers a comparison of the datasets used for each analysis.

**Table 3.2**
**Data layers used in analyses**

| GCWW Datasets | Source | CEC Datasets | Source |
|---|---|---|---|
| 2005 Field Survey | GCWW | Operable Generators | EIA |
| Bridges from Kentucky Transportation Cabinet | KYDOT | Proposed Generators | EIA |
| BUSTR: LUST database (geocoded) | OEPA | Natural Gas Pipeline (Interstate) | EIA |
| ESRI | ESRI | Bridges | USGS/USC |
| Field Survey and NKADD database | GCWW | Rail Crossing | USGS/FRA |
| Historical Landfills - Orphan Sites from Solid Waste Branch | KYDEP | CERCLIS Facilities | EPA |
| Inland Sensitivity Atlas - Boat | EPA | NPDES Facilities | EPA |
| Inland Sensitivity Atlas - Marinas | EPA | OIL Facilities | EPA |
| Inland Sensitivity Atlas-Oil Storage Facilities | EPA | RCRA (Active) Facilities | EPA |
| Kentucky Facility Index (KFI) | KYDEP | TRI Facilities | EPA |
| KPDES - PCS | KYDEP | Airports | FAA |
| ODNR-DMRM | OEPA | Dams Over 50 Feet in Height | USACE- |
| OEPA Landfill GIS Layer | OEPA | Inland Waterway Ports | USACE |
| OEPA-DSW Surface Impoundment | OEPA | Aboveground Storage Tank | Corona |
| Ohio EPA-DERR MSL GIS layer | OEPA | Coal Pile | Corona |
| Ports from Corps of Engineers | USACE | Power Plant | Corona |
| RCRAINFO (Non-TDS) | EPA | Silos | Corona |
| Tier II from Emergency Management Division | GCWW | Aboveground Storage Tanks | OHSFM |
| US EPA Envirofacts | EPA | Active Permits for Concentrated Animal Feeding Operations | OHEPA |
| EPA-BASINS NPDES Facilities (2001) | EPA | Biosolids Application | OHEPA |
| USGS Geonames | USGS | Combined Sewer Overflow | OHEPA |
| | | Industrial Mining Operations | OHDNR |
| | | NPDES Individual Permit | OHEPA |
| | | Surface Coal Mining | OHDNR |
| | | Underground Coal Mining | OHDNR |
| | | Underground Industrial Mining | OHDNR |
| | | Available Mine Maps All | KYDMS |
| | | Kentucky Inter-System Operational Permit Points | KYIA |
| | | Lift Stations | KYIA |
| | | Oil and Gas Wells | KYGS |
| | | Package Treatment Plants | KYIA |
| | | Water Tanks | KYIA |

(continued)

**Table 3.2 (Continued)**

| | | | |
|---|---|---|---|
| | | Water Treatment Plants | KYIA |
| | | Wastewater Treatment Plants | KYIA |
| | | Wastewater Treatment Plant Outfalls | KYIA |

***Differences in Common Data Categories***

Several of the datasets used by both methodologies produced significantly different results, likely because reporting programs and datasets have evolved over time and may be different than at the time GCWW examined them. One such example is GCWW's NPDES permit search in 2001 using the EPA's Better Assessment Science Integrating Point and Nonpoint Sources (BASINS) tool. While this tool still exists, it is unclear how GCWW structured the 2001 data search. Because of the 14-year time difference in datasets, as well as utilizing a different tool, GCWW identified 24 NPDES POIs of which CEC identified 10 current matches. Similarly, the GCWW method included a search on the EPA Envirofacts database. It is unclear when this search was performed. Given a potential multi-year gap between the dates of the CEC and GCWW database searches, information for RCRA, NPDES, CERCLIS, and other datasets available through this portal could vary considerably. Unlike the BASINS query, the Envirofacts dataset contained points from multiple EPA programs. CEC utilized all available layers to assess matches and was able to match 109 out of 168 Envirofacts POIs listed in the GCWW dataset, including 69 out of 87 data points from the Permit Compliance System (PCS). GCWW's RCRA dataset included 5 points identified as treatment, disposal and storage (TDS) facilities while CEC's assessment included all RCRA facilities (including TDS) and identified 60 sites. Of the unmatched POIs in the Envirofacts dataset, 22 referred to air pollution permitting, a category not included in the CEC analysis. Lift stations were another category of POI common to both assessments but with unusually different results; GCWW identified 7 lift stations as part of the 2005 field survey while CEC identified 30 lift stations from a KY dataset (no data were available for OH). CEC-identified lift stations did include all seven of the lift stations identified by GCWW.

***Data Categories Unique to One Method***

CEC's method in part utilized data sets that were created based on a specific feature or set of features, such as CAFOs, biosolids application sites, mines, among others. GCWW utilized this approach as well. Because the data sets available and chosen at the time of each study varied considerably, there were instances in both methodologies where POIs unique to a given data set were identified. For example, past source water assessments focused on point and non-point source contaminants for a source water, while the CEC methodology focused on acute, point source threats to source water. Many of the POIs related to non-point sources, such as golf courses, cemeteries and large parking lots, were facilities identified in GCWW's 2005 Field Survey or identified through local knowledge or visual identification. This type of POI accounted for approximately 379 sites identified by GCWW. GCWW also included data related to marinas, impoundments, underground storage tanks and several categories of commercial establishments, such as auto parts stores and dry cleaners, which were not included in CEC's method, but may be available as GIS datasets from state agency or other sources. For example, underground storage tank data are available from the BUSTR database but were not included as part of the CEC methodology because they were deemed less likely to pose an acute threat to source water than ASTs or other surface features.

CEC's method also included unique datasets that did not appear in GCWWs approach, such as power generators, TRI facilities, airports, large dams, coal piles, silos, CAFOs, biosolids application sites, mines, package treatment plants and wastewater treatment plant outfalls. In comparing the two methods, however, many POIs reflect the identification of the same site by different means. For example, a power generating station was identified by CEC through a data set from the Energy Information Administration (EIA) that features operable power generating stations. This same site was identified by GCWW through Envirofacts and their 2005 Field Survey. In many instances, POIs identified by GCWW were done so through the 2005 Field Survey and were not necessarily categorized in the same way as a data set based on a specific feature.

CEC's methodology also included special requests for information about the location of specific potential contaminants at stationary and mobile sources. For example, CEC staff requested Tier II hazardous chemical storage reports and hazardous commodity flow data for railroads from the Emergency Management Agencies in Campbell County, Kentucky and Clermont County, Ohio. These data, if received, will be added to the GIS as new POIs and will make available information about the location and quantities of specific potential contaminants within the Zone of High Concern and at the two facilities chosen for the pilot study.

### *Differences in Image Analysis Methods*

The last set of differences between the two methodologies was related to the types of sources identified through aerial imagery analysis. Some of the categories of POIs that GCWW identified through imagery analysis, such as unlisted ASTs, slurry ponds and quarries, would be captured by the CEC imagery analysis methodology if it were applied to the full zone of concern. In fact, at just the two sites where image analysis was completed (covering less than one square kilometer), CEC identified 57 ASTs, none of which were included in the OH State Fire Marshall's AST dataset. However, other sites such as parking lots, cemeteries and golf courses were not part of the CEC visual analysis methodology. These types of facilities are readily visible in the images, though, and could be added to the protocol if desired.

### *Data Summaries*

Of the 901 POIs identified by GCWW, 249 were also identified by CEC. A complete imagery analysis of the Zone of High Concern by CEC staff would likely increase this number, as would the addition of datasets not currently part of the CEC methodology, such as underground storage tanks (Table 3.3). The CEC data analysis of the Zone of High Concern identified a total of 681 points of interest (Table 3.4). This leaves 432 POIs identified by CEC that were not captured by GCWW's analysis.

**Table 3.3**
**Summary of POIs in GCWW data layers**

| Source | Facilities (total) | Facilities Matched by CEC Datasets |
|---|---|---|
| Bridges from Kentucky Transportation Cabinet | 33 | 33 |
| BUSTR: LUST database (geocoded) | 103 | 33 |
| ESRI (Golf Courses) | 2 | 0 |
| Field Survey and NKADD database | 11 | 2 |
| Historical Landfills - Orphan Sites from Solid Waste Branch | 2 | 2 |
| Inland Sensitivity Atlas-Oil Storage Facilities | 3 | 2 |
| Inland Sensitivity Atlas - Boat Access | 20 | 9 |
| Inland Sensitivity Atlas - Marinas | 11 | 2 |
| Kentucky Facility Index (KFI) | 7 | 0 |
| KPDES - PCS | 12 | 10 |
| ODNR - DMRM | 5 | 1 |
| OEPA-DSW Surface Impoundment GIS Layer | 10 | 5 |
| OEPA Landfill GIS Layer | 2 | 1 |
| OH EPA - DERR | 1 | 0 |
| Ports from Corps of Engineers | 4 | 3 |
| RCRAINFO (Non-TDS) | 5 | 1 |
| Tier II from Emergency Management Division | 4 | 1 |
| US EPA Envirofacts | | |
| AIRS/AFS | 17 | 2 |
| PCS | 87 | 69 |
| RCRIS | 56 | 30 |
| RCRIS AIRS/AFS | 2 | 2 |
| TRIS | 3 | 3 |
| TRIS RCRIS AIRS/AFS | 3 | 3 |
| USGS Geonames (Cemeteries) | 15 | 0 |
| EPA-BASINS NPDES Facilities (2001) | 24 | 10 |
| 2005 Field Survey | | |
| Monitoring Wells | 4 | 0 |
| Cemeteries | 30 | 0 |

(continued)

**Table 3.3 (Continued)**

| | | |
|---|---|---|
| Parking Lots | 99 | 0 |
| Auto Parts Store | 19 | 0 |
| Junk Cars/Yards | 17 | 0 |
| ASTs | 123 | 18 |
| Lift Station | 7 | 7 |
| Gasoline Pump | 60 | 0 |
| Dry Cleaner | 5 | 0 |
| No Description | 61 | 0 |
| Miscellaneous | 34 | 0 |
| Total | 901 | 249 |

**Table 3.4**
**Summary of POIs in CEC data layers**

| Source | Dataset | William H. Zimmer Station | ADM | Zone of Critical Concern | Zone of High Concern | Extended Zone |
|---|---|---|---|---|---|---|
| EIA | Operable | 0 | 0 | 0 | 0 | 419 |
| EIA | Operable | 1 | 0 | 1 | 0 | N/A* |
| EIA | Proposed | 0 | 0 | 0 | 1 | 43 |
| USGS/USCB-TIGER | Bridges | 0 | 0 | 15 | 129 | N/A |
| USGS/FRA | Rail Crossing | 0 | 0 | 19 | 3 | N/A |
| EPA | CERCLIS | 0 | 0 | 0 | 0 | 274 |
| EPA | NPDES Facilities | 0 | 0 | 30 | 146 | 21615 |
| EPA | OIL Facilities | 1 | 0 | 1 | 0 | 734 |
| EPA | RCRA (Active) | 1 | 0 | 6 | 53 | 18138 |
| EPA | TRI Facilities | 1 | 0 | 3 | 4 | 2387 |
| FAA | Airports | 0 | 0 | 0 | 3 | 435 |
| USACE-NID | Dams Over 50 | 0 | 0 | 1 | 2 | 417 |
| USACE | Inland Waterway | 3 | 1 | 29 | 3 | 876 |
| Corona | Aboveground | 57 | 0 | N/A | N/A | N/A |
| Corona | Coal Pile | 1 | 0 | 2 | N/A | N/A |
| Corona | Power Plant | 1 | 0 | N/A | N/A | N/A |
| Corona | Silos | 0 | 12 | N/A | N/A | N/A |

(continued)

**Table 3.4 (Continued)**

| | | | | | | |
|---|---|---|---|---|---|---|
| OHDNR | Aboveground Storage Tanks | | | | | |
| | Active | 0 | N/A | 0 | 3 | 574 |
| | Inactive | 0 | N/A | 2 | 11 | 1822 |
| | Unknown | 0 | N/A | 0 | 0 | 29 |
| OHDNR | Active Permits for Concentrated Animal | 0 | N/A | 0 | 0 | 10 |
| OHDNR | Biosolids Application | 0 | N/A | 0 | 38 | 7192 |
| OHDNR | Combined Sewer Overflow | 0 | N/A | 0 | 0 | 261 |
| OHDNR | Industrial Mining Operations | 0 | N/A | 1 | 0 | 704 |
| OHDNR | NPDES Individual Permit | | | | | |
| OHDNR | Industrial | 1 | N/A | 3 | 3 | 591 |
| OHDNR | Public | 1 | N/A | 2 | 10 | 976 |
| OHDNR | Surface Coal Mining | 0 | N/A | 0 | 0 | 6257 |
| OHDNR | Underground Coal Mining | 0 | N/A | 0 | 0 | 44 |
| OHDNR | Underground Industrial Mining | 0 | N/A | 0 | 0 | 19 |
| KYDMS | Available Mine Maps All Seams | N/A | 0 | 0 | 0 | 17964 |
| KYIA | Kentucky Inter-System Operational Permit | N/A | | | | |
| KYIA | Lift Stations | N/A | 0 | 6 | 24 | 381 |
| KYGS | Oil and Gas Wells | N/A | 0 | 1 | 2 | 24458 |
| KYIA | Package Treatment Plants | N/A | 0 | 2 | 20 | 133 |
| KYIA | Water Tanks | N/A | 0 | 0 | 0 | 344 |
| KYIA | Water Treatment Plants | N/A | 0 | 0 | 0 | 26 |
| KYIA | Wastewater Treatment Plants | N/A | 0 | 0 | 1 | 36 |
| KYIA | Wastewater Treatment Plant Outfalls | N/A | 0 | 2 | 18 | 53 |

**N/A = not assessed*

### *GCWW / CEC Method Comparison Conclusions*

The comparison of GCWW and CEC methodologies highlighted the advancement in GIS data availability over the last decade as well as the unique strengths of the "boots on the ground" field visit and GIS data assessment approaches for a comprehensive source water assessment effort. The best tool will enable local source water protection managers to keep abreast of dynamically updated GIS data and incorporate additional, site-specific information collected through site visits, phone calls, spill response exercises or other means. A dynamic GIS information system should allow the user to be part of the data acquisition process and should enable users to keep track of their field visit and other data in a more accessible and searchable format than paper files.

## DATA GAPS

Two types of data gaps are described below:

- Gaps in the locations or details of sites considered primary indicators of the potential for contamination and
- Gaps in contaminant property data.

### Gaps in Site and Detail Data

CEC's search for a common set of data indicating potential sources of acute contamination was largely successful for the main data categories, even across state boundaries. In most states, additional relevant data were identified during web searches or discussions with state agencies and other organizations. Table 3.5 displays the result of the initial data search for all six states within the zones of concern of the five pilot sites. The most notable data gap was the lack of AST data for IL, IN and KY. A list of 101 AST inspections completed from January 1 – March 21, 2015 was obtained from the IL State Fire Marshall but did not contain AST contents or capacities. It also did not include all ASTs in IL, just those that were inspected during this three-month period. These data were deemed too incomplete to incorporate into the assessment. The locations of slurry ponds were not available in any state; this type of data may be best generated through aerial image analysis. For several of the pilot sites, GIS data from past source water assessments were not available. Tier II data were requested from LEPCs, generally at the county level, with varying levels of success. Some provided data, some did not respond and some responded that requests could only be made for specific facilities and not geographic areas, such as the zone of concern or county.

**Table 3.5**
**Gaps in data availability by state (X: data obtained; P: data partially obtained; -: data gap)**

| Dataset | IL | IN | KY | OH | PA | WV |
|---|---|---|---|---|---|---|
| Sites noted in previous source water assessments | - | - | X | X | - | X |
| Above ground and underground storage tanks | P | - | - | X | X | X |
| NPDES sites | X | X | X | X | X | X |
| RCRA sites | X | X | X | X | X | X |
| TRI sites | X | X | X | X | X | X |
| OIL sites | X | X | X | X | X | X |
| CERCLIS sites | X | X | X | X | X | X |
| Oil and gas operations | - | X | X | X | X | X |
| Mining operations | P* | X | X | X | X | X |
| Waste management sites | X | X | - | - | X | X |
| Pipelines | X | X | X | X | X | X |
| Power plants | X | X | X | X | X | X |

(continued)

**Table 3.5 (Continued)**

| | | | | | | |
|---|---|---|---|---|---|---|
| Refineries | X | X | X | X | X | X |
| Airports | X | X | X | X | X | X |
| Hazardous cargo transport routes | X | X | X | X | X | X |
| Railroad stream crossings | X | X | X | X | X | X |
| Road stream crossings | X | X | X | X | X | X |
| EPCRA Tier II reporting facilities | P | - | - | P | P | P |
| Slurry ponds | - | - | - | - | - | - |

* coal mines only

Another notable data gap was the lack of information about the chemical contents and capacities of many of the ASTs, especially those identified through the aerial image analysis. Reasons for a lack of data on AST contents are that the data are suppressed due to security concerns or business concerns, that AST contents are not regulated (for reasons described below) and data are not reported, or that AST operators have failed to register a tank. The large number of ASTs visible in aerial images but not appearing in state AST databases indicated that many tanks were unregistered. Many of these may not meet the registration requirements of the state. The AST registration thresholds for the six states in the pilot sites' zones of concern ranged from no registration or inspection requirements to more expansive definitions of size and contents. In general, state AST regulations tended to focus on hazardous substances as defined on one or more federal lists and petroleum products. For chemicals not included on these lists, very little information exists about locations or storage volumes. The contents of residential propane tanks (which CEC analysts were instructed not to mark) and PWS water storage tanks can be reasonably accurately estimated from their shape and location in the image. For other ASTs identified during image analysis, however, phone calls or site visits may be the only way to determine the contents and capacities.

**Gaps in Contaminant Property Data**

As noted above, contaminant data pertinent to prioritization and spill response were collected for contaminants known to be stored in zones of concern. Some of the desired contaminant attributes could not regularly be found using the main data sources such as TOXNET and CAMEO (Appendix F). A deeper dive into the primary literature could uncover additional research or other analyses to fill some of these data gaps, but this level of effort was beyond the scope of the current project. For some critical data gaps, this additional level of research effort may be warranted in the future. As part of the SOP described in Appendix E, data gaps were classified as individual chemical properties were researched. In some cases, a contaminant property was not applicable to a particular chemical. In other cases, a credible data repository indicated that the data did not exist. These data gaps were recorded as "Data Unavailable"; however, there is a possibility that more recent primary studies have been conducted since the data repository was created. For most data gaps, there was no reason to believe that the data did not exist, but they were not discovered following application of the SOP and were indicated as "Data Not Found".

The most common data gaps were treatability, RfD, method detection limit, Henry's law constant and no observed adverse effect levels. The EPA Treatability database was the sole source used for treatability data. Since that database presently has treatment data for only 64 contaminants, treatability data for a number of contaminants in the CEC's database were not found. Filling treatability data gaps entails a comprehensive search of the primary and secondary literature, assessment of all relevant studies and some level of meta-analysis of the data. Because this effort requires a substantial level of effort and can only be done by individuals with a relatively specialized knowledge, additional research on treatability is only merited for the contaminants of specific concern to utilities.

# CHAPTER 4
# RECOMMENDATIONS

## HOW TO USE THE METHODOLOGY AS PART OF A SOURCE WATER PROTECTION PROGRAM

The data collection and management methodology described in this report will help drinking water utilities enhance and maintain their source water characterizations, emergency preparedness and response capabilities. Using the capabilities of a GIS with a standardized data update schedule will also enable utilities to review and revise their priorities, detection capabilities and emergency response plans on a regular schedule.

More specifically, utilities can use the contaminant locations, chemical properties and data gaps identified with this methodology to prioritize specific sites or facilities for follow-up data collection, remedial action or emergency response planning activities. Information about permits, stored chemicals, compliance history, etc. can be used to prepare for site visits. For site visits to large properties with limited time, completing an analysis of aerial images prior to a site visit could help to focus the visit and discussion on particular storage or transportation areas of interest. If site access is not granted, aerial image analysis may be the only method for evaluating chemical storage locations and other potential sources of contamination.

The method described in this report also generates information that will support communication with significant upstream facility owners, the emergency response community, customers and the general public. For instance, contact information obtained from multiple data sources could be used to initiate or maintain communications with managers of significant facilities upstream. Utilities may wish to conduct tabletop spill response exercises at priority sites to better understand critical data gaps, resource availability and communication needs during an incident.

A review of known potential contaminants upstream and their physical characteristics could inform early warning and event detection activities and technology choices as well as laboratory sampling capabilities that may be required during a contamination event. These same data will be invaluable during an incident when time is short and fast access to time of travel estimates, detection and treatability guidance, fate and transport factors, health effects and safety precautions is critical. The chemical and physical properties of a contaminant can be searched in an emergency situation to help identify the location and/or identity of a chemical spill. For example, because chemical vapors are invisible, odor may be one of the most important early detection signals and may be related to spilled chemicals or breakdown product(s). Access to a shared contaminant database will increase the chance that breakdown products or chemicals previously unknown within the drainage area will be documented and accessible during an incident. Similarly, sudden changes in color or taste can indicate a change in water quality that could be investigated using an appropriately designed contaminant database.

No system can provide perfect up to date information about every contaminant that may become a problem, and even with a sophisticated geospatial platform and toolset, unexpected events may occur. However, with comprehensive planning, the methodology outlined in this report will allow a utility to prioritize its limited resources to prepare for the most likely threats.

Finally, because the contaminant database compiled for this project accumulated the information for chemicals at many intakes, the assessment of chemical research gaps will help to inform a larger discussion within the drinking water community about which data gaps are most critical to fill and how to accomplish this work.

## OTHER AVAILABLE TOOLS AND RESOURCES

### DWMaps

DWMAPs is a EPA online mapping tool designed to provide drinking water information, such as water supply companies, drinking water sources and sources of pollution. The functionalities include identifying water suppliers by area served, finding sources of potential contamination and accessing Clean Water Act surface water impairment information. The tool includes federal data related to major point source and discharge permitting programs (NPDES, RCRA, TRI, CERCLA, TSCA, CSO, pipelines, hazardous materials routes, and pesticide and fertilizer use by country). DWMAPs does not provide access to datasets owned and managed by states or other entities, such as state storage tank databases, Tier II Hazardous Chemical Inventory, results of image analysis, etc. Information about specific chemical properties is not provided through the interface. DWMaps is currently available to the general public. In the future, additional functionality may be provided to certain user groups, such as drinking water utilities.

### RiverSpill

RiverSpill is a GIS model developed by the US Forest Service to predict the time of travel, dispersion and decay for a pollutant introduced to water. It uses the USGS Enhanced River Reach file and SDWIS to link results to the locations of public water supplies and population served (Samuels, 2001). RiverSpill is now part of the ICWater modeling package, which can be obtained by federal, state and local government agencies from the Defense Threat Reduction Agency (DTRA).

### ICWATER

The Incident Command Tool for Drinking Water Protection (ICWater) was developed by the US Forest Service to support incident response needs for drinking water. It is based on ESRI ArcMap technology linked to the RiverSpill model and provides real-time assessments of contaminant travel and dispersion. The tool links RiverSpill to the NHD stream network as well as the locations of drinking water intakes, hazards and critical infrastructure (Ryan D. , 2007). The ICWater modeling package may be requested by federal, state and local government agencies from the Defense Threat Reduction Agency (DTRA) (Samuels R. , 2015).

### WCIT

The EPA’s Water Contamination Information Tool (WCIT) is a secure online tool that provides contaminant information for chemical, biological and radiological contaminants. It may be accessed by drinking water utilities, primary enforcement agencies, federal officials, public health agencies and water associations. Requests for access can be made online. The data included describe a contaminant’s physical properties, behavior in water and health effects as well as field and laboratory analytical methods and water infrastructure decontamination methods (USEPA, 2015).

## EPlan

EPlan is an online tool developed by the EPA and the University of Texas at Dallas to provide on-site hazardous chemical information (Tier II, RMP, 311 and 302 data) to first responders and emergency preparedness personnel. It can also be used by states as a Tier II reporting system (UT Dallas, 2015).

## THE WAY FORWARD

The methodology developed for this project allows utilities to use an information systems approach to create dynamic source water assessments that identify a greater range of potential contaminant threats than past assessments and have clear data update and maintenance schedules. Potential sources of contaminants can be displayed visually and analyzed in many different ways (e.g., proximity to intakes, groups of chemicals, time of travel, etc.). Using GIS technology coupled with government and private databases and a shared contaminant database, a wealth of information can be integrated in a single user-friendly platform. Powerful query tools allow questions to be asked of the data and maps, providing users with an understanding of complex relationships and the ability to make data driven decisions.

With this basic system in place, new functionality can be incorporated to increase the usefulness of the system to drinking water utilities. For example, the current method focuses on acute threats to source water while past source water assessments also included non-point source threats. Many datasets related to non-point source threats, such as nutrient and pathogen loading estimates, are now available as GIS datasets that could easily be incorporated into this type of geospatial platform. While this project focused on surface water intakes, similar analyses are being performed for groundwater systems. With a data collection effort focused on groundwater contamination, the same system could be used to evaluate threats to groundwater wells or new locations for wells or surface water systems.

Several other possibilities were identified during the scoping and requirements workshops that bear repeating here. Communication tools to improve source water protection efforts as well as emergency response could be added to the toolsets. Features such as mail merge data for upstream AST, Tier II or downstream notification facilities could largely automate regular contact with facility owners and help to ensure that current data are available when they are needed.

The wealth of data about potential sources of contamination, chemical properties and facility contact information compiled in a GIS could be an essential component of tabletop exercises that bring together first responders, drinking water utilities and facilities of particular interest. These exercises typically strengthen the relationships needed to respond effectively in an emergency, highlight where resources can be shared or are needed, and how data can be most effectively used and shared during an incident. A geospatial platform is an ideal place to store information as it becomes available and manage secure access to sensitive information.

Work on the five pilot projects also highlighted the importance of planning strategic site visits, recording the data in an easily accessible and searchable format and staying involved with LEPCs to facilitate data acquisition for chemical hazards and to maintain strong relationships with the local emergency response community.

# APPENDIX A
# ANNOTATED LIST OF DATA RESOURCES

* Indicates that acquiring the dataset may require a FOIA or manual data request.

## OHIO

| Ohio Dataset | Description | Source | Access Format | Update Frequency |
|---|---|---|---|---|
| Above Ground Storage Tanks* | Locations, contents and capacities of above ground storage tanks. | OH SFM | MS Excel | As needed |
| Oil & Gas Wells | Oil & Gas Wells - active, inactive, horizontal and directional. | OH DNR | Web Service and Shapefile | TBD |
| Oil & Gas Fields | Oil & Gas Fields of Ohio. | OH DNR | Web Service and Shapefile | TBD |
| Mines of Ohio | Mine openings, abandoned underground mines, surface mine operations, active underground mine extents. | OH DNR | Web Service and Shapefile | TBD |
| Industrial Mineral Locations | Locations of industrial mineral operations, including ownership and contact information. | OH DNR | Web Service | TBD |
| Combined Sewer Overflow Locations | Combined sewer overflow locations (CSOs) permitted by the State of Ohio. Locations have not been verified and may represent locations of relief structures rather than outfalls. | OH EPA | | TBD |
| Concentrated Animal Feeding Operations | Locations of concentrated animal feeding operations (CAFOs) permitted under state and federal regulations with species and maximum number of animals permitted | OH EPA | | TBD |
| Harmful Algal Bloom Advisories | Provides site, date, monitoring results and type of advisory (recreational, public or health) for recent HAB advisories | OH EPA | | Monday through Friday |

(continued)

**(Continued)**

| | | | | |
|---|---|---|---|---|
| Individual NPDES Permits | NPDES permit points for individually permitted facilities (i.e., those not covered under a general permit) | OH EPA | | |
| OSIP II Imagery | 6 inch to 1-foot resolution aerial imagery of Ohio collected from 2011 through 2014 (partial state coverage). | OH OIT | Web Service, GeoTIFF, MrSID, Esri GRID and ASCII GRID | Irregular |
| OSIP I Imagery | 1-foot resolution aerial imagery of Ohio collected from 2006 through 2010 (partial state coverage). | OH OIT | Web Service, GeoTIFF, MrSID, Esri GRID and ASCII GRID | Irregular |

## ILLINOIS

| Illinois Dataset | Description | Source | Access Format | Update Frequency |
|---|---|---|---|---|
| All Coal Mines | Shapefiles of coal mine points and polygons, including active mines, as of April 2014. | ISGS | Shapefile | TBD |
| Active Landfills | Locations of active landfills | IL EPA | Shapefile | TBD |
| Concentrated Animal Feeding Operations | Locations of CAFOs | IL EPA | Tabular | TBD |
| Leaking Underground Storage Tanks* | The Leaking Underground Storage Tank Incident Tracking ("LIT") database identifies the status of all Illinois LUST incidents reported to the Illinois Emergency Management Agency ("IEMA") and to the Illinois Environmental Protection Agency. | IL EPA | MS Access DB | Daily |
| NPDES Permits | Locations of NPDES permitted facilities | IL EPA | Shapefile | TBD |
| IL DOT Orthophotography | 6 inch to 1-foot aerial imagery for the State of IL, excluding the 7 Chicago Metropolitan Area counties, collected in 2011. | IL DOT | GeoTIFF and JPEG 2000 | Irregular |

## INDIANA

| Indiana Dataset | Description | Source | Access Format | Update Frequency |
|---|---|---|---|---|
| Mines – Surface | Locations of surface coal mines | INGS | Shapefile | TBD |
| Active Industrial Mineral Operations (2010) | Locations of operations for industrial minerals producers that were active in 2010 in Indiana. These operations extract and (or) produce the following commodities: cement, clay and shale, crushed stone, dimension limestone, dimension sandstone, gypsum, lime, peat, and sand and gravel. | INGS | Web Service and Shapefile | TBD |
| Best Available Imagery (2010) | Best imagery from 2005-2011 at a resolution of 1 foot per pixel or better. | Indian Spatial Data Portal | Web Service and Raster | TBD |
| Pipelines Oil Gas (1988) | Location and extent of known natural gas, crude oil, and refined products pipelines in Indiana. | INGS | Web Service and Shapefile | TBD |
| Institutional Control Sites (2015) | When any amount of contamination above a residential closure level is left on a property, a legal measure called an Institutional Control (IC) may be needed. An IC protects human health and the environment by restricting property activity, use, or access. Attributes include facility names and addresses, the type(s) of use-restrictions and IC's, and a URL that links directly to a digital copy of the IC document. | INDEM | Web Service and Shapefile | TBD |
| Confined Feeding Operations (2013) | Shows swine, chicken, turkey, beef or dairy agribusinesses that have large enough numbers of animals that IDEM regulates for environmental concerns, as defined by IC 13-18-10 of the Indiana Code. Provided by personnel of the Indiana Department of Environmental Management, Office of Land Quality. Data are current as of June 25, 2013. | INDEM | Web Service and Shapefile | TBD |

(continued)

**(Continued)**

| | | | | |
|---|---|---|---|---|
| Composting Facilities | Shows leaf, limb, or grass collection sites where a compost product is created, as defined by IC 13-20-10 of the Indiana Code. | INDEM | Web Service and Shapefile | TBD |
| Waste Disposal Storage Handling | Shows waste site locations for the disposal, storage, and handling of solid and hazardous waste | INDEM | Web Service and Shapefile | TBD |
| Industrial Sites (2013) | Shows the locations of access points to industrial waste site locations, along with unique identifiers for each location. | INDEM | Web Service and Shapefile | TBD |
| Landfill Boundaries (2014) | Shows boundaries for open dump sites, approved landfills, and permitted landfills in Indiana. | INDEM | Web Service and Shapefile | TBD |
| Septage Sites | Shows septage waste site locations. Septage Sites are permitted septage (septic tank waste) sites where the waste is land applied | INDEM | Web Service and Shapefile | TBD |
| Tire Sites | Tire Sites contain tires either for processing, for storage, or transport, as well as some illegal tire dumps | INDEM | Web Service and Shapefile | TBD |
| Treatment Storage Disposal Sites | Treatment Storage and Disposal Facilities (TSD) are facilities that may treat, store, and (or) dispose hazardous waste | INDEM | Web Service and Shapefile | TBD |
| Water NPDES Facilities | State-permitted wastewater facilities and associated information such as the name of the facility, contacts, and a variety of mailing addresses | INDEM | Web Service and Shapefile | TBD |
| Water NPDES Pipe Locations | Shows National Pollutant Discharge Elimination System (NPDES) Program pipe locations | INDEM | Web Service and Shapefile | TBD |
| Abandoned Quarries | Shows abandoned quarry locations that have locational information that is adequate to determine coordinate values. | INGS | Web Service and Shapefile | TBD |

(continued)

**(Continued)**

| | | | | |
|---|---|---|---|---|
| Mineral Production Sites (2010) | Locations of industrial mineral operations that were active in 2010 in Indiana. These operations extracted and (or) produced the following commodities: cement, clay and shale, crushed stone, dimension limestone, dimension sandstone, gypsum, lime, peat, slag, and sand and gravel. | INGS | Web Service and Shapefile | TBD |
| Petroleum Wells | Locations of all known petroleum test wells, as contained in the Indiana Geological Survey Petroleum Database Management System (PDMS) | INGS | Web Service and Shapefile | TBD |
| Ethanol Production Facilities | Locations of ethanol production facilities in Indiana. Attributes include the name and address of the facility, and information regarding ethanol production capacity. | INGS | Web Service and Shapefile | TBD |
| Land Parcels | Land parcels maintained by county agencies, as part of the Indiana Data Sharing Initiative | INDHS, INOT, INGIC | Web Service and Shapefile | TBD |

## KENTUCKY

| Kentucky Dataset | Description | Source | Access Format | Update Frequency |
|---|---|---|---|---|
| Kentucky Geological Survey Oil and Gas Wells | Oil and Gas well in Kentucky for which there are longitude and latitude coordinates | KGS | Shapefile | Weekly |
| Water Treatment Plants | Location of Water Treatment Plants in Kentucky | Kentucky Infrastructure Authority | Shapefile | TBD |
| Water Tanks | Location of Water Tanks in Kentucky | Kentucky Infrastructure Authority | Shapefile | TBD |
| Wastewater Treatment Plants | Location of Wastewater Treatment Plants in Kentucky | Kentucky Infrastructure Authority | Shapefile | TBD |
| Sewer Lines | Location of Sewer Lines in Kentucky | Kentucky Infrastructure Authority | Shapefile | TBD |
| Wastewater Treatment Plant Outfalls | Location of Wastewater Treatment Plant Outfalls in Kentucky | Kentucky Infrastructure Authority | Shapefile | TBD |
| Lift Stations | Location of Lift Stations in Kentucky | Kentucky Infrastructure Authority | Shapefile | TBD |
| Package Treatment Plants | Location of Package Treatment Plants in Kentucky | Kentucky Infrastructure Authority | Shapefile | TBD |
| Kentucky Inter-System Operational Permit Points | KY Division of Water permits to transfer wastewater from one system to another for treatment. | Kentucky Infrastructure Authority | Shapefile | TBD |

## PENNSYLVANIA

| Pennsylvania Dataset | Description | Source | Access Format | Update Frequency |
|---|---|---|---|---|
| Storage Tank Locations* | Aboveground and underground storage tanks regulated under Chapter 245 pursuant to the Storage Tank and Spill Prevention Act. Storage tanks currently contain, have contained in the past, or will contain in the future, petroleum or a regulated hazardous substance. Above ground storage tanks that either contain hazardous materials or have a capacity greater than 21000 gallons are removed. Summary information is available online. Detailed information (including specific contents and capacities) is available via request. | PA DEP | MS Excel, Web Service and Shapefile | Quarterly |
| Coal Pillar Locations (Oil and Gas) | Coal Pillar Locations are pillars of coal that must remain in place to provide support for an oil and gas well on the surface. | PA DEP | Web Service and Shapefile | Quarterly |
| Commercial Hazardous Waste Operations | Locations of permitted commercial hazardous waste operations | PA DEP | Web Service and Shapefile | TBD |
| Conservation Wells (Plugged) | The plugged conservation well layer identifies the permitted surface location of oil and gas conservation wells that have been plugged. | PA DEP | Web Service and Shapefile | Weekly |

(continued)

**(Continued)**

| | | | | |
|---|---|---|---|---|
| Conservation Wells | The conservation well layer identifies the permitted surface location of oil and gas conservation wells that have not been plugged. These include active, regulatory inactive, orphaned, and abandoned wells. A conservation well is any well which penetrates the Onondaga horizon, or in those areas in which the Onondaga horizon is nearer to the surface than thirty-eight hundred feet, any well which exceeds a depth of thirty-eight hundred feet beneath the surface. | PA DEP | Web Service and Shapefile | Weekly |
| Oil and Gas Encroachment Locations | Includes information about the following PA DEP Encroachment Location for Oil & Gas sub-facility types: Bridge, Culvert, Stream Bank Protection, and Intake structure. | PA DEP | Web Service and Shapefile | Quarterly |
| Oil and Gas Locations (Conventional/Unconventional) | All the Oil & Gas Wells in Pennsylvania that the DEP has locational information on. | PA DEP | Web Service and Shapefile | Daily |
| Oil and Gas Locations | Includes information about the following PA DEP Oil and Gas Location sub-facility types: Land Application, Well, and Pit. Some sub facility types are not included in this layer due to security policies. | PA DEP | Web Service and Shapefile | Quarterly |

(continued)

**(Continued)**

| | | | | |
|---|---|---|---|---|
| Historic Oil and Gas Locations | This dataset was created using georeferenced mine maps of various/unknown accuracy and various/unknown coordinate systems to various base maps, including but not limited to USGS topographic maps and PAMAP aerial photography. The locations were then digitized using the georeferenced mine maps. These locations are provided for informational purposes only and should not be sole means of decision making and are in no way a substitute for actual field observations. | PA DEP | Web Service and Shapefile | TBD |
| AML Inventory Points | Approximate location of Abandoned Mine Land Problem Areas containing public health, safety, and public welfare problems created by past coal mining. It is a subset of data contained in the Office of Surface Mining (OSM) Abandoned Mine Land Inventory. | PA DEP | Web Service and Shapefile | Continually |
| Captive Hazardous Waste Operations | Location of captive hazardous waste operations statewide | PA DEP | Web Service and Shapefile | TBD |
| Coal Mining Operations | Includes information about the following PA DEP Coal Mining Operation sub-facility types: Deep Mine, Discharge Point, Mineral Preparation, NPDES Discharge Point, Post Mining Treatment, Refuse Disposal Facility, Refuse Reprocessing Facility and Surface Mine. | PA DEP | Web Service and Shapefile | Quarterly |
| Coal Pillar Locations (Mining) | Coal Pillar Locations are pillars of coal that must remain in place to provide support for a coal mine. | PA DEP | Web Service and Shapefile | Quarterly |
| Industrial Mineral Mining Operations | Includes information about the following PA DEP Industrial Mineral Mining program sub-facility types: Deep Mine, Discharge Point, Mining Stormwater GP and Surface Mine. | PA DEP | Web Service and Shapefile | Quarterly |
| Active Underground Mining Permit Boundaries | This dataset contains the digitized underground permit area boundaries of the active underground bituminous mines in Pennsylvania based from the Location Maps submitted with the permit applications and permit renewal applications. | PA DEP | Web Service and Shapefile | As Needed |

(continued)

**(Continued)**

| | | | | |
|---|---|---|---|---|
| AML Inventory Polygons | This data set portrays the approximate location of Abandoned Mine Land Problem Areas containing public health, safety, and public welfare problems created by past coal mining. It is a subset of data contained in the Office of Surface Mining (OSM) Abandoned Mine Land Inventory. This layer identifies AML Polygons representing specific areas too large to be represented by points within the entire, AML Inventory Site, examples include AML dangerous highwalls. | PA DEP | Web Service and Shapefile | Continually |
| Anthracite Coal Mine Permit Update | Anthracite Coal Mine Permits. | PA DEP | Web Service and Shapefile | As Needed |
| Anthracite Coal Refuse Update | Anthracite Coal Refuse and Prep plant. | PA DEP | Web Service and Shapefile | As Needed |
| Bituminous Coal Mine Permit Update | Bituminous Coal Mine Permits. | PA DEP | Web Service and Shapefile | As Needed |
| Bituminous Coal Refuse Update | Bituminous Coal Refuse. | PA DEP | Web Service and Shapefile | As Needed |
| Digitized Mine Areas | This dataset includes the Mined Out Areas of the various coal seams in Pennsylvania. The information in this dataset was gathered from mine maps at these various locations so that the data can be readily available to environmental professionals. | PA DEP | Web Service and Shapefile | Monthly |
| Industrial Mineral Mine Permit Update | Industrial Minerals mine permits. | PA DEP | Web Service and Shapefile | As Needed |
| Longwall Mining Panels by Status | This dataset identifies the footprint of these longwall panels relative to the surface. | PA DEP | Web Service and Shapefile | Daily |

(continued)

**(Continued)**

| | | | | |
|---|---|---|---|---|
| Mine Drainage Treatment / Land Recycling Project Locations | Locations of mine cleanup projects including abandoned coal refuse pile reclamation, abandoned deep mine reclamation, abandoned mine drainage treatment, abandoned oil and gas well reclamation, abandoned surface mine reclamation and internal monitoring points | PA DEP | Web Service and Shapefile | TBD |
| Municipal Waste Operations | Location of municipal waste operations | PA DEP | Web Service and Shapefile | TBD |
| Water Pollution Control Facilities | Locations of facilities related to the Water Pollution Control Program, including agricultural activities, compost/processing, conveyance systems, discharge points, groundwater monitoring points, internal monitoring points, land discharges, manure management, outfall structures, pipelines or conduits, pump stations, storage units and treatment plants (sewage or industrial wastewater) | PA DEP | Web Service and Shapefile | TBD |
| Water Resources | Locations of facilities related to the Water Use Planning Program, including discharges, ground water withdrawals, interconnections, storage, surface water withdrawals and water allocation permits. | PA DEP | Web Service and Shapefile | TBD |
| Tier II Reports – Bucks County | Received by request from the Bucks County Emergency Management Agency. Tier II reports document the type and amount of hazardous materials stored at a facility. | LEPC | MS Excel | Continually |
| Downstream Notifications | Letters sent by above ground storage tank owners to water utilities within 20 downstream miles. The letters include the storage capacity and contents of each tank. Some letters provide location information, such as maps, site sketches or descriptions. | Aqua | PDF | Continually |

## WEST VIRGINIA

| West Virginia Dataset | Description | Source | Access Format | Update Frequency |
|---|---|---|---|---|
| Above Ground Storage Tanks | Locations and points of contact for WV facilities with above ground storage tanks. This is a brief and largely incomplete dataset, seemingly derived from information contained in Industrial Multi-Sector Permits. | WV DEP | Web Service and Shapefile | TBD |
| Coal Refuse Fills | Locations of coal refuse fills | WV DEP | Shapefile | Daily |
| Underground Mining Limits | This data represents the boundary within which underground mining may occur for individual permits. They do not include exclusion zones within a boundary where mining is prohibited. They also do not depict actual workings, either planned, or existing. The WVDEP does not receive maps of underground works on a routine basis, but uses mining limits as an indicator of mining that may be occurring in an area. | WV DEP | Shapefile | Daily |
| Oil and Gas Permits | Locations, points of contact, permit information and links to online oil and gas well logs for oil and gas permits in WV. | WV DEP | Web Service and Shapefile | TBD |
| NPDES Concentrated Animal Feeding Operations | Locations of permitted CAFOs | WV DEP | Web Service | TBD |
| NPDES Ind POTW (Publicly Owned Treatment Works – Municipal) | Permitted municipal POTW locations. | WV DEP | Web Service | TBD |
| NPDES Individual (Industrial Facilities) | Locations of individually permitted NPDES facilities | WV DEP | Web Service | TBD |

(continued)

**(Continued)**

| | | | | |
|---|---|---|---|---|
| NPDES Sanitary Landfill Leachate | Locations of sanitary landfill leachate systems with NPDES permits | WV DEP | Web Service | TBD |
| NPDES Solid Waste Landfill | Locations of NPDES permitted solid waste landfills | WV DEP | Web Service | TBD |
| NPDES Private Sewage Treatment Facilities | Locations of NPDES permitted other (private sewage treatment) facilities | WV DEP | Web Service | TBD |
| NPDES Municipal POTWs | Locations of permitted municipal NPDES POTW facilities | WV DEP | Web Service | TBD |
| NPDES Sewage Collection Discharge to POTW | Locations of permitted NPDES discharge from sewage collection systems to POTWs | WV DEP | Web Service | TBD |
| Parcels | Boundaries and points of contact for WV land parcels (partial state coverage). | WVU | Web Service and Shapefile | TBD |
| Mining Permits | Locations, points of contact and permit information for mining permits in WV. | WV DEP | Web Service and Shapefile | TBD |
| Mining Permit Boundaries | WV Mining permit boundaries, points of contact, permit information and links to online mining permit information. | WV DEP | Web Service and Shapefile | TBD |
| Potential Sources of Contamination in Zones of Critical Concern | Locations, names, descriptions, industry categories and threat to surface water rankings for PSCs, as identified by WV BPH, in the ZCCs. | WV BPH | MS Access DB | Irregular |
| Zones of Critical Concern (Regulatory) | The current regulatory boundaries for PWS zones of critical concern. | WV BPH | MS Access DB | Irregular |
| Electricity Power Plants | Locations of WV power plants. | WV DOT | Web Service and Shapefile | TBD |

(continued)

**(Continued)**

| | | | | |
|---|---|---|---|---|
| Tier II Reports | Received by request from the Kanawha Putnam Emergency Planning Committee and the Summers County Office of Emergency Management. Tier II reports document the type and amount of hazardous materials stored at a facility. | LEPCs | MS Excel | Continually |
| WV Sheriff's Association Aerial Imagery | In 2010-11 the oblique aerial photography company Pictometry captured 12" pixel resolution for the entire state and 4" pixel resolution for populated areas. In 2012 re-flights of the 4" imagery were done for select areas. | WVU | Web Service | Irregular |

## USA

| National Dataset | Description | Source | Access Format | Update Frequency |
|---|---|---|---|---|
| TRI Facilities | The Toxic Release Inventory System (TRIS) contains information about more than 650 toxic chemicals that are being used, manufactured, treated, transported, or released into the environment by certain industry groups, including information about waste management and pollution prevention activities. | US EPA | Web Service | Continually |
| RCRA (Active) Facilities | The Resource Conservation and Recovery Act Information System (RCRAInfo) tracks many types of information about generators, transporters, treaters, storers, and disposers of hazardous waste. | US EPA | Web Service | Continually |
| NPDES Facilities | Location and facility identification information from EPA's Facility Registry System (FRS) for the subset of facilities that link to the Permit Compliance System (PCS) or the National Pollutant Discharge Elimination System (NPDES) module of the Integrated Compliance Information System (ICIS). PCS tracks NPDES surface water permits issued under the Clean Water Act. This system is being incrementally replaced by the NPDES module of ICIS. | US EPA | Web Service | Continually |
| OIL Facilities | The Oil database contains information on Spill Prevention, Control, and Countermeasure (SPCC) and Facility Response Plan (FRP) subject facilities to prevent and respond to oil spills. | US EPA | Web Service | Continually |

(continued)

**(Continued)**

| | | | | |
|---|---|---|---|---|
| CERCLIS Facilities | The Comprehensive Environmental Response, Compensation, and Liability Information System (CERCLIS) is EPA's inventory of abandoned, inactive, or uncontrolled hazardous waste sites regulated under the Comprehensive Environmental Response, Compensation, and Liability Act (CERCLA). It records information about all aspects of hazardous waste sites from initial discovery to listing on the National Priorities List (NPL). | US EPA | Web Service | Continually |
| Airports | Aircraft landing facilities in the US with physical and operational characteristics. Data are derived from FAA's National Airspace System Resource Aeronautical Data Product | Federal Aviation Administration | Shapefile | Annual |
| Alternative Fuels | Locations of fueling facilities offering fuel other than gasoline in the US | US DOE | Shapefile | Annual |
| Dams Over 50 ft | Locations of large dams over 50 feet tall in the US | US ACE | Shapefile | Annual |
| National Bridge Inventory | Locations of bridges on US highways, state and county roads and federal lands. | Federal Highway Administration | Shapefile | Annual |
| US Army Corps of Engineers Ports | Physical information about commercial facilities at the principal U.S. Coastal, Great Lakes and Inland Ports. The data consist of listings of port areas' waterfront facilities, including information on berthing, cranes, transit sheds, grain elevators, marine repair plants, fleeting areas, and docking and storage facilities. | US ACE | Shapefile | Annual |
| Railroad Grade Crossings | Locations of railroad grade crossings | Federal Railroad Administration | Shapefile | Annual |

(continued)

**(Continued)**

| | | | | |
|---|---|---|---|---|
| Military Installations | The dataset depicts the authoritative boundaries of the most commonly known Department of Defense (DoD) sites, installations, ranges, and training areas in the United States and Territories. These sites encompass land which is federally owned or otherwise managed. | Military Surface Deployment and Distribution Command – Transportation Engineering Agency | Shapefile | Annual |
| Hazardous Material Routes | Hazardous material routes developed using the 2004 First Edition TIGER/Line files. The routes are described in the National Hazardous Material Route Registry (NMHRR). | Federal Motor Carrier Safety Administration | Shapefile | Annual |
| Railway Network | Spatial network of rail lines within the United States | Federal Railway Administration | Shapefile | Annual |
| U.S. Army Corps of Engineers Navigable Waterway Network | Comprehensive network of navigable waterways. | US ACE | Shapefile | Annual |
| Roads National Database | All roads in the nation | US Census Bureau | ESRI FGBD | Annual |
| NLCD 2011 Land Cover | 2011 land cover for the coterminous US | USGS | Raster | TBD |
| Power Plants | Generator-level specific information about existing and planned generators and associated environmental equipment at electric power plants with 1 megawatt or greater of combined nameplate capacity. | Energy Information Administration | Tabular | Annual |
| National Hydrography Dataset | A digital vector dataset that contains lakes, ponds, streams, rivers, canals, dams, watersheds *by USGS Hydrologic Unit Code (HUC)) and stream gages. | USGS | Web Service | Irregular |
| Bing Aerial Imagery and Basemaps | A collection of multi-sourced, high-resolution aerial imagery, basemaps. | Microsoft | Web Service | Continually |
| Google Aerial Imagery and Basemaps | A collection of multi-sourced, high-resolution aerial imagery, basemaps. | Google | Web Service | Continually |
| TIGER/Line Shapefiles | Boundaries, roads, address information, water features, and more. | US Census Bureau | Shapefiles and DBFs | Annually |

# APPENDIX B
# TIME OF TRAVEL ANALYSIS

The current time of travel analysis incorporated into the WaterSuite geospatial platform assumes the user will want to calculate the time of travel for a spill that has occurred within the area of interest, defined geographically using the agreed upon Zone of Critical Concern (ZCC). Spills sites are likely to be identified as either locations within the ZCC already identified as Points of Interest (‘POI’) or other locations within the ZCC that are not specifically identified as a POI.

For locations within the ZCC, the workflow options are as follows:

- A user clicks on a POI marker, identifying the POI location as the spill site and initiating a time of travel calculation from the latitude/longitude information stored in the POI’s attributes.
- A user clicks on any area within the ZCC to identify the spill site. The latitude/longitude information is determined by WaterSuite or the ESRI mapper, and this information is used to initiate a time of travel calculation.
- A user clicks on an existing POI to access pre-calculated travel times based on the reference gage’s historical streamflow across a range of percentiles. The time of travel values for these percentiles are stored within WaterSuite and remain available regardless of the current status of the EPA’s WATERS service and the USGS’s REST service, whose use in the time of travel calculation is outlined below. This allows the user to compare on demand travel time calculations with baseline travel times representative of various historical flows when both web services are running, while providing some estimate of travel time if one or both web services are unresponsive. If a user knows the current streamflow near the intake and how this flow compares to historical values, they may accordingly be able to roughly estimate travel time.

## METHODOLOGY

The calculation of a spill’s time of travel to the downstream drinking water intake implements the work of Jobson and others at the USGS (Jobson, 1996; Ries, 2006). These equations predict travel times to the leading edge of a plume, peak concentration and 10% of peak concentration (i.e., trailing edge) at the probable and maximum probable velocities.

The first step in the calculation of the travel time is the identification of a *reference gage*. For each implementation, each zone of concern or area of interest, a nearby USGS gage was identified as an appropriate source of current streamflow conditions. The appropriate donor gage was determined by compiling and screening a list of nearby¸ active USGS gages whose contributing areas share similar attributes to the contributing area of the project’s drinking water intake. Attributes taken into consideration during the screening may include: total distance from the drinking water intake, total contributing area, ecoregion classification (level III), basin compactness, dominant geology (Hunt, 1979; Reed and Bush, 2001), and land use and land cover types within the gage’s contributing area. Additional attributes currently under study for inclusion in the selection process include: average temperature and precipitation (PRISM dataset), latitude and longitude of the watershed’s centroid, average baseflow index of the watershed, number of dams present on the main channel, Strahler stream order, percentage of different Hydrologic Soil Groups within the watershed, and watershed slope.

With a reference gage identified, the time of travel calculation proceeds using the combined results of querying several web services hosted by the EPA and USGS. The EPA's WATERS ('WATERS') web services (USEPA, 2014) provide "REST like-access to the WATERS database services." These database services provide access to the latest enhanced edition of the National Hydrography Dataset (NHDPlus v2.1), allowing users to perform complex analyses based on querying the underlying data stored within the NHD. The USGS's REST ('UREST') service (USGS, 2015) provides access to almost all available data (real time, historical statistics, site information, etc.) for any gage maintained by the USGS. Both web services provide significant hydrologic analysis capabilities of which only a small subset is used in this methodology.

The time of travel tool allows for on demand calculation of travel times using site specific flow data anywhere within the drinking water intake's upstream contributing area. Additionally, travel times have been pre-calculated for each point of interest using the reference gage's historical streamflow across a range of percentiles; specifically, the 0.5$^{th}$, 1$^{st}$, 5$^{th}$, 10$^{th}$, 25$^{th}$, 50$^{th}$, 75$^{th}$, 90$^{th}$, 95$^{th}$, 99$^{th}$ and 99.5$^{th}$ percentiles. The travel times from each point of interest for these percentiles are stored within the GIS tool and available regardless of the current status of the WATERS service and the UREST service, whose use in the TOT calculation is outlined below. This allows the user to compare on demand travel time calculations with baseline travel times representative of various historical flows when both web services are running, while providing some estimate of travel time if one or both web services are unresponsive. If a user knows the current streamflow near the intake and how this flow compares to historical values, they may accordingly be able to roughly estimate travel time.

The TOT calculation, in both the on demand and pre-calculated scenarios proceeds as follows:

1. Determine the spills' likely flow path to the stream.
    a. Method: WATERS Point Indexing web service (Raindrop Tool indexing path analysis) - the spill's likely flow path is identified.
    b. Returns:
        i. Total flow path distance
        ii. An indexed path representing the flow path location and geometry
        iii. The flow path end point in the NHD network, identified using the reach code, reach measure, and comid – a unique id for each NHD network member

2. Determine the contributing areas of the DW intake and of the location where the spill enters the stream network.
    a. Method: WATERS Navigation Delineation web service - the upstream contributing area for the specified point is calculated.
    b. Returns:
        i. Total contributing area in km$^2$
        ii. Contributing area geometry

3. Determine the distance the spill will travel along the reach (NHD network) to the DW intake.

a. Method: Waters Navigation web service (point to point option) - the NHD network is traversed from the location where the spill entered the network to the DW intake
b. Returns:
    i. Total flow path distance
    ii. The flowlines IDs for network members in the traversal path
    iii. The traversal path geometry

4. Determine current instantaneous and mean annual streamflow measurements based on the identified reference gage measurements.
    a. Method: UREST web service - the reference gage's streamflow measurements are retrieved.
    b. Returns:
        i. Mean annual streamflow measurement.
        ii. Instantaneous streamflow measurement.
        iii. Using these measurements, we calculate the following based on the methodologies of Jobson (1996) and Ries (2007):
            1. Instantaneous streamflow for the average reach contributing area
            2. Mean annual for the average reach contributing area
            3. Instantaneous streamflow at intake
            4. Mean annual streamflow at intake

5. Determine travel times
    a. Method: Calculations follow the methodology outlined in *The Prediction of Traveltime and Longitudinal Dispersion in Rivers and* Streams (Jobson, 1996). Some of the more important values calculated include:
    b. Returns:
        i. Probable and maximum probable velocities.
        ii. Travel times for the leading edge ($T_l$), peak concentration ($T_p$), and time for the concentration to reach 10% of the peak concentration ($T_{D10}$), for both the probable and maximum probable velocities.

## ASSUMPTIONS AND LIMITATIONS

Inherent in using the Jobson methodology are several assumptions about river mixing and flow characteristics that simplify the modeling of pollutant and tracer transport and allow for the use of regression equations to predict travel time and concentration. These include the assumptions that vertical and lateral dispersion fully take place upstream of the measurement location, so that longitudinal dispersion of the pollutant is the dominant mixing process being measured or predicted. Vertical mixing is said to occur rather rapidly in streams while lateral mixing is thought to be complete within the first several kilometers downstream of the entry point (Jobson, 1996). The pollutant concentration is then modeled as a scalene triangle, with the leading edge of the pollutant plume arriving at approximately 89% of the travel time to the peak concentration. The relationship between velocity and discharge is different for each stream, but the generally assumed to take the form $V = KQ^a$, with the constant $K$ derived for each reach and exponent $a$ often set to 0.34 based on the investigation of this relationship by researchers (Jobson, 1996; Jobson, 1989). Finally, there is an assumption that additional contributory water sources, such as tributaries, outfalls from treatment plants, etc., do not significantly alter the properties of the modeled streamflow as to disrupt the process of longitudinal mixing and render predictions impossible.

In determining the discharge (instantaneous or otherwise) at the spill location and the intake, as well as at the midpoint of the pollutant's traversal along the stream, an assumption is made about the relationship between the upstream contributing area and the discharge at the points of interest. Both Jobson (1996) and Ries (2007) use a scaling equation of the form $\mathbf{Q_u = (A_u/A_g)^b Q_g}$, with the exponent *b* set to a value of 1 Jobson's work with different values chosen in the NSS program based on the basin or geographic area of interest. In the NSS calculator, this equation is not used for estimating instantaneous flows in ungaged basins; this may be due to stricter assumptions and requirements on the transference of characteristics between basins for certain temporal domains (e.g., hourly or daily flows versus annual means). However, the equation is used in the Jobson report when estimating instantaneous streamflows, with the exponent b set to 1; this is likely due to the more lax assumptions about the transference of data and characteristics between basins being necessary to complete the TOT calculations.

Recognized limitations in the current implementation of the tool consist of challenges in automatically identifying appropriate stream donor gages, concerns about the availability of web services providing data used in the travel time calculation, and inability for the methodology to correctly handle estuaries, reservoirs, and dammed river sections. Additionally, the tool does not currently track the travel time necessary for the spill to follow its predicted overland flow to the stream; this will be addressed in future updates, with current proposed solutions focused on implementing kinematic wave equations for modeling shallow sheet flow over a wide plane (using an implementation of Manning's equation). Related to this are improvements that could be made to the prediction of the overland spill path through the use of higher resolution digital elevation models (DEMs); current overland flow path predictions are based on flow direction grids calculated from nationally available 30m DEMs. Similarly, the delineation of upstream contributing areas relies on a combination of 30m DEMs and the medium resolution National Hydrography Dataset (NHD), which is accessed through the web services discussed above. Referencing the high resolution NHD dataset and higher resolution DEMs will refine upstream contributing area delineations and therefore improve streamflow predictions; this may be possible in the future depending on improvements to the web services accessed by the tool or through incorporation of these datasets within WaterSuite and internalization of the web services' capabilities.

# APPENDIX C
# PROJECT SCOPING DOCUMENT

Water Research Foundation Project 4581: Using Geographic Information System (GIS) to Identify and Catalog Potential Contaminant Stored in Tanks Upstream of Drinking Water Intakes

Prepared by:
Jeffrey Rosen, Margaret Kearns & Tim Bartrand
Corona Environmental Consulting
6 Old Country Way
Scituate, MA 02066

January 8, 2014

## PROJECT DESCRIPTION

The Water Research Foundation's Project Number 04581, "Using Geographic Information Systems (GIS) to Identify and Catalog Potential Contaminants Stored in Tanks Upstream of Drinking Water Intakes," will deliver a methodology for identifying and characterizing chemical storage facilities and other potential source water contamination sites and a functional software tool that provides access to these data for the emergency response community and the drinking water community. As part of this effort, a data gaps analysis will identify those datasets that are difficult to access or otherwise unavailable. The project also includes application of the methodology and toolset to four pilot sites in West Virginia, Pennsylvania and Illinois.

Project 4581 will be conducted in 4 steps: (i) scoping; (ii) preliminary toolset development; (iii) requirements analysis; (iv) pilot studies. Generally, scoping, the topic of this report, entails determining ***what*** the toolset will do and what pilot studies will entail. Requirements analyses will determine ***how*** the system will work and what functionality it will have. The requirements analysis will be completed with project partners in early February 2015.

## PURPOSE AND REPORT ORGANIZATION

This report summarizes discussions at four webinars conducted in December 2014 to obtain input on the scope (generic features and capabilities) of the information system as well as goals and data sources for the four pilot studies. The report presents the scoping effort methodology and background materials developed for the webinars, summarizes discussions (results) about generic system features and pilot study objectives, then provides a summary of scoping webinar findings. Some needs and opportunities that were discussed are beyond the scope of the current effort, but are summarized in the Results. The Summary provides an outline of the scope of work for the four pilot sites and overall goals for the information system that will be developed for this project.

## METHODS

The webinars were conducted as a facilitated discussion and discussions captured in writing by CEC staff for all of the webinars and recorded for three out of the four webinars. An agenda for the meeting is presented in Appendix A with meeting participants listed in Appendix B. The project team attending the webinars included representatives from the Water Research Foundation (WRF), the American Water Works Association, West Virginia-American Water Company, Aqua America and Corona Environmental Consulting (CEC). The webinars were also attended by some members of the project advisory committee, the project steering committee and WRF project managers. Utility representatives and staff of the Metropolitan Washington Council of Governments participated in the scoping discussions and will participate in the requirements analysis in anticipation of a future project using the same methods and information system in the Potomac River watershed.

The goal of the four scoping webinars was to discuss the generic features, capabilities and primary data sources for the information system and define the pilot study goals. The webinars also included a demonstration of the current state of the information system or updates to the system since participants had last viewed it. The first two webinars focused on the two pilot sites for Aqua America (Shenango and Kankakee plants), the third focused on the West Virginia-American Water Company pilot sites (Huntington and Weston plants) and the fourth meeting focused on obtaining input from drinking water utilities and staff at the Metropolitan Washington Council of Governments.

### Annotated List of Datasets

To inform discussions about available data for each pilot project, a list of data resources and fact sheets for key data sets in each relevant state were distributed to the project team (Appendix C). The focus area for data collection included all states within the upstream drainage area of each of the four pilot site intake locations. Potential new data sources that were discussed during the scoping webinars are listed in the Scoping Webinar Results section.

CEC began the data collection process by developing a list of types of data that were likely to be available and useful for characterizing acute threats to source waters (Table 7). These included data sources for threats as well as a set of navigational and imagery data. Many of the relevant federal datasets were documented and obtained through prior work with the West Virginia-American Water Company and Aqua America. Some data sets of national extent (federal data sets) were obtained from federal agencies, though, searches focused on collection of state data.

CEC staff began the research for state and local data with an internet search to identify dataset owners and major GIS data clearinghouses in each state within the focus area. In CEC's experience, state data clearinghouses maintained by state agencies (departments of environmental protection departments of transportation, natural resources divisions, universities or other key agencies) are the best sources for data at the level of detail required for identifying water intake vulnerabilities. Phone calls to state agencies, universities, local authorities and other groups were made to obtain access to data in GIS format and to better understand the structure and content of the datasets. In addition to this structured approach, CEC also relied on word of mouth from project partners, state agency and university staff and others to identify unique and relevant data sources.

Searches were conducted to obtain the key data sets presented in Table 7. For each dataset, basic data content, ownership, update frequency and contact information were recorded.

The list of datasets that were identified through this process was distributed to scoping webinar participants prior to the meetings and served as a starting point for data availability discussions for each pilot site.

**Desired datasets to characterize acute source water threats**

| Desired Dataset |
| --- |
| • Above ground storage tanks |
| • Sewage treatment plants |
| • Airports |
| • Power plants |
| • Underground storage tanks |
| • Concentrated animal feeding operations (CAFOs) |
| • Local businesses / NAICS codes |
| • Military facilities |
| • Agricultural chemical sales |
| • Known potential sources of contamination |
| • Hydraulic fracturing permits |
| • Discharge permits |
| • Industrial multi-sector permits |
| • Hazardous waste permits |
| • Oil and gas permits |
| • Mining permits |
| • EPCRA Tier II hazardous chemicals reports |
| • Emergency response plans |
| • High resolution aerial imagery |
| • Digital elevation models or LIDAR imagery |
| • Topography |
| • Hydrography |
| • Gaging stations |
| • Dams |
| • Roads |
| • Railroads |
| • Bridges |
| • Floodplains |
| • Parcels |

## SCOPING WEBINAR RESULTS

### Annotated List of Datasets

The initial search for datasets yielded the data summarized in Appendix C. Six national scale datasets describing different types of acute contamination threats were identified. The availability of state level data varied by state. The drainage areas for the four pilot sites include parts of West Virginia, Pennsylvania, Illinois, Indiana and Ohio. For most states, datasets are available that provide information about oil and gas wells, mines, above and below ground storage tanks, railroads and airports. Imagery data is also widely available, although the coverage and resolution varies. Other types of data were more difficult to locate, such as power plants, sewage treatment plants, and military sites.

Some datasets will require additional effort to access or are currently unavailable to the project team. Access to EPA's Water Contaminant Information Tool (WCIT), for example, is only available to drinking water and wastewater utilities, state primacy agencies and federal officials. Other entities may be granted access by the EPA based on their need. Several other datasets require Freedom of Information Act (FOIA) requests or other special requests to gain access. These include Tier II hazardous chemicals reports, emergency response plans and some state above ground storage tank data. Furthermore, distribution of some elements of data sets is limited and the identity, location and volume of some chemicals may not be available due to security or proprietary concerns. For instance, chemicals that meet certain requirements may be claimed as "Trade Secrets" on Tier II hazardous chemical reports. The locations and specific chemical identities of the compounds are not publicly available. Data on chemicals of interest that are regulated under the Chemical Facility Anti-terrorism Standards (CFATS) act may also have limitations on distribution due to homeland security concerns. Access to information created under this program, termed chemical terrorism vulnerability information (CVI), may be granted to emergency responders or others with a need to know during an emergency but is otherwise only available to authorized users with a need to know. Other datasets with restricted or difficult access are noted in Appendix C. Data gaps and other useful datasets and data repositories were discussed at each of the scoping webinars and are discussed below.

### System Features

The general features of the prototype information system were demonstrated during each of the scoping webinars. Prototype features and functionalities include common GIS tasks, basic reporting, contaminant property search, aerial image datasets and links to oblique images and the ability to create new data points. Scoping session participants noted that the same general features could be used for wastewater treatment plant operations, though no specific projects were discussed. Participants in the scoping webinars also suggested several other features that would be useful for the pilot site investigations:

1. **User experience** - The user's experience of the information system should be carefully considered to be as intuitive and easy to understand as possible, particularly when interacting with the contaminant database. Operators must be able to quickly and easily understand the information and make decisions. A graphical dashboard was recommended for this purpose.

2. **Early access to contaminant data** - Participants requested that the contaminant database be accessible to the project group shortly before the requirements analysis begins in early February.
3. **Tablet operating system** - Access to the system via tablets will make it more useful for field staff.
4. **Changes over time** - The ability to track changes over time, such as points that have been reviewed by staff and notes about specific data features, is a key requirement.
5. **Automated alerts** - For changes in real-time or newly updated data, automated messages could alert staff to new developments.
6. **Creation of new datasets** - Participants requested the ability to add source water sampling points or other types of data points that are not part of an existing dataset.
7. **Existing data source connections** - At multiple scoping sessions, the requirement to be able to connect the new information system to existing LIMS or other databases or data streams was highlighted.
8. **Sub-watershed contaminant report -** It is important for users to have the ability to identify sources of contamination in an upstream sub-watershed based on water quality monitoring results downstream. This feature would generate a list of potential contaminants for a specific sub-basin.
9. **User-selected area report** - The ability to generate a report of features within a user-selected polygon was also requested.
10. **Road/rail/pipeline highlights** - The ability to highlight particular sections of road, rail or pipeline of particular concern was suggested.

The final set of detailed features that will be included in the information system will be determined during the requirements analysis in February.

**West Virginia American Water Pilot Sites**

The two pilot sites for West Virginia American Water are the Weston and Huntington facilities. The Weston source water intake is in north central West Virginia downstream of the Stonewall Jackson Lake Wildlife Management Area on the West Fork River. The Huntington plant has two intakes relatively close together on the Ohio River. One intake is upstream of the confluence of the Guyandotte and Ohio Rivers and one is downstream. Generally, the downstream intake is the plant's primary water source unless water quality for the Guyandotte River requires activation of the upstream intake. Because of the size and high concentration of industrial activities of the Ohio River, the combined zone of critical concern is much larger for the Huntington intakes than for the other pilot sites. Approximately half of this zone is in Ohio while the other half is in West Virginia. The zone of critical concern is based on ORSANCO's Zone 1 definition[1], which includes a quarter mile buffer from the river banks upstream 25 miles, or the five-hour travel time. The zone of critical concern also stretches one quarter mile downstream of the intake. Searches for above ground storage tanks and other key contaminant sites are focused on zones of critical

[1] West Virginia Department of Health and Human Resources, Bureau for Public Health, Office of Environmental Health Services, Source Water Protection Unit. June 2003. State of West Virginia Source Water Assessment and Protection Program Source Water Assessment Report, WVAWC Huntington District, Cabell County, PWSID: WV3300608.

concern because time to react to spills and other contamination events occurring in the zone of critical concern are short.

***Weston***

The areas of focus for this site were water quality at the Stonewall Jackson dam, oil and gas activities and the I-79 interstate road crossing upstream of the intake. The Stonewall Jackson Lake has an active marina and is a recreational destination managed by the US Army Corps of Engineers (ACOE). ACOE staff communicate with WV American Water staff when a large change in flow is planned, which helps WVAW staff anticipate changes in raw water quality (particularly turbidity). Water quality and stream flow data may be available from the ACOE and/or the golf course nearby, where several groundwater wells are located, but no water quality monitoring is conducted at the marina. At the plant intake, manganese levels can be high due to lake turnover. High manganese during turnover is observed in other reservoirs and likely related to cycling of manganese by microorganisms in or near reservoir sediments. The River Alert Information Network (RAIN) maintains basic pH, temperature and conductivity data that is used by West Virginia American Operators and should be incorporated into the new information system if possible. Oil and gas wells are common in this area, as are capped historical wells. Interstate 79 runs past the Weston plant atop a steep incline where several truck accidents have reportedly occurred. Spills in this stretch of highway can potentially be carried directly downhill and enter the water in the vicinity of the intake. Although the scoping group was unaware of specific sources of information about hazardous substances carried on highways, they noted that the Kanawha Putnam Emergency Planning Council (KPEPC) compiled a month long snapshot of the types of hazardous materials that were transported through Kanawha County. Similar studies may be available from other emergency planning committees.

The group also discussed several other issues. Cattle operations are common in this area but have not had a significant impact on source water quality in the past. A campground located between the Stonewall Jackson dam and the Weston intake has caused bacteriological issues in the past that were easily resolved. The plant has not had any problems with microcystin.

***Huntington***

The Huntington plant draws water from the Ohio River, which is a much larger system than the other three pilot sites. The scoping group noted that, in general, an assessment of vulnerabilities along the whole river may be of interest, though too involved to complete within the current budget. The focus areas for this site were above ground storage tanks, mobile sources (trains, ships and barges), oil and gas well activity, abandoned mines, industrial facilities, water quality monitoring and interstate coordination of data sharing and emergency communications. Industries, particularly those with above ground storage tanks, along the Ohio River are a focus area, and several West Virginia facilities with multiple above ground storage tanks were seen during the system demonstration. Mobile sources present a unique consideration for this intake because of the heavy commercial traffic on the Ohio River. The Ohio River Valley Water Sanitation Commission (ORSANCO) compiled a list of past shipping spills that may be useful, and Corona is investigating the possibility of using the Coast Guard's Automatic Identification System (AIS) to provide real-time locations of particular vessels. The coast guard system is likely to identify ships carrying hazardous cargo, although more detailed cargo information may not be available. All activity upstream of the intakes and downstream of the R.C. Byrd Lock and Dam at

Apple Grove, WV is of interest. Rail traffic has increased considerably in recent years, and past discussions with railroad staff suggest that information about rail cargo will also be limited, although the rail companies do have an inventory of each container's contents. The focus for rail traffic will be coordination of this information during an emergency situation and/or determining the commonly carried hazardous materials on specific rail lines.

Both Ohio and West Virginia have active hydraulic fracturing operations in this region, and accidents and spills have occurred in the recent past. Data for the locations of active hydraulic fracturing operations is desired, though may be of limited use if it is not updated frequently, as the hydraulic fracturing process is relatively short for each well. The locations of injection well sites (for injection of brines used in hydraulic fracturing) are also of interest.

Abandoned mines are to be considered because of the potential for acid mine drainage. This issue was highlighted by flash flooding in an underground Southern Ohio Coal Company mine in Meigs County, Ohio, which released a large quantity of acidic water that caused high levels of iron and aluminum to tributaries of the Ohio River in 1993. Arsenic and selenium often associated with areas affected by acid mine drainage. The WV Office of Abandoned Mine Lands and Reclamation may be a good source of data. Tailings piles are also of concern in West Virginia, particularly as sources of zinc and arsenic, although the scoping group was not aware of an existing source of data for tailing pile locations. They may be visible in aerial images or pH data may be used as a clue to their presence upstream.

The ability to monitor water quality in real-time and detect contaminant events was discussed at length. West Virginia American Water has ordered continuous monitoring equipment in order to establish baseline water quality conditions in the river and is also interested in upstream event detection. The group discussed options for event detection software packages that could be incorporated into the information system. Several commenters requested the ability to transfer or share data between a facility's LIMS system and the information system being developed under this project. ORSANCO also maintains several datasets that may be of use. For example, they have a detailed list of potential contaminants in the Ohio River and an organics detection system that identifies for 30 of the most commonly found compounds. From an operational perspective, information about the location of a contaminant plume is also desirable because a different response may be necessary depending on which side of the river the plume travels down; in the Ohio River, lateral mixing may occur over long distances, though the group was not aware of any upstream mixing studies.

The ability and extent to which utilities and other entities along the Ohio River share data is an important topic for the Huntington site. ORSANCO has water quality and other data that may be useful, including real-time water quality and flow data and logs of historical spills and contaminant data. Data sharing agreements may need to be revisited to better manage emergency situations and communication among responders. There may be opportunities to improve interstate communication in emergencies, in particular, through the information system that is being developed.

Other topics that were discussed but of lesser or undetermined importance were pipeline crossings, coal ash slurry ponds and nuclear power plants.

### Aqua Pennsylvania Pilot Sites

The two pilot sites for Aqua America are the Shenango, PA and Kankakee, IL facilities. The Shenango plant intake is on the Shenango River in western Pennsylvania and the primary water provider for Sharon, PA and its environs. The plant uses the Actiflo treatment process, which means that coagulation and sedimentation occur very quickly and there is a relatively short retention time during which operators can react to changes in influent water quality. The plant has no alternate sources or interconnections and relies on storage during emergency situations. Travel time between the reservoir and plant filter is between 3 and 5 hours normally but can be as short as 2 hours under less typical high flow release conditions. For these reasons, accurate real-time knowledge of water quality is especially critical at this plant.

The Kankakee facility has surface and lower level intakes on the Kankakee River downstream of its confluence with the Iroquois River. Upstream of the Kankakee plant intake the Kankakee River is primarily spring fed and the water has a relatively high hardness for a surface water. The Iroquois River water is much more turbid than the Kankakee. The water treatment plant intake is on the upstream side of the city of Kankakee, IL; therefore, runoff from the city is not of concern. The watershed is predominantly agricultural and plant staff did not report significant problems related to source water quality. During low flows, the source water can come almost entirely from the Kankakee River with very little mixing of Iroquois River water, which is generally more turbid. During higher flows, more mixing occurs, and turbidity can change by several orders of magnitude within a few hours. Under some flow conditions a portion of the raw water bypasses softening (lime). When raw water is turbid, 100% of the water is treated in all processes.

### *Shenango*

The most pressing concerns for this site are rapid detection of changes in water quality, Shenango Reservoir water quality and management, leachate from a capped landfill just upstream of plant intakes, the Sharon Wastewater Authority sewage treatment plant's lift station and associated historical overflow conditions just upstream, an upstream rail yard and road crossings.

Because of the short overall retention time at the Shenango plant, source water monitoring and early warning systems are especially critical. Plant staff currently monitor turbidity, pH, algae (via Algae Watch) and chlorine dioxide (one of several oxidants used at the plant) using continuous analyzers. Other parameters analyzed are taste and odor, TOC, organics and UV254. These parameters are monitored via grab sampling though the plant staff working to appropriately expand using continuous monitoring options for these and other best practices.

The US ACOE operates the Shenango Reservoir and maintains a minimum release of 100 cubic feet per second from its low level outlet. The dam is not designed to spill over the top. There are no visible above ground storage tanks near the reservoir. Aqua maintains no control over the dam and reservoir and the US ACOE typically notifies Aqua of planned flow changes and "dam Ops" from the reservoir. A useful addition to this communication would be continuous (real-time) flow and water quality monitoring at the dam. Currently, water quality in the Shenango Reservoir is unknown separate from downstream tributaries prior to the intakes and none of the information about reservoir operations comes directly to Aqua's control room other than phone calls. Aqua staff has recently considered the possibility of requesting wall space for monitors at the dam. A significant water quality data gap faced by Aqua is algae concentrations and community makeup (i.e., whether algae are toxic) in the reservoir.

There is a closed and capped landfill immediately upstream of the Shenango plant's intake that produces leachate of unknown quality and quantity at this time. The landfill owner is petitioning for permission to release the leachate to the river rather than sending it to a treatment plant, as is currently required. While looking at the site in the prototype system during the scoping webinar, participants noted that it is in the CERCLIS database and was previously listed on the

National Priorities list for the Superfund program. Additional information about the quality of this leachate is needed and may be available in federal CERCLIS documents, industrial pre-treatment agreements or monitoring for the receiving wastewater treatment plant.

Another site of particular concern is the Sharon Wastewater Authority's lift station, which has an overflow that enters the river just upstream of the water treatment plant. As the landfill leachate is currently discharged to the Sharon Wastewater Authority's collection system, this is a potential cause and component of lift station overflows. It may be possible for Aqua staff to request permission to install alarms or real-time TOC monitors and level detectors on or near the lift station that would provide at least 10-15-minute notice of a spill event. Monitoring at this location could also provide a clear picture of source water quality downstream of the reservoir including most major tributaries and the potential landfill leachate.

The Shenango Rail Yard is about 1.5 miles upstream of the intake. Trains run 24 hours a day and Aqua staff has no current way of knowing what cargo is carried through this site or on nearby tracks. There is currently no SOP for emergency communication between the rail yard and Aqua.

The PA DEP will be working on source water protection at the Shenango plant soon, and the new information system should be able to upload this information and share other datasets with the DEP's system. This data sharing is important because some of the contaminant data collected for the pilot sites may be useful for downstream facilities. Within the Shenango facility, regular links could be established with the LIMs system.

### *Kankakee*

The most pressing concerns for this site are water quality impacts from intensive agriculture. Examples of those impacts are atrazine use and elevated ammonia concentrations. In prior source water assessments for the Kankakee plant, data were more readily available for the portions of the zone of critical concern in Illinois than for portions in Indiana. One of the few source water problems that the Kankakee plant has experienced in the past is a high demand for chlorine (possibly related to raw water ammonia loads) as well as taste and odor episodes. In addition, atrazine was recently detected in the source water for the first time in five years. Atrazine and ammonia sources should be investigated.

Plant staff felt that the IL Environmental Protection Agency's (EPA) source water assessments were well done in the past, although the assessments stopped at the Indiana border. Aqua staff requested the inclusion of NPDES permitted facilities in Indiana identified during the last source water assessment; however, no additional information about potential threats in Indiana was collected. IL EPA is currently revisiting the source water assessment at Aqua's request but does not have plans to include any additional data from Indiana at this time. Plant staff noted that two thirds of the watershed is in Indiana, including large areas of intensive agriculture, several interstate highways and a planned extension of I-65.

## Metropolitan Washington Council of Governments

The Metropolitan Washington Council of Governments (COG) was included in the scoping process to broaden input from the drinking water community during the early stages of development of the information system. CEC is working with COG to finalize a contract to do similar assessment using this information system for the Potomac River watershed. Thus, the scoping session for COG focused on the most critical features and uses for the information system that is being built.

The most pressing concerns for COG members were acute threats close to metropolitan Washington, such as pipelines, road/rail crossings and storage tanks. Information about these sources would be used to update source water assessments, prepare for spill response and prioritize standard and emergency monitoring efforts and equipment needs. It could also be used to guide source water protection work.

The group also discussed security concerns with specific data sources, such as Tier II forms, which are used to report the storage locations and volumes of hazardous substances. Some of these substances are regulated under federal Homeland Security legislation, such as the Chemical Facility Anti-Terrorism Standards (CFATS), or can be claimed as trade secrets, which limits the availability of information about the location and storage volume for these hazardous chemicals. The group discussed options for setting up data sharing protocols to allow access to this information during an emergency if it is not otherwise available. The National Dam Safety Review Board (NDSRB)is working to complete an interagency agreement for data sharing during emergency situations, and it may be worthwhile to discuss the issue with the NDSRB.

**New Data Sources**

The data sources below were mentioned during the scoping sessions as critical sources of information relevant to the key concerns at the pilot sites. CEC will contact the data owners to obtain the relevant data if possible and incorporate them into the new information system.

- Stream flow gaging and water quality monitoring stations
- Flood zones
- Time of travel models
- Landfill leachate characterization (for municipal solid waste landfills in general and for the landfill upstream of the Shenango pilot site specifically)
- US ACOE water quality monitoring and/or discharge data from flow control structures (for Stonewall Jackson, WV, and Shenango, PA, reservoirs)
- Railroad hazardous cargo and/or emergency management procedures
- Kanawha Putnam Emergency Planning Committee (KPEPC) hazardous materials transport study
- US ACOE AIS ship tracking data, if readily available
- ORSANCO water quality and historical spill data
- Hydraulic fracturing sites
- Hydraulic fracturing brine injection wells
- Abandoned mines
- Coal tailings piles
- Greenville Water Authority water quality data (for the Little Shenango River upstream of the Shenango pilot site)
- Agricultural data
    - Atrazine use
    - The Association of American Plant Food Control Officials (AAPFCO) data (fertilizer and trace metals)

During the COG scoping session, several data sources were discussed for the Potomac River watershed, which is not included in the scope of the current project, but will be part of a future project for CEC and COG:

- 2002 District of Columbia Source Water Assessment report
- Chesapeake Bay Program – many relevant datasets
- Interstate Commission on the Potomac River Basin (ICPRB) – time of travel model, river miles
- Water quality monitoring data (primarily from biomonitoring at locations along the Potomac river)
- Water Contaminant Information Tool (WCIT) – This EPA database houses information about "contaminants of concern for water security" but is not accessible to the private sector.

## SUMMARY AND CONCLUSIONS

The four scoping meetings identified the types of analyses the project partners consider most useful for the GIS system (in general) and the issues of greatest concern at each of the four pilot sites. Some of the discussions extended beyond the scope of the current project but will undoubtedly be revisited in future projects. The scope of work that can be delivered under the current effort is summarized below.

### Information System

The goal of the information system is to provide access to regularly updated datasets related to acute threats to source water vulnerability in a cloud-based, open-source GIS system. The system will need to perform basic GIS functions, such as creating new datasets, reporting and summarizing, as well as provide a platform for tracking changes in data over time. The user experience should be geared toward plant operators, source water protection and emergency response staff.

Near-term information system developments will result in a functional model that will be used during requirements analyses. Features that will be available for evaluation and consideration at the requirements analysis include integration of a contaminants database linked to potential sources of contamination, links to real-time USGS stream gage data, and ability to select and display key data that may be used to identify or investigate potential sources of contamination.

### Pilot Projects

The same basic analysis will be conducted for all four pilot plants along with a tailored investigation of particular sites or issues of concern for each plant. For all of the pilot sites, CEC will collect basic federal and state data related to acute threats and conduct an analysis of aerial images within the zones of critical concern to identify above ground storage tanks and other visible potential threats to source water quality. The data will be loaded into the information system and tools will be developed to complete the analysis within the open source software. The methodologies for data collection and analysis will be documented so that other utilities can complete the same level of analysis without using the information system developed in this project.

The objectives for each pilot site were discussed during the scoping meetings. Particular sites and issues of concern (outside the general scope described above) will be investigated by

CEC and incorporated into the final report and information system if possible. CEC will attempt to find locate, quality assure and load the following data into the information system:

***Weston***

***Stonewall Jackson dam***

- US ACOE reservoir management and water quality data
- Groundwater wells' water quality data from the nearby golf course
- Marina activities and any water quality data that may be available

***Oil and gas wells***

- Current and historic locations

***Interstate 79***

- Potential hazardous materials transported

***Huntington***

***Industrial Facilities***

Chemical storage and use at in the zone of critical concern

***Trains***

- Hazardous cargo
- Emergency contacts
- Emergency response SOPs

***Ships & Barges***

- AIS tracking (locations and hazardous cargo flags), if readily available

***Oil and gas wells***

- Active hydraulic fracturing locations

***Injection wells***

- Location and content

***Mines***

- Active and abandoned mine locations
- Tailings piles locations
-

***Water quality monitoring data***

- ORSANCO
- RAIN
- West Virginia American Water raw water monitoring data, if made available

***Shenango***

***Shenango Reservoir***

- US ACOE reservoir management and water quality data

***Landfill leachate composition or constituents of concern***

***Rail yard***

- Hazardous cargo
- Emergency contacts
- Emergency SOPs

***Road crossing locations***

***Sharon wastewater authority sewage lift outfall location and historic discharges***

- Frequencies and dates of spills
- Confirm landfill leachate routing and possible information the wastewater utility possesses on the discharge

***Kankakee***

***Agricultural runoff***

- Atrazine loadings
- Ammonia loadings

***Potential sources of contamination in Indiana***

***Road crossings***

***IL EPA sites of concern***

# APPENDIX D
# REQUIREMENTS REPORT

REQUIREMENTS ANALYSIS

Water Research Foundation Project 4581: Using Geographic Information System (GIS) to Identify and Catalog Potential Contaminant Stored in Tanks Upstream of Drinking Water Intakes

Prepared by:
Jeffrey Rosen, Margaret Kearns & Tim Bartrand
Corona Environmental Consulting
6 Old Country Way
Scituate, MA 02066

March 16, 2015

## REQUIREMENTS ANALYSIS

### Project Description

The Water Research Foundation's Project Number 04581, "Using Geographic Information Systems (GIS) to Identify and Catalog Potential Contaminants Stored in Tanks Upstream of Drinking Water Intakes," will deliver a methodology for identifying and characterizing chemical storage facilities and other potential source water contamination sites and a functional software tool that provides access to these data for the emergency response community and the drinking water community. As part of this effort, a data gaps analysis will identify those datasets that are difficult to access or otherwise unavailable. The project also includes application of the methodology and toolset to four pilot sites in West Virginia, Pennsylvania and Illinois.

Project 4581 will be conducted in 4 steps: (i) scoping; (ii) preliminary toolset development; (iii) requirements analysis; (iv) pilot studies. Generally, the requirements analysis, the topic of this report, entails determining specific needs for the information system, such as forms, reports, security, administration and more detailed functionality.

### Purpose and Report Organization

This report summarizes discussions at two requirements analysis workshops that were held with project partners and advisory committee members in February 2015. The report presents the requirements workshop methodology, the process for prioritizing requirements, and summarizes the discussions (results) about specific system features. The Summary provides an outline of the high priority and other necessary requirements for the information system that will be developed based on feedback received at these workshops. Although the requirements discussed in this report are specific to a geospatial information platform in development by CEC, they illustrate practical needs expressed by drinking water professionals and suggest general methodologies for meeting those needs, whether using the information platform or through other means.

### Methods

Prior to the requirements workshops, CEC developed a suite of pilot information system tools to demonstrate the potential for the final system to meet the anticipated needs of the water community based on needs articulated during past projects and discussions. Once the pilot tools were in place, two requirements analysis workshops were held to elicit input from project team members. Each of the two day workshops was organized to support group brainstorming, refine the ideas generated and prioritize the resulting list (Appendix A). The workshops were broken up into thematic sessions covering a range of topics necessary to fully describe the desired information system's requirements:

- Business requirements (what should the tool do?)
- Viewing, creating and analyzing GIS data
- Communications tools
- Reporting
- Training & support
- System administration
- Security

Each session began with an introduction to the theme and demonstrations of pilot system features related to the session's theme. A facilitated brainstorming session followed the demonstrations, and ideas were documented as they were discussed and refined by the group. The evening after the first day of the workshop, CEC staff reviewed the list and identified requirements that were core to the scope of the information system, as defined during previous scoping sessions, and those that were not already anticipated for the first version of the system. On the second day, workshop participants reviewed the list of not currently planned requirements and discussed general priority levels for them, categorizing them from highest priority essential functions to "desirements", i.e. features that would be useful to the drinking water community but not absolutely essential in the first version of the information system. Following the end of the second workshop, CEC combined and summarized the lists from both workshops.

## RESULTS

### Workshops

Two requirements analysis workshops were held in January and February 2015, one in Charleston, WV and one in Washington, D.C. The workshops were attended by representatives from all of the project partners and the two advisory committees (Appendix B). The pilot system demonstration features effectively stimulated discussion on each session's themes and included tools to complete the following tasks:

- Visualize points, polygons & images
- View oblique imagery
- View related data in popup windows
- Create bookmarks & galleries
- Visualize changes over time
- View three-dimensional data
- Create new data
- Edit data
- Track changes
- Search functions
- Export spreadsheets
- Produce online reports with text, maps & charts
- Export single & multiple chemical reports
- Generate charts and tables of potential sources of contamination
- Communicate with clients or other users
- Integrate social media
- Manage users

### Discussion of the Requirements

Workshop participants suggested requirements that can be grouped into several major themes:

- Methods and tools for improved spill response

- Methods and tools for generating automated reports and facilitating routine communications
- Information/data sharing with entities outside the utility
- Easy access to routinely-used data
- Tracking changes and updates in watershed features.

The requirements suggested in these categories point to underlying needs of utility staff that are currently only partially met or not met at all. Requirements related to these categories are described below and methodologies for meeting these requirements within the geospatial information platform or via other tools and approaches are discussed.

***Methods and Tools for Improved Spill Response***

Participants expressed consistent and specific interest in tools and data management to support spill response. Requirements related to spill response do not pertain directly to pilot studies that are being conducted as a part of WRF project 4581, although they are a logical extension of the process of identifying and characterizing potential sources of contamination (PSCs). Broadly, the spill response requirements identified during the workshops can be grouped as engineering requirements (where was the spill, how long until the spill reaches the intake) and communications requirements.

Engineering/science requirements related to spill response cover analyses on spills from the time of the spill to the time the contaminant has passed a drinking water treatment plant intake. Participants requested that spill locations, temporary sample locations and other points of interest related to a spill could be added to the information system just as PSCs are currently added. Adding locations of interest on the fly implies a number of additional underlying requirements and suggests that, at present, tools for maintaining geospatial spill data and relating those data to other key data are lacking. An important aspect of maintaining spill locational data is establishing a consistent frame of reference for spill, sample and other data. River miles appear to be the best reference system for referring to spill locations, though definitions of river miles can differ even for plants with intakes on the same source water.

Time of travel (TOT) modeling requirements differ widely among utilities. Some utilities have access to highly-resolved and calibrated time of travel models and prefer incorporating those models into the information system, or conducting time of travel analyses outside the information system. Other utilities, particularly those whose intakes are on lower order streams, have either no TOT models or rudimentary or out of date models. For those utilities a requirement to develop basic TOT estimation methods and tools would be a significant benefit. Even a relatively simple TOT model will require numerous underlying system features such as ability to delineate the upstream drainage from any point, ability to associate a particular stream location with appropriate gauge data and overland path models that connect spill locations with water courses.

Other engineering and science spill response requirements included collection and maintenance of data related to laboratory capacity, ability to retrieve actionable contaminant data quickly, and ability to capture/record information/data related to event (e.g., gauge data, sample results). Once again, this requirement implies that the ability to store and manage spill-related data might be lacking.

Spill response requires many flows of information from utilities. Utilities must communicate with the party responsible for the spill (for example, an above ground tank owner), entities in the emergency response plan, regulators, and perhaps most importantly customers. In general, spill response communication requirements were related to rapid access to reliable data that could be incorporated into spill messaging and communications vehicles already in place at utilities. For example, one requirement is the ability to make a static map that is pushed out to a mailing list. Data within the information system varies in its ability to be disseminated. Some data will be open (e.g., data originating from an open federal data source) while other data might be proprietary or sensitive and their distribution might be limited. When data are output for use in communications, open and restricted data must be clearly distinguished.

### *Methods and Tools for Reporting and Routine Communications*

Requirements pertaining to routine communications included inverse communication to tank owners (utility outreach to tank owners) and automatic generation of reports that summarize the zone of critical concern (or other region) through automatic generation of data that provide an overview of upstream risk (as a component of ongoing source water characterization). Inverse communication with tank owners should establish an open path between tank owners and utilities such that information is updated frequently enough to be relevant during a spill. Outreach to tank owners could be periodic or based on water quality observations or other triggers. In the event of a suspected spill these communications could be extended to emergency responders.

Another communications requirement was that the data, methods and tools in development in this project could also be used to generate updates to source water assessment reports. Information system features that would facilitate automatic generation of reports include the ability to group by source type, ownership, or contaminants. The requirement includes automatic generation of pie charts and tabular data typical of those used in source water assessments.

Another requirement, albeit a relatively low priority, was the ability to enlist external agencies and individuals in data collection and development of the information system. Engaging external system users poses additional underlying requirements related to access, security and quality control and assurance. Examples of external communication requirements (from the external party, through the information system and to the utility) are crowdsourcing as a source of PSC data and access to the information system by staff of the National Rural Water Association (circuit riders).

### *Management of the system within the utility and with outside entities (general community, emergency responders)*

Numerous system access requirements were posed. Those requirements allow access control during normal operations and additional options for use during events, table top exercises and other instances when access is temporarily changed. As currently envisioned, the information system will have an administrator at each utility or group of utilities. The administrator will define user types and assign access to members of each user type group.

In certain circumstances such as event response or table top exercises temporary access and other changes might be required. For example, requirements related to table top exercises are that

- Data used and generated during table top exercises be flagged and segregated from other data

- An administrator should have the ability to assign temporary access to the information system
- Precautions should be in place that maintain data securely and prevent export of the data to the extent possible.

Another requirement suggests that the system be able to "turn on" periodic archives of the data during an event. To be of greatest use archived data should be created in folders with meaningful names and in a structure that is intuitive.

### *Easy Access to Routinely Used External Data*

Water utilities use external data routinely and workshop participants proposed several requirements related to easier access to those data and integration of the data with an information system. Specific mention was made of connections to USGS gauge data and NOAA hydrographs. Additional data that could be integrated into the information system are water quality data (collected online or via grab samples and collected by any of the stakeholders), Tier II data and groups of contacts. One requirement generalized this grouping of requirements by requiring that the information system be able to incorporate external data (of any kind) supplied by the users. This requirement will be considered, but is very difficult to implement due to wide variability in the format and nature of data users might wish to incorporate.

### *Monitoring Changes in the Watershed*

A key feature of the methodology proposed in WRF project 4581 is the continual update of data. As a practical matter, system users, plant operators and others will be very concerned with changes in their watershed. This need was captured in a requirement that the information system have the capacity to generate reports of changes and updates for features within a given region. This requirement must be addressed thoughtfully to avoid confronting users with a very long daily list of changes that might not be significant.

## SUMMARY AND NEXT STEPS

The two facilitated workshops effectively documented the needs and desires of a variety of drinking water utilities and project partners for identifying potential sources of contamination upstream of drinking water intakes and preparing for and responding to emergency incidents. The most important topics identified were methods and tools for improved spill response, reporting and communication, information sharing with outside entities, accessing routinely-used data and tracking changes and updates in watershed features. The development of a system that can address these features will improve utilities' ability to prioritize and track source water protection activities as well as plan for and respond to emergency situations.

The next steps for this project are to develop the information system and complete the four pilot studies. With the addition of Greater Cincinnati Water Works as a fifth pilot study, CEC will have the opportunity to compare the new methodology and system developed for this project to current and past source water protection efforts to document its strengths and weaknesses.

## RAW LIST OF REQUIREMENTS

| Meeting | Requirement | Priority |
|---|---|---|
| WV | Enter sub-meter GPS coordinates of an exact spill location (or other location) and create a map feature that has properties associated with it. | 1 |
| WV | Spill response functionality. This should be an operational system that can be turned on and then accumulates information as it comes in. Allows visualization and information sharing with responders. | |
| WV | Take and save screenshots. Allow sharing with a team. E.g., share as a text message. Contents could be saved as a bookmark. Along with creating a bookmark, generate a static picture of a view that can then be shared. Alternatively, the system publishes a point-in-time copy of the view that the user bookmarks. This creates an interactive view with static content but functionality. | |
| WV | "Turn on" periodic archives of the data during an event. Make sure there is an archive of all updates to the data base (on an ongoing basis, not just during events). Archives created in meaningful folders. | |
| WV | Flexibility in how points are created. This could be river miles (could be per US ACE definitions), lat/long, street intersections, based on mouse clicks, etc. | 1 |
| WV | Ability to create a new category of point and define the attributes for the new point category. There could be a point type "other" for new point creation. Alternatively, add "spill site" as an added point type. Attributes are time, material, plants between the spill site and your intake, etc. | 1 |
| WV | Inverse communication to tank owners (utility outreach to tank owners). Establish an open communication path. For example, create a mail merge list to allow periodic communication or automatically generate the communications. Also, reach out to PSC owners based on water quality observations. This should extend to first-responders to remind them to inform the utility in the event of a spill. The critical point is the ability to export mail merge list or similar list of contacts to whom a communication can be sent. | 1 |
| WV | Incorporate hydraulic atlas data. | Note |
| WV | User can load a data set managed somewhere else. | 1 |
| WV | Facilitate data gaps analysis. Develop an understanding of what's out there. | 1 |
| WV | Flag conflicting information regarding a point. | 1 |
| WV | Send alert notification for any automatic update that impacts a field-verified point. | 1 |
| WV | If the information system is used in a tabletop exercise, there should be a tabletop mode that prevents accumulation of pretend data but allows the system to record the tabletop exercise. | |

| Meeting | Requirement | Priority |
|---|---|---|
| WV | Tools to help prioritize follow-up to fill data gaps. Prioritization could be by contaminant properties, locations, geographic regions, volumes, etc. | |
| WV | Tools to identify data gaps by category. | |
| WV | When there are inconsistencies in data records, the information system automatically flags the data and the data are validated by the facilities responsible. | |
| WV | Establish parent-child relationship between tanks and a site. | 1 |
| WV | Government entities can access federal pipeline data. Private entities cannot get access. Pointing private utilities to the data source is a step in the right direction | Note |
| WV | Ability to make a static map that is pushed out to a mailing list. Similar functionality to SwiftReach. | |
| WV | Selectively notify customers (e.g., limit notifications to a portion of the distribution system or distinct geographic regions such as pressure zones). | |
| WV | Communication of water quality events to downstream utilities. | |
| WV | Notify groups (external to the utility or within the utility) when they are granted access to the system during an emergency response. | |
| WV | Label data available for external use as confidential, sensitive, restricted, etc. For example, when data are exported they are accompanied by a disclaimer or notification saying that the data are confidential or otherwise restricted. Could have an option that all reports and exported data are marked as confidential unless the user specifies otherwise. | 1 |
| WV | Ingestion and attribution (including access restrictions) of external data. | 1 |
| WV | Automatically generate reports of data in a user-specified region, ZCC or whole watershed. Have a dropdown list of choices of data that can be reported/exported. A WV Senate Bill 373 requirement is to report health ratings, location (not available for public disclosure for tanks in the ZCC), and whether the tank is fit for service. | 1 |
| WV | User-defined fields should be associated with created points to enable tracking, follow-up, prioritization, etc. | 1 |
| WV | Ability to generate a facility report (i.e., everything at a particular site or facility). | 1 |
| WV | Click on a point on a stream and identify upstream or downstream PSCs or facilities. | 1 |
| WV | Water utility identifies downstream intakes that might be impacted by their chemical tanks in the event of a spill and generates a report for those intakes indicating PSCs. Generate notifications to downstream intakes in the vent of a spill. | 1 |
| WV | Ability to create pre-defined map with legend, north arrow, confidentiality disclosure, scale, etc. Predefined map is for a given region and for a user-selected subset of data. | 1 |
| WV | User can create a template that allows mapping (with all the usual map features) of defined features in a selected region. | 1 |

| Meeting | Requirement | Priority |
|---|---|---|
| WV | Adjust the hue and saturation for maps (e.g., satellite image base maps) in the user interface. | 1 |
| WV | ORSANCO has a contaminants "bible" that has very good treatment data and good one-sheet summaries provided by system. Try to gain access and use data if possible. | |
| WV | In a multiple contaminants report, add distance to intake, whether the PSC is from a facility with a permit (e.g., TRI) or was created by the user. | 2 |
| WV | Create a standard report using user-defined parameters. Have the ability for the user to define their own parameters. | 1 |
| WV | Develop and document a methodology for prioritizing contaminant data collection (a deliverable under the WRF project). Include in the report additional benefits for capturing contaminant data. | 1 |
| WV | Searches of contaminant database by: chemical; level (hazard); any parameter listed on the state tank registry; tank construction/material; tank status [e.g., not fit for service]. | |
| WV | Control access in the field. Field staff can view, add and annotate the data. | |
| WV | Likely system users: the GIS team; intermittent access for managers; water quality team; and treatment team. | Note |
| WV | Capture pipelines. Important attributes include size and pressure, contents, and locations of shut-off valves. Storm sewers can be important as conveyances to MS4 outfalls but will not be mapped in this effort. An alternative to positively identifying all pipeline attributes is to establish a communication link between the utility and pipeline owner. | 3 |
| WV | When there are significant changes to ownership, activities, land uses etc., those are flagged. Flagged changes are pushed to end users as messages. | 3 |
| WV | Integrate disparate flows and characteristics for tributaries in system and estimate the time of travel from any point to the intake. | 1 |
| WV | Click on a point on any tributary and calculate the distance to the intake. | 1 |
| WV | Determine the distance between two points on the stream network. | 1 |
| WV | Four major areas - (i) knowing what's there (ii) emergency response (iii) forensics and documentation for after the fact and (iv) table top exercises. | 1 |
| WV | The information system will maintain contacts by individual system, contacts by agency and contacts relevant to location (by geopolitical boundaries). User can query the contacts by geography, mouse click, selecting a region, etc. | 1 |
| WV | Show floodplains. Annotate whether sites have flood emergency response plans or flood protection in place. | 1 |
| WV | Source water protection planning - dynamic and live system. Have a library of contaminants in the event of a spill. | 1 |
| WV | Identify MS4 outfalls, SSO overflows and CSO overflows. | 1 |

| Meeting | Requirement | Priority |
|---|---|---|
| WV | Provide access to USGS gauge data (raw). Hyperlink directly to USGS gauge sites. Direct access to NOAA predictive hydrographs (via hyperlink). Click on gauge point on the map to get basic data and access to hyperlink to NOAA or USGS live feed. | 1 |
| WV | Easy access/query of relevant data for use in event communication. In this scenario, the information system produces the relevant data and another set of tools is used to generate the communications (e.g., emails). | 1 |
| WV | Push data to agencies the zone of critical concern (or other coverage) for incorporation into their own emergency response tools. | 1 |
| WV | Ingest Tier II data into the system. | 1 |
| WV | Sort data that are selected and reported. For example, sort by location within the ZCC or based on the health rating. | 1 |
| WV | Export PSCs or other data as a spreadsheet. | 1 |
| WV | Automated generation of reports that summarize the zone of critical concern (or other region). Summarize data that provide an overview of upstream risk. Refer to examples used during the workshop (drawn from the Neshaminy SWAP). Grouping by source type. Grouping by ownership. Grouping by contaminant type or contaminant identity. Generate tables and pie charts. Allow user to select what h/she wants using radio buttons or similar controls. | 1 |
| WV | Views that may be saved and used for exporting user-defined maps, such as bookmarks and/or galleries. | 1 |
| WV | Reports that show the data source for each contaminant datum. | 1 |
| WV | Report of tank status for tanks in the ZCC. | 1 |
| WV | Searchable comment fields. | 1 |
| WV | Hands-on training via in-person session. Training should use live data. One in-person training session per organization. | 1 |
| WV | Before training sessions, confirm that everyone getting trained has access on their device. | 1 |
| WV | Aqua training in Bryn Mawr (suggested). Kankakee staff could attend or attend via webinar. | 1 |
| WV | 24/7 support is not necessary. Project team will be available for support in the first year. | 1 |
| WV | Test all systems prior to any training (WVAW, Aqua, GCWW by webinar). Test at training locations (Charleston + Bryn Mawr). | 1 |
| WV | Let general public provide point locations of PSCs. For utility staff, identify points that should be updated or verified based on observations in the field. Staff can add new points based on crowdsourced observations or field observations. Data access control on crowdsource tools is critical. | |
| WV | For the public sourcing of data, acknowledge public inputs and provide explanation regarding the choice to make the data available publicly or not. | |
| WV | In source water planning, the information system can identify stakeholders and facilitate communication. | 3 |

| Meeting | Requirement | Priority |
|---|---|---|
| WV | Inventory of physical assets and other emergency response assets. Emergency availability of resources: water, samplers, labs, etc. This might duplicate processes already in place with EMA. | 3 |
| WV | Maps or animation showing contaminant progressing down the river. | 3 |
| WV | Integrate ICWater. | 3 |
| WV | Surrogate measures for chemical presence/absence. For example, would the contaminant also show up as a change in TOC or conductivity? | 3 |
| WV | Visualization of algal blooms and risk of toxins, MIB, *geosmin* and associated monitoring data. Similar tools and visualization (time evolution) as in the Ebola demo. Visualization could be done for other evolving threats or concerns. | 3 |
| WV | Review UKWIR contaminant database and incorporate into data collection if possible. | 3 |
| WV | Tools that allows communication with the general public. Crowdsourcing for inputting potential sources of contamination. General public has access to a portion of the tools and data. Good example would be access to river watch groups. | 3 |
| WV | Give external agencies access to the system to inspect and/or download select sets of data. | 3 |
| WV | Snipping tool for capturing images of partial screen shots. | 3 |
| WV | Inventory of who is sampling (upstream), where good sample locations might be, and what resources are needed and available for sampling. Tool can facilitate network of water users on a stream. | 3 |
| WV | Functional roles: View only; full functionality except for establishing roles and rights; and administrator with full functionality plus right to establish roles and rights. Need a non-disclosure agreement level of access. Only people with signed NDAs can see the data. | 1 |
| WV | An attribute of users is whether or not they've signed an NDA. If they have that attribute, they get access to data (though the access still differs according to role). | 1 |
| WV | Limit the ability to print to a pdf and assign limited distribution to extracted data. Grey out the "save as" and similar buttons except for the administrator role. | 2 |
| WV | Need defined user roles such as (i) administrator, (ii) manager, (iii) operator, (iv) outside cooperator, (v) public, and (vi) emergency [multiple levels]. | 1 |
| WV | Confidential data are tagged as confidential and their use is controlled. | 1 |
| WV | Administrator can see a log of changes in the system. | 1 |
| WV | Ability to grant access to data sets by user. Access differs according to user role, whether user has appropriate NDAs in place and the nature of the data source. | 1 |
| WV | When users are first added to the system all permissions are "off." They are turned on by an administrator. | 1 |

| Meeting | Requirement | Priority |
|---|---|---|
| WV | Administrator can assign access to data layers by group. Match the security level of data sets with the security level of user roles/groups. | 1 |
| WV | External access users should have passwords/access with an expiration date. | 2 |
| WV | Administrator can create a profile for each user and each data set. The profiles determine access to data and functionality. Profiles can include expiration dates. | 1 |
| WV | NDAs are associated with individual data sets. Administrator sets whether or not an NDA is required for a data set. Data sets have an attribute related to their use during an emergency. | 1 |
| WV | A user signs a single NDA that covers all data covered by NDAs associated with individual data. This is the preferred alternative. | 1 |
| WV | There are certain data sets that should be locked and require clear intent to allow changes (i.e., data available for edit only to certain users). | |
| WV | The administrator within each utility will have full rights to grant access to users and to access to specific data. Administrator can also use the system. | |
| WV | The administrator within any system has the right to create other administrators within a limited number. | |
| WV | Assign a specific right for loading local GIS data. | |
| WV | Create a user-defined user role (i.e., a role that is created by an administrator). | |
| WV | Only the system administrator or someone designated by the administrator (e.g., GIS technician) can add data to the system (users can propose adding layers, but can't actually do it). Administrator finalizes, approves and makes the dataset available within the system. | |
| DC | Maintain a pedigree for the data (i.e., changes are tracked and users are alerted when changes occur). | 1 |
| DC | Integrate downstream notification data when updating and maintaining storage tank data in the system. | 2 |
| DC | Time of travel model from spill to intake. Including timing that allows informed intake closing decisions. DC region utilities will use ICPRB spill model. | 3 |
| DC | Ability to determine lat/long of a spill site and enter those data to become an event. | 1 |
| DC | Rapidly retrieve all information about a contaminant (e.g., treatability, physicochemical characteristics, toxicity, analytical procedures, etc.). | 1 |
| DC | Maintain information on lab capacity for analysis of specific contaminants; link to EPA lab database; document regional capacity for analysis; look into Water Laboratory Alliance (EPA) as a potential data source | 1 |
| DC | Identify and visualize established points for sampling (e.g., intakes, bridges). | |

| Meeting | Requirement | Priority |
|---|---|---|
| DC | Identify and visualize sampling points (create these on the fly) and associate points with water quality data entered into the system. | 1 |
| DC | Create spill location points and associate points with source of information. | 1 |
| DC | Create generic potential sites of contamination (e.g., a waste pile). These could be point, line or polygon. Associate the feature with the data source. Associate feature with QA data for the point. | |
| DC | Connect real-time water quality data to fixed sampling points, including an interface. | 3 |
| DC | Communicate system information with other responders (e.g., EPA, Coast Guard, etc.) Link to or integrate with existing contingency and emergency response plans. | 2 |
| DC | Organize points of contact and contact information for emergency response plans, contingency plans, and tactical response plans. Better to use positions instead of names. | 3 |
| DC | Integrate (link to) emergency response plans, contingency plans, tactical response plans and standard FEMA ICS forms. The key is to maintain meta-data related to these forms to help users quickly clarify overlapping jurisdiction and to identify members for various groups/associations. The DC Region Homeland Security currently plays this role. | |
| DC | Ability to create groups of contacts. | |
| DC | Gaps analysis. What data are missing and important? Which agency should pursue filling gaps? The system should have a tool for identifying the gaps but will not answer which agency should pursue filling the gaps. System should have the capacity to compare gaps with respect to their importance. For example, large volume tanks might be higher priority than smaller tanks in determining their contents. Gaps could be determined and compared for contacts, tanks, contaminant data, etc. The most important gaps to fill are tank contents. | 1 |
| DC | Develop and integrate methodologies for comparing data gaps and prioritizing data collection. | |
| DC | Assemble the key data components that would be used in a risk estimate for a specific intake-contaminant combination. This includes location (i.e., proximity to intake). | 1 |
| DC | Conduct general risk categorizations based on gross parameters such as proximity, toxicity, volumes. | 1 |
| DC | User input for weighting prioritization and level of risk (i.e., user-input data for contaminants). | 2 |
| DC | Show tweets and other social media in the system. Could supplement existing utility communication capabilities and tracking. This is an information gathering tool that is integrated with emergency response. | 2 |
| DC | Plug in all regional messages and public notification work/tools. | |
| DC | Link to portions of the CDC toolbox. | |
| DC | Associate photos/media and notes with a point created in the system. | 1 |

| Meeting | Requirement | Priority |
|---|---|---|
| DC | Ability to extract lat/long from photos or other data used to create a point. | |
| DC | Search by NAIC or SIC code (or other industrial code), contaminant, proximity (e.g., in river miles) from any point, river reach (as defined by user), contaminant characteristic (e.g., odor), facility name, compliance status (e.g., under a consent order), watershed, geopolitical jurisdiction, and risk ranking. | |
| DC | Capture/record information/data related to an event (e.g., gauge data, sample results, etc.). | |
| DC | Use an event to validate the flow model. | |
| DC | Search by updates done over the last X hours. | |
| DC | Incorporate MARFC flood prediction data. | |
| DC | For the National Capitol Regional, communication and coordination tools are already in existence. System can provide components of reports that will be used in extant communication tools/plans/channels. | Note |
| DC | Push out maps showing spill locations and spill progress. | 1 |
| DC | System will produce content (important information related to a spill) and push it out for use in existing communication tools. For example, reports of contaminant information generated as pdfs, plain text or in another format can be attached to emails or incorporated into documents. | 1 |
| DC | System will create bookmarks (views with certain geography, specific layers turned on and with annotations). | 1 |
| DC | Ability to query system for list of changes since last check. Alternatively, the system will push out alerts at a particular frequency or changes could be shown on a dashboard. | 1 |
| DC | Data change dashboard that summarizes changes since a certain time. User chooses whether or not to look at them. Changes are categorized (e.g., by the field that was changed). Differentiate between data connected to PSCs and those that are not as well as those that are in the ZCC or other designated buffer. | 1 |
| DC | Notifications on addition of or changes to contaminant data. | 1 |
| DC | Create a map that can be published in another GIS system. | 1 |
| DC | Templates for maps or a bookmark created for public distribution (e.g., agree upon which layers should be exported). | 2 |
| DC | Links for sharing views or bookmarks with others. | 1 |
| DC | Automatically generate components of a source water assessment or protection plan. Account for requirements that differ by state. | 3 |
| DC | Report of contact information for tank owners, parcel owners and emergency responders. This could be done for a specified region (e.g., within the ZCC or a lassoed region). | 1 |
| DC | Report of contaminants that is editable (e.g., Word doc to be edited by the end-user). | 1 |

| Meeting | Requirement | Priority |
|---|---|---|
| DC | Export data and/or tables as Word documents or Excel spreadsheets. | 1 |
| DC | Ability to import data (e.g., data from a utility's GIS). This function could require CEC facilitation. | 2 |
| DC | Ability to create and save an ad hoc report (e.g., all facilities storing petroleum products). | 1 |
| DC | Push data corrections to the originating agency. | 1 |
| DC | Hands-on training. | Note |
| DC | Create training modules (train the trainers) and recordings that can be used to demonstrate for public officials and other stakeholders (e.g., to show at the partnership for the utilities not currently participating). | |
| DC | Face-to-face meetings to train trainers. | Note |
| DC | Create and record webinars that show how to use the system. | |
| DC | Create webinar snippets that show how to use specific system elements/tools/functions. | |
| DC | Support provided 5 days per week is probably sufficient. The objective is to answer questions within a day. | Note |
| DC | Create an FAQ and make it accessible. | |
| DC | Create and maintain a user chat room. | |
| DC | Make sure you can restore access 24/7 (e.g., automatically regenerate forgotten passwords). | |
| DC | Typically train one or two GIS people and one or two technical people per utility (probably 4 people for COG). | Note |
| DC | For WSSC, train a minimum of two people. | Note |
| DC | Aqua training will likely be in Bryn Mawr. | Note |
| DC | For Fairfax, ideally hold a hands-on training for 4-5 people plus a webinar involving more people and a presentation (or recorded presentation) for management describing the purpose and functionality of the system. | Note |
| DC | Ability to assign roles (e.g., master administrator, administrator, operator, engineer, reviewer, etc.). Three levels are probably sufficient (i.e., administrator, individuals who enter and edit local information and operator). Give a few users a wide ability to disseminate data from the system. User levels are assigned conservative default permissions. Utility administrators control data access of "master administrator". | 1 |
| DC | Administrators can create new roles. | |
| DC | Ability to assign permissions (read, write, edit, delete). | 1 |
| DC | Ability to tag data with respect to the degree to which they may be disseminated. | 1 |
| DC | Designation of data as coming from public sources (and generally available for use) (i.e., does the public have access to the data outside the system?). | 1 |
| DC | Designation of "master user" who helps administer the system for a group of utilities. | |

| Meeting | Requirement | Priority |
|---|---|---|
| DC | Default access for data is that nobody can view it. Viewing permissions must be turned on. | 1 |
| DC | For source water planning and protection, only a select few people focused on source water protection are included as users in the system. Public access (or other wider access) does not contribute to the mission of source water protection. | Note |
| DC | As a rule, general public should not be given access to the tool. However, there could be an emergency mode that allows access to the emergency response community for very specific sets of information. Emergency response partners are unlikely to use the information system but could benefit from reports dumped from the information system. | Note |
| DC | National Rural Water Association circuit riders may be sources of PSC site data and somehow users of the system. | 3 |
| DC | Situational awareness dashboard for decision-makers to use during an event. CEC contacts group already working on situational awareness dashboard creators (see Steve B's email). | 3 |
| DC | System administrators may know other system administrators and can make data available to users outside their utility. | 1 |
| DC | An emergency mode that allows all utilities to see everyone's selected data. This is controlled by any administrator from any of the member utilities. All users are notified that the emergency mode has been activated. Activation of emergency mode is logged (who initiated it, when, etc.). | 2 |
| DC | Ability to create expiration dates for emergency mode or access. | 3 |
| DC | Tag data and actions as occurring during an event. Maybe just tag a range of dates and times as event period. | 3 |
| DC | Tag data generated during table top exercises and demonstrations as not real. | 3 |
| DC | Primary goal is to get a live version of the source water assessment (current state of practice is to keep static assessments). | Note |
| DC | Train users to create and use tags for data (e.g., tag data as occurring during an event, demonstration or tabletop exercise). | Note |
| DC | Record data owner for each data set. Any changes need to go through the owner. | |

# APPENDIX E
# CONTAMINANT DATA COLLECTION STANDARD OPERATING PROCEDURE

**Standard Operating Procedure**

**Contaminant Database, WaterSuite**

6 Old Country Way

Scituate, MA 02066

781.378.0271

www.coronaenv.com

## Table of Contents

# ACRONYMS

ATSDR- Agency for Toxic Substances and Disease Registry

CAS – Chemical Abstracts Services
CASRN – CAS Registry Number
CDR – Chemical Data Reporting Rule
CEC – Corona Environmental Consulting

DOT – Department of Transportation
DTB – Drinking water Treatability Database

EPA – United States Environmental Protection Agency

GIS – Geographic Information System

IRIS - Integrated Risk Information System

LD50 – Lethal Dose 50

MSDS – Material Safety Data Sheet

NAICS – North American Industry Classification System

PSC – Potential Source of Contamination

ORD – Office of Research and Development

QA/QC – Quality Control/Quality Assurance

RfD – Reference Dose

SOP - Standard Operating Procedure

TOXNET - TOXicology Data NETwork

ZCC – Zone of Critical Concern

# LIST OF KEY TERMS

Data Type

Data Type refers to the type of format required for a specific data point in the database. Data type examples include: integer (whole numbers), text, date and time, decimal etc.

Lethal Dose 50

The amount of a chemical required to kill 50% of a population of animals in a given time

Unique Key

Uniquely identifies each row in a table and is comprised of a set of columns. No two distinct rows in a table can have the same combination of column values. Required for record uniqueness.

## INTRODUCTION AND BACKGROUND

This Standard Operating Procedure (SOP) describes the methods Corona Environmental Consulting (CEC) will use to manage and present data to support work on the contaminant database which is a part of a larger system referred to as the WaterSuite.

WaterSuite is an online geospatial platform and information system that links federal, state and local data indicating potential sources of contamination (PSC) to a contaminant database developed specifically for the drinking water community. Each potential source of contamination may have a different contaminant or set of contaminants of concern. For each contaminant identified within a zone of critical concern (ZCC) for a client, chemical, physical, toxicological, medical and safety information will be collected and included in the WaterSuite contaminants database, following the standardized methodology described in this document.

Ultimately, the WaterSuite system will provide clients with a suite of standard Geographic Information System (GIS) tools and functions as well as federal, state and local data related to source water protection. The GIS viewer will link known locations of a particular contaminant with information about its physical properties, fate and transport, toxicity, health effects and emergency response needs. Queries and reports on these data sets will be able to be generated including spreadsheets, maps and GIS files for use in reports or other communications. As new tools are developed within WaterSuite, (i.e. time of travel and prioritization routines) contaminant data collection efforts may need to be modified. Any modifications will be documented in an updated version of this SOP.

## PURPOSE AND OBJECTIVE

The purpose of the SOP is to provide guidance for CEC staff and universities supporting data collection efforts for the contaminant database and to ensure that all data are entered and Quality Assurance/Quality Control (QA/QCed) in an appropriate manner. This approach will help ensure that the required data are collected and are of the appropriate quality. It will also standardize and facilitate data collection.

## ROLES AND RESPONSIBILITIES

CEC will oversee the data collection process supported by the various universities that will have students entering the data. Once CEC obtains a set of contaminants the first step will include identifying a Chemical Abstracts Service Registry Number (CASRN) if it is not listed in the source data. CAS Common Chemistry can be used to confirm a CASRN or to confirm a chemical name. The user can search for substances in Common Chemistry by either the CASRN or by the chemical name.

After a CASRN is confirmed, alternative names will be checked prior to sending a contaminant request to the various universities. If a contaminant is already in the database under an alternative name CEC will document this and that contaminant will not be sent to the universities for data entry. An alternative names list was developed primarily from one main data source, CAS Common Chemistry (http://www.commonchemistry.org). If a contaminant is not listed in CAS Common Chemistry, supplemental sources will be consulted including CAMEO, etc.

Next CEC will evaluate if any of the contaminants are already in the database by cross-referencing the alternative names list in the database. Once alternative names are evaluated CEC will assign datasets to the students at the various universities who will be responsible for data collection and entry. Students will enter data under the direction of point person at each university.

In addition, each university will also have a responsibility to QA/QC a portion of the data entered by a different university. This process is detailed in Data Mining Procedures. QA/QC.

## DATA SOURCES

Currently, CEC utilizes several main databases to populate the majority of the data parameters including:

### CAS 'Common Chemistry'

*Chemical Abstracts Service, a division of the American Chemical Society* http://www.commonchemistry.org

Common Chemistry is provided by Chemical Abstracts Service, as a public service. CAS is a division of the American Chemical Society. The Common Chemistry database contains the CAS Registry Number, chemical names, molecular formulas, and structures for approximately 7900 chemicals. These substances of widespread general public interest are included if they are of global commercial use or importance and have been cited 1,000 or more times in the CAS databases. In addition, all of the elements of the Periodic Table are also in the Common Chemistry database, although not all of the elements may meet the 1,000 references threshold.

### TOXNET (TOXicology data NETwork) Database

*Toxicology and Environmental Health Information Program, Division of Specialized Information Services of the National Library of Medicine, TOXNET (TOXicology Data NETwork)* http://toxnet.nlm.nih.gov

TOXNET is comprised of a set of databases covering chemicals and drugs, risk assessment, regulations, toxicology, diseases and the environment, environmental health, occupational safety and health. In addition, TOXNET also provides links to PubMed, The National Library of Medicine (NLM's) which allows access to biomedical literature. TOXNET includes specific chemicals, mixtures, and products, as well as unknown chemicals. TOXNET also lists special toxic effects of chemicals in humans and/or animals and includes references from the scientific literature.

The databases that are available through TOXNET include: ChemIDplus, Chemical Carcinogenesis Research Information System, Carcinogenic Potency Database, Comparative Toxicogenomics Database, Haz-Map, Genetic Toxicology, Integrated Risk Information System (IRIS), International Toxicity Estimates for Risk, Hazardous Substances Data Bank (HSDB), The Household Products Database, LactMed, Toxics Release Inventory, TOXMAP, and TOXLINE. In general, HSDB, IRIS, and ChemIDPlus are the more utilized data sources for the data mining process. Each data source is described in more detail in app X.

### CAMEO Chemicals

*National Oceanic and Atmospheric Administration, Office of Response and Restoration. n.d. CAMEO Chemicals, version 2.4.1 rev.3.* http://cameochemicals.noaa.gov

CAMEO Chemicals is a library with thousands of datasheets containing response-related information and recommendations for hazardous materials that are frequently transported, used, or stored in the United States. This database has been designed as a reactivity prediction tool (demonstrating the tendency of substances to undergo chemical change) and to evaluate reactivity alerts about individual chemicals for people involved in hazardous material incident response and planning. In order to predict reactivity, each chemical has been assigned to one or more reactive groups based on its chemistry. The user is able to create a “My Chemicals” section where reactivity predictions are based on the reactions that might occur between reactive groups. This tool is designed to predict pair-wise chemical reaction(s), and may be useful to predict chemical reactions between drinking water and a given chemical. For data mining purposes the data analyst will collect general information including UN numbers, DOT labels, reactivity, health effects, hazard ratings etc., CEC will not be utilizing the My Chemicals section of CAMEO at this time.

### Integrated Risk Information System

*United States Environmental Protection Agency(EPA), Office of Research and Development (ORD), National Center for Environmental Assessment.* http://www.epa.gov/iris/

IRIS is a human health assessment program with a web accessible database maintained by the EPA National Center for Environmental Assessment, a division of ORD. The IRIS database contains information on more than 550 chemical substances including information on health effects that may result from exposure to environmental contaminants. IRIS provides health effects hazard identification and dose-response assessment information.

### Agency for Toxic Substances and Disease Registry (ATSDR)

United States Department of Health and Human Services, Agency for Toxic Substances and Disease Registry http://www.atsdr.cdc.gov

The Agency for Toxic Substances and Disease Registry (ATSDR) is a part of the U.S. Department of Health and Human Services. ATSDR is a federal public health agency that conducts public health assessments and studies in an effort to provide information to help prevent harmful exposures and diseases related to toxic substances. The ATSDR contains information on more than 550 chemicals substances including information on health effects that may result from exposure to environmental contaminants. ATSDR has several toxicological profiles for contaminants, ToxFAQs, and a toxic substances portal.

### EPA Drinking Water Treatability Database

*United States Environmental Protection Agency, Drinking Water Treatability Database* http://iaspub.epa.gov/tdb/pages/general/home.do

The EPA's Drinking Water Treatability Database (TDB) can be used to identify effective drinking water treatment processes for a contaminant of interest. It provides an overview of treatment processes that were found to be effective for the removal of a given contaminant. When available the database also describes treatment processes that are considered to be ineffective for the removal of a contaminant from drinking water. Data are based on peer-reviewed journals and conferences, symposia, research reports, theses, and dissertations. Currently there are over 50 contaminants in the database including those regulated in drinking water, on the CCL, of water security interest, of pesticide registration interest, and endocrine disruptors and pharmaceuticals. The database includes more than 25 treatment processes used by drinking water utilities.

### Material Safety Data Sheets

A Material Safety Data Sheets (MSDS) is a document designed for workers in the event that they may be exposed to hazardous materials. MSDS contain information about the nature of a chemical, such as physical and chemical properties and provide information on the potential hazards (health, fire, reactivity and environmental) and how to work safely with the chemical product. MSDS are particularly useful when populating the database for a mixture.

### Additional Data Sources

In an effort to populate all data fields for a contaminant, additional data sources will be necessary to fill in data gaps. Some examples of additional credible data sources may include:

- EPA Computational Toxicology Research
- US Department of Health and Human Services
- NIOSH Pocket Guide to Chemical Hazards
- Right to Know Hazardous Substance Fact Sheets (various states i.e. NJ) http://web.doh.state.nj.us/rtkhsfs/indexfs.aspx
- National Primary Drinking Water Regulations – Technical Fact Sheets http://www.epa.gov/ogwdw/pdfs/factsheets/voc/tech/toluene.pdf
- Merck Index
- UKWIR toxicity database
- National Environmental Methods Index
- Chemspider
- NIST Chemistry WEbBook
- OSHA Occupational Chemical Database
- EPA ChemFact
- HAZ MAT tool
- EPA NCOD http://www.epa.gov/enviro/html/data_source.html
- http://water.epa.gov/drink/index.cfm -

### Scientific Literature

In an effort to populate all data fields for a given contaminant it may be necessary to perform literature reviews. This will be necessary to fill in any data gaps. First a data gaps analysis will be conducted by CEC and additional costs and effort involved to fill in gaps will be reviewed. CEC will work with clients and evaluate the drinking water community needs to prioritize which gaps get filled.

### Data Mining Procedure

During data mining, CEC will collect a variety of information to populate the contaminant database. Several pieces of information will be collected from various data sources. Each parameter of interest for a given contaminant is listed below.

| **General Information** |
|---|
| Name, CASRN, UN/NA, DOT Label, Alternative Names, Manufacturer, Purpose, Regulatory Programs |
| **Physical Properties** |
| Color, Odor, Taste |
| **Fate and Transport** |
| Specific Gravity, Solubility, Vapor Pressure, Henry's Law Constant, Partition Coefficient, Decay Constant, Reactivity, |
| **Detection and Treatability** |
| Method Detection Limit, Treatability |
| **Hazard Ratings** |
| National Fire Protection Agency |
| **Health Effects and Medical Treatment** |
| Health Effects, Medical Treatment, Protective Equipment |
| **Toxicity** |
| LD50, NOAEL, RfD |

The workflow process for capturing contaminant data in the Contaminant Database consists of the following steps:

- Verify CAS or blend, alternative names, UN/NA numbers, and DOT hazard labels
- Collect Physical and Chemical data and a portion of the fate and transport data
- Medical, Health effects and Protective equipment
- Evaluate Toxicity data
- Hazard Info
- Detection and Treatability

#### *Verify CAS or mixture, alternative names, UN/NA numbers, and DOT*

The first steps of documenting a new contaminant include verifying the CAS number if it is not already identified. CEC will use CAS Common Chemistry to confirm the CASRN and chemical name. After a CAS is confirmed, alternative names will be checked prior to sending a contaminant request to the various universities. If a contaminant is already in the database under an alternative name CEC will document this and that contaminant will not be sent to the universities for data entry. An alternative names list was developed primarily from one main data source, CAS Common Chemistry (http://www.commonchemistry.org). If a contaminant is not listed in CAS Common Chemistry, supplemental sources will be consulted including CAMEO, etc.

After a CAS number is verified the next step is to collect UN/NA numbers and any applicable DOT labels. If a contaminant is a mixture there will not be an associated CASRN. Mixtures will have two features. First CEC will identify if something is a mix and assign the mix ID (i.e. MIX-#). This will be the unique identifier for the particular mixture.

### *Classifying Mixtures/Identifying Components of a Mixture*

Mixtures will be categorized as 1) a defined mix/well characterized (all components defined), 2) semi-defined mix/partially defined (include a majority component), or 3) undefined. Mixtures will also be classified by the major chemical components each listed with the CASRN.

A mixture will have its own chemical and physical properties. It will also have health effects, medical treatment, and protective equipment and NFPA ratings. After a mixture has been identified, individual components of that mixture should be evaluated as components may hit intake at different times so there is a need to add the missing contaminant from that mixture. Documentation will occur to identify if the individual components already exist in the contaminant database. Each mixture will be classified as fully evaluated, or not fully evaluated by the data analyst.

Sometimes the components will be listed as proprietary, which need to still be documented as a component. If the percentage of the mixture that is proprietary is listed it should be documented in the database as the percentage of mixture by weight when possible.

UN/NA numbers and DOT hazard classifications should also be listed for a contaminant. CAMEO and TOXNET typically have UN/NA and DOT designations for each contaminant.

### *Collect Physical and Chemical data*

For physical and chemical data, the user will be directed to look up the contaminant in TOXNET first, followed by CAMEO Chemicals. If information about a contaminant exists at both TOXNET and CAMEO, the data should be collected from both sites.

### *Medical, Health Effects, Protective Equipment, and Hazard Information*

The same data mining process should be followed to collect medical, health effects, protective equipment, and hazard data.

### *Evaluate Toxicity data*

Three main pieces of toxicity data are collected for each contaminant, LD50, NOAEL, and RfD. TOXNET (ChemIdPlus) is the first source to check followed by IRIS, and ATSDR Toxic Substance Portal and Toxicological Profiles.

### *General Data Entry Consideration*

The data analyst may encounter a few issues during the data mining process.

- *Scientific notation.* If a data point is listed in scientific notation the data analyst should enter the data as presented in the original source. The data wizard will account for any necessary data conversions.
- *No Min/max listed for a parameter*. If a single value is encountered in the source and the data entry form asks for minimum and a maximum value, the data wizard will account for this.
- *Source vs reference*. A data source will be TOXNET, CAMEO, a MSDS for a given contaminant and a reference is the original source such as an author and date of the study. In the event an original source is not listed the URL of the data source will be used.
- *Data from two sources conflict*. List each data point and their data analyst will be prompted with a flag from the data wizard indicating there is a data discrepancy. The analyst will be able to provide a comment and a CEC staff member will review the data conflict.
- *Multiple data points for a single field.* If multiple data points exist for a parameter, the data analyst will enter each data point in separate data entries.
- *Data may be unavailable in the standard sources, causing data gaps*. A review will be conducted by each data analyst as a part of the data mining process for incomplete fields/data gaps. To aid this effort, users will be prompted with a form-based mechanism (check box) to indicate if the primary data sources were reviewed for an empty data parameter. A second tier of data mining may occur for incomplete data fields. CEC will determine when the second tier of data collection efforts will occur. This effort may include literature reviews.

### *Data Collection*

Each unique piece of information about a contaminant is referred to as a *parameter* in this SOP. Each parameter is detailed in individual tables below. A description of each parameter, sources to check and a detailed description of all data fields are included in each table.

<u>CASRN Number</u>

The CASRN number is a unique numerical identifier assigned to each chemical by the Chemical Abstracts Service (CAS).

Sources to check: CAS Common Chemistry, CAMEO, TOXNET

| **Field** | **Description** | **Data Type** | **Example** |
|---|---|---|---|
| CASRN | CAS Reference Number | Number | 1634-04-4 |

| | | | |
|---|---|---|---|
| Source | Database, data source | Text | CAMEO |
| Reference | Citation | Text or hyperlink | http://cameochemicals.noaa.gov/chemical/330 |

DOT Hazard Label

The Department of Transportation labels certain materials based on their hazardous properties.

| **Field** | **Description** | **Data Type** | **Example** |
|---|---|---|---|
| DOT Code | Hazard Label | Text | Flammable Liquid |
| Source | Database, data source | Text | CAMEO |
| Reference | Citation | Text or hyperlink | http://cameochemicals.noaa.gov/chemical/330 |

UN/NA Number

UN numbers, assigned by the United Nations Committee of Experts on the Transport of Dangerous Goods, are universally accepted four-digit numbers used to identify hazardous materials or classes of hazardous materials. They are identical to North American (NA) numbers.

| **Field** | **Description** | **Data Type** | **Example** |
|---|---|---|---|
| UN/NA Number | UN/NA Number | Number | 1013 |
| Notes | Qualification of UN/NA number | Text | Anhydrous |
| Source | Database, data source | Text | CAMEO |
| Reference | Citation | Text or hyperlink | http://cameochemicals.noaa.gov/chemical/330 |

Alternative Names

This parameter contains a list of common synonyms and trade names for each contaminant. This table allows the user to identify if a contaminant already exists in the database.

Sources to check: CAS Common Chemistry, CAMEO, NIOSH Pocket Guide

| **Field** | **Description** | **Data Type** | **Example** |
|---|---|---|---|
| Alternative Names | Synonyms for contaminant | Text | Carbon Dioxide, [Refrigerated Liquid] |
| Source | Database, data source | Text | CAMEO |
| Reference | Citation | Text or hyperlink | http://cameochemicals.noaa.gov/chemical/330 |

Purpose

The Chemical Data Reporting (CDR) Rule requires manufacturers and importers to report information on the contaminants they produce domestically or import into the United States. The CDR database currently contains the 2012 submission, which provided 2011 manufacturing, processing, and use data for reportable contaminants. This table contains information from the US EPA Non-confidential 2012 Chemical Data Reporting information on chemical production and use in the United Sates. The next submission period for the 2016 CDR will not be collected until June to September 2016.

Sources to check: TOXNET and MSDS

| Field | Description | Data Type | Example |
|---|---|---|---|
| Purpose | Use | Text | Used for heating oil and diesel engines |
| Source | Database, MSDS, etc. | Text | TOXNET (HAZ-MAP) |
| Reference | Citation | Text or hyperlink | CHEMINFO MSDS |

Manufacturer

Manufacturer information contained in TOXNET is primarily obtained from the US EPA Chemical Data Reporting (CDR) report 'Non-confidential 2012 Chemical Data Reporting information on chemical production and use in the United Sates.' There may be multiple production sites listed, general manufacturer information, etc. For our purposes only manufacturer information will be collected at this time. Production site information will not be mined for the CEC database Each piece of information is accompanied by an address for the site.

Sources to check: TOXNET for all fields with the exception of NAICS. CEC will input NAICS codes.

| Field | Description | Data Type | Example |
|---|---|---|---|
| Manufacturer | Manufacturer | Text | Eastman Chemical Co, Inc. |
| Street Address line 1 | Street Address line 1 | Text | 200 South Wilcox Dr. |
| Line 2 | Line 2 | Text | |
| City | City | Text | Kingsport |
| State | State | Text | TN |
| Zip Code (5) | Zip Code (5) | Number | 37660 |
| Zip Code (4) | 4 number suffix in zip code | Number | 5147 |
| Country | Name of country | Text | USA |
| Phone | Phone Number | Number | (781)929-9934 |

| | | | |
|---|---|---|---|
| NAICS Code | NAICS Code | Number | 325211 |
| NAICS Description | NAICS Description | Text | Plastics Material and Resin Manufacturing |
| Source | Database, MSDS, etc. | Text | TOXNET (HSDB) |
| Reference | Citation | Text or hyperlink | [US EPA; Chemical Data Reporting (CDR)…. * |

Regulations

This parameter is populated with lists from regulatory programs. The user will not be prompted to enter any data in the regulatory table. Currently the regulations table Emergency Planning and Community Right to Know Act (EPCRA)/ Comprehensive Environmental Response Compensation and Liability Act (CERCLA)/Clean Air Act (CAA, section 112(r)) Consolidated List of Lists is in the regulations table. Other regulations to be included are as follows: Clean Water Act List of Priority Pollutants, Toxics Release Inventory (TRI) Toxic Chemical List, Persistent Bioaccumulative Toxic (PBT) Chemicals, National Waste Minimization Program Priority Chemicals, Chemical Facility Anti-Terrorism Standards (CFATS), Emergency Response Notification System, SARA Priority List of Hazardous Substances, Toxic Organics (TTO) POTW industrial pre-treatment standard, UCMR3 candidate monitoring list, possibly dw specific reg. Regulation information is necessary to have in the case of an emergency.

Sources to check: N/A

| **Field** | **Description** | **Data Type** | **Example** |
|---|---|---|---|
| Section 302 (EHS) TPQ | Extremely Hazardous Substances, Threshold Planning Quantity | Number | 1,000/10,000 |
| Section 304 EHS RQ | Extremely Hazardous Substances – Reportable Quantity | Number | 5,000 |
| CERCLA RQ | Comprehensive Environmental Response, Compensation & Liability Act of 1980 – Reportable Quantity | Number | 5,000 |
| Section 313 | Section 313 - Toxics Release Inventory Reporting | Number | 313 |

| | | | |
|---|---|---|---|
| RCRACODE | Resource Conservation and Recovery Act | Text | U007 |
| CAA 112(r) TQ | Section 112(r) of the Clean Air Act | Number | N/A |

Color

At times, water can have an unpleasant appearance. This may not pose a public health threat but a sudden change in color could indicate a potential public health concern.

Sources to check: TOXNET, CAMEO

| **Field** | **Description** | **Data Type** | **Example** |
|---|---|---|---|
| Color | Color of contaminant in a given form | Text | Colorless to light pink crystalline solid |
| Notes | Additional Pertinent Information | Text | Phenol liquefies by mixing with about 8% water. |
| Source | Database, MSDS, etc. | text | TOXNET (HSDB) |
| Reference | Citation | Text or hyperlink | NIOSH Pocket Guide to Chemical Hazards. |

Odor

Odor threshold measurement may be considered as having at least two parameters (quality and intensity). The qualitative aspect may be described in associative terms i.e., "like-dislike” or familiar qualitative experiences like fatty, sour, sweet, etc. (cite ref). Odor may also be assessed by other categories including discernable, ‘irritating’ concentration, odor threshold, and odor detection limit. Odor can be an early detection that a there is a contaminant release.

Sources to check: TOXNET, CAMEO

| **Field** | **Description** | **Data Type** | **Example** |
|---|---|---|---|
| Odor | Odor of contaminant | Text | Odorless |
| Numeric Qualifier | Numeric Qualifier | Text | > |
| Odor Threshold Minimum | Odor Threshold Minimum | Number | 1 |
| Numeric Qualifier | Numeric Qualifier | Text | > |
| Odor Threshold Maximum | Odor Threshold Maximum | Number | 1 |

| | | | |
|---|---|---|---|
| Odor Threshold Units | Odor Threshold Units | Text | mg/cu m |
| Odor Threshold Type | Odor Threshold Type | Text | threshold |
| Notes | Additional Pertinent Information | Text | |
| Source | Database, MSDS, etc. | Text | TOXNET (HSBD) |
| Reference | Citation | Text or hyperlink | [O'Neil, M.J. (ed.). The Merck Index - …. |

Taste

A change in a water supply's taste is not necessarily a health concern. However, a change in taste indicates a change in water quality that should be investigated.

Sources to check: TOXNET, CAMEO

| **Field** | **Description** | **Data Type** | **Example** |
|---|---|---|---|
| Taste (Qualitative) | Description of taste of a contaminant | Text | "nasty," bitter, nauseating and similar to rubbing alcohol |
| Numeric Qualifier | Numeric Qualifier | Text | > |
| Taste Minimum | Taste Minimum | Number | 24 |
| Numeric Qualifier | Numeric Qualifier | Text | > |
| Taste Maximum | Taste Maximum | Number | 37 |
| Taste Units | Taste Units | Text | µg/litre |
| Notes | Notes | Text | N/A |
| Source | Database, MSDS, etc. | Text | MTBE in Drinking-water - Background Document for development of WHO. |
| Reference | Citation | Text or hyperlink | Dale et al. |

Specific gravity

Specific Gravity is the ratio of mass of a solid to mass of an equal volume of water at 4 degrees C or otherwise noted. It is a measure of density relative to the density of a reference substance. If a substance has a specific gravity greater than 1 it will sink, less than 1 it will float. This is an important piece of information to have in an emergency situation.

Sources to check: TOXNET, CAMEO,

| Field | Description | Data Type | Example |
|---|---|---|---|
| Minimum Numeric Qualifier | Minimum Numeric Qualifier | Text | N/A |
| Specific Gravity Minimum | Specific Gravity Minimum | Number | 1.06 |
| Maximum Numeric Qualifier | Maximum Numeric Qualifier | Text | N/A |
| Specific Gravity Maximum | Specific Gravity Maximum | Number | 1.06 |
| Reference Temperature of Material | Reference Temperature 1 | Number | 20 |
| Reference Temperature of Material Units | Reference Temperature 1 Units | Text | Deg C |
| Reference Temperature of Water | Reference Temperature 2 | Number | 20 |
| Reference Temperature of Water Units | Reference Temperature 2 Units | Text | Deg C |
| Measurement Type | Measurement Type | Text | N/A |
| Notes | Notes | Text | N/A |
| Source | Database, MSDS, etc. | Text | DOW MSDS 2013 |
| Reference | Citation | Text or hyperlink | ASTM D4052 |

<u>Solubility</u>

Solubility is the quality or state of being soluble. It is the maximum quantity of a substance that may be dissolved in another.

Sources to check: TOXNET, CAMEO

| Field | Description | Data Type | Example |
|---|---|---|---|
| Solubility (Qualitative) | Solubility (Qualitative) | Text | Slightly soluble in water |
| Numeric Qualifier | Numeric Qualifier | Text | N/A |
| Minimum Solubility | Minimum Solubility | Number | 2.024X10+3 |
| Numeric Qualifier | Numeric Qualifier | Text | > |
| Maximum Solubility | Maximum Solubility | Number | 2.024X10+3 |
| Solubility Units | Solubility Units | Text | Mg/L |

| | | | |
|---|---|---|---|
| Reference Temperature | Reference Temperature | Number | 25 |
| Reference Temperature Units | Reference Temperature Units | Text | Deg C |
| Measurement Type | Measurement Type | Text | (est) |
| Notes | Notes | Text | In water |
| Source | Database, MSDS, etc. | Text | TOXNET (HSDB) |
| Reference | Citation | Text or hyperlink | US EPA; Estimation Program Interface (EPI) Suite. |

Vapor Pressure

Vapor pressure of a liquid is the pressure exerted by its vapor when the liquid and vapor are in dynamic equilibrium. It is necessary to specify the temperature because vapor pressure increases with temperature. It is an indication of a liquid's evaporation rate.

Sources to check: TOXNET, CAMEO

| **Field** | **Description** | **Data Type** | **Example** |
|---|---|---|---|
| Numeric Qualifier | Numeric Qualifier | Text | N/A |
| Minimum Vapor Pressure | Minimum Vapor Pressure | Number | 0.092 |
| Numeric Qualifier | Numeric Qualifier | Text | > |
| Maximum Vapor Pressure | Maximum Vapor Pressure | Number | 0.092 |
| Vapor Pressure Units | Vapor Pressure Units | Text | mm Hg |
| Reference Temperature | Reference Temperature | Number | 25 |
| Reference Temperature Units | Reference Temperature Units | Text | Deg C |
| Measurement Type | Measurement Type | Text | Extrapolated |
| Notes | Notes | Text | N/A |
| Source | Database, MSDS, etc. | Text | TOXNET |
| Reference | Citation | Text or hyperlink | Daubert, T.E., R.P. Danner… |

Henry's Law Constant

Henry's Law is one of the gas laws and states "At a constant temperature, the amount of a given gas that dissolves in a given type and volume of liquid is directly proportional to the partial pressure of that gas in equilibrium with that liquid." HLC will be useful to utilities in the removal of contaminant in the event of a spill.

Sources to check: TOXNET (not always listed and usually at the bottom of the chemical/physical properties under "other")

| **Field** | **Description** | **Data Type** | **Example** |
|---|---|---|---|
| Numeric Qualifier | Numeric Qualifier | Text | N/A |
| Minimum HLC | Minimum Vapor Pressure | Number | 0.5 |
| Numeric Qualifier | Numeric Qualifier | Text | > |
| Maximum HLC | Maximum Vapor Pressure | Number | 0.5 |
| HLC Units | Vapor Pressure Units | Text | atm-cu m/mol |
| Reference Temperature | Reference Temperature | Number | 25 |
| Reference Temperature Units | Reference Temperature Units | Text | Deg C |
| Measurement Type | Measurement Type | Text | Estimated |
| Notes | Notes | Text | N/A |
| Source | Database, MSDS, etc. | Text | TOXNET |
| Reference | Citation | Text or hyperlink | US EPA; Estimation Program Interface (EPI) Suite |

Octanol/Water Partition coefficient ($K_{ow}$)

According to the EPA, the octanol/water partition coefficient (log $K_{ow}$) is a coefficient representing the ratio of the solubility of a compound in octanol (a non-polar solvent) to its solubility in water (a polar solvent). Higher $K_{OW}$ values indicate more non-polar compounds.

Sources to check: TOXNET, CAMEO, varied sources.

| **Field** | **Description** | **Data Type** | **Example** |
|---|---|---|---|
| Log Kow Minimum | Log Kow Minimum | Number | 0.47 |
| Log Kow Maximum | Log Kow Maximum | Number | 0.47 |
| Measurement Type | Measured or Estimated | Text | (est) |
| Source | Database, MSDS, etc. | Text | TOXNET (HSDB) |
| Reference | Citation | Text or hyperlink | US EPA; Estimation Program |

| | | | Interface (EPI) Suite. |
|---|---|---|---|

Reactivity

Reactivity can be defined as the rate at which a chemical substance tends to undergo a chemical reaction. Qualitative descriptions of reactivity are available in CAMEO

Sources to check: CAMEO then TOXNET

| **Field** | **Description** | **Data Type** | **Example** |
|---|---|---|---|
| Reactant Description | Reactant Description | Text | 2,2-DIBROMO-3-NITRILOPROPIONAMIDE is incompatible with bases, metals, oxidizing agents, acids. Dangerous gases may accumulate as a result of ignition and fire. |
| Source | Database, MSDS, etc. | Text | CAMEO |
| Reference | Citation | Text | http://cameochemicals.noaa.gov/chemical/18093 |

Method detection limit (MDL)

The EPA defines a MDL as the minimum concentration of a substance that can be measured and reported with 99-percent confidence that the analyte concentration is greater than zero. It is determined from the analysis of a sample in a given matrix containing the analyte.

Sources to check: EPA

| **Field** | **Description** | **Data Type** | **Example** |
|---|---|---|---|
| Detection Method | Description of Detection Method | Text | US EPA 502.2 |
| Detection Limit | Detection Limit | Number | 0.01 |
| Detection Limit Units | Detection Limit Units | Text | µg/L |
| Source | Database, MSDS, etc. | Text | EPA |
| Reference | Citation | Text or hyperlink | |

Treatability

This parameter includes referenced information on the control of contaminants in drinking water.

Sources to check: EPA Drinking Water Treatability Database, TOXNET (HSDB) cleanup methods, varied.

| Field | Description | Data Type | Example |
|---|---|---|---|
| Treatment Description | Treatment types found to be effective | Text | The following treatment method(s) have proven to be effective for removing antimony to below 0.006 mg/L or 6 ppb: coagulation/filtration and reverse osmosis. |
| Source | Database, MSDS, etc. | Text | EPA - Basic Information about Antimony in Drinking Water |
| Reference | Citation | Text or hyperlink | http://water.epa.gov/drink/contaminants/basicinformation/antimony.cfm |

National Fire Protection Association (NFPA 704) Hazard Ratings

NFPA 704 is the standard system for the identification of the Hazards of Materials for Emergency Response.

Sources to check: CAMEO

| Field | Description | Data Type | Example |
|---|---|---|---|
| Health Hazard | Danger, warning, caution categories | Number | 1 |
| Flammability | Flammability categories | Number | 2 |
| Instability Rating | Notes | Number | 0 |
| Special Hazard Rating | Water reactive, oxidizing agent, etc. | Text | N/A |
| Notes | Additional Relevant Information | Text | N/A |
| Source | Database, MSDS, etc. | Text | CAMEO |
| Reference | Citation | Text or hyperlink | NFPA, 2010 |

Health Effects

This parameter describes the associated health effects from a particular contaminant. It is relevant for an emergency situation.

Sources to check: TOXNET and CAMEO

| Field | Description | Data Type | Example |
|---|---|---|---|
| Health Effects | Description of various health effects from a contaminant | Text | Strong skin irritant, moderate eye irritant |

| | | | |
|---|---|---|---|
| Source | Database, MSDS, etc. | Text | TOXNET (HSDB) |
| Reference | Citation | Text or hyperlink | http://toxnet.nlm.nih.gov/ |

Protective Equipment

This parameter describes the necessary protective equipment if handling a contaminant spill. It is relevant for an emergency situation.

Sources to check: TOXNET and CAMEO

| Field | Description | Data Type | Example |
|---|---|---|---|
| Protective Equipment Description | Description of protective equipment | Text | Protective gloves; goggles or face shield. |
| Source | Database, MSDS, etc. | Text | CAMEO |
| Reference | Citation | Text or hyperlink | USCG, 1999 |

Medical Treatment

This parameter describes the associated necessary medical treatment after exposure to a particular contaminant. It is relevant for an emergency situation.

Sources to check: CAMEO and TOXNET

| Field | Description | Data Type | Example |
|---|---|---|---|
| Medical Treatment Description | Description of necessary medical treatment if exposed to a contaminant | Text | INGESTION: do not induce vomiting. |
| Source | Database, MSDS, etc. | Text | CAMEO |
| Reference | Citation | Text or hyperlink | USCG, 1999 |

LD50

Lethal Dose50: The amount of a chemical required to kill 50% of a population of animals in a given time. Multiple entries can exist for a certain animal as well as multi animal studies for a given contaminant. Each data point should be entered as an independent record.

Sources to check: Start with entire search of TOXNET but most likely to be in TOXNET (HSDB) or TOXNET (ChemIDPlus), then IRIS, and ATSDR.

| Field | Description | Data Type | Example |
|---|---|---|---|

| | | | |
|---|---|---|---|
| Numeric Qualifier | Numeric Qualifier | Text | > |
| Minimum LD50 | An LD50 represents the individual dose required to kill 50 percent of a population of test animals | Number | 2000 |
| Maximum LD50 | An LD50 represents the individual dose required to kill 50 percent of a population of test animals | Number | 2000 |
| Exposure Method | Route of contaminant exposure | Text | Oral, dermal, ingestion |
| Animal Model | Species | Text | Rat |
| Animal Model Sex | Type of animal | Text | Male, female |
| Notes | Other defining characteristics | Text | N/A |
| Source | Database, MSDS, etc. | Text | N/A |
| Reference | Citation | Text or hyperlink | DOW MSDS 2013 |

<u>NOAEL</u>

In toxicology the no observed adverse effect level (NOAEL) indicates the exposure level of an organism, through experiment or observation, at which there is no biologically or statistically significant increase in the frequency or severity of adverse effects in the exposed population when compared to its appropriate control. (ref: United States Environmental Protection Agency). Examples of significant increase in adverse effects would include alteration of morphology, functional capacity, growth, development or life span.

Sources to check: TOXNET (ChemIdPlus), IRIS, ATSDR

| **Field** | **Description** | **Data Type** | **Example** |
|---|---|---|---|
| NOAEL | No observed adverse effect level | Number | 180 |
| NOAEL Units | Units for NOAEL | Text | mg/kg-day |
| Exposure | Level of exposure at which no adverse effects are found | | |
| Effects | Effects | Text | Reduced body weight and clinical signs |

| | | | |
|---|---|---|---|
| Exposure Amounts | Exposure Amounts per body weight per day | Text | 0, 60, 180, or 540 mg/kg-day |
| Animal Model | Animal Model | Text | Rabbits |
| Animal Model Condition | Animal Model Condition | Text | pregnant |
| Animal Model Sex | Animal Model Sex | Text | Female |
| Notes | Notes | Text | |
| Source | Database, MSDS, etc. | Text | TOXNET |
| Reference | Citation | Text or hyperlink | SIDS. Screening Information Data Set for High Production Volume Chemicals… |

Reference Dose

Definition: Reference dose (RfD) is an estimate (with uncertainty limitations) of a daily oral exposure to the human population (including sensitive subgroups) that is likely to be without an appreciable risk of harmful effects during a lifetime. RfDs may be derived from a NOAEL, LOAEL, or benchmark dose. Uncertainty factors are generally applied to reflect limitations of the data used.

Sources to check: IRIS, TOXNET (IRIS), ATSDR

| Field | Description | Data Type | Example |
|---|---|---|---|
| CASRN | Unique numerical identifier assigned by CAS | Number | 71-43-2 |
| Reference Dose | An estimate of a daily oral exposure to the human population that is likely to be without an appreciable risk of harmful effects during a lifetime | Number | 4.0 x 10-3 |

| | | | |
|---|---|---|---|
| Reference Dose Units | Units for RfD | Text | mg/kg-day |
| Experimental Doses | A statistical lower confidence limit on the dose or concentration at the Benchmark Dose (BMD) | Text | BMDL = 1.2 mg/kg/day |
| Critical Effects | The most sensitive and relevant endpoint | Text | Decreased lymphocyte count (Human occupational inhalation study) |
| Uncertainty Factor | Usually reflects the limitations of the data used | Number | 300 |
| Modifying Factor | It provides additional uncertainty not included in the UF | Number | 1 |
| URL | URL of source | Text | http://www.epa.gov/iris/subst/0238.htm |
| Source | Database, MSDS, etc. | Text | ASTDR |
| Reference | Citation | Text or hyperlink | CDC |

Mix Identification

This table will identify if a contaminant is a mixture.

| **Field** | **Description** | **Data Type** | **Example** |
|---|---|---|---|
| Chemical Mixture ID # | Name of mixture | Text | Non Chlorinated Brakleen |
| CASRN (MIX-#) | ID assigned to each mixture, unique identifier for blend | Text | MIX-1 |
| Source | Database, MSDS, etc. | Text | CRC Safety Datasheet |
| Reference | Citation | Text or hyperlink | Brakleen® Non-Chlorinated Brake Parts Cleaner MSDS |

Chemical Mixture

Each component of a mixture will have an associated CAS number and percentage of contaminant by weight relative to each mixture. For example, Acetone, CASRN 67-64-1, 80% minimum and 90& maximum relative to the mixture.

Sources: N/A

| **Field** | **Description** | **Data Type** | **Example** |
|---|---|---|---|
| Chemical Mixture-Chemical Crosswalk ID # | Contaminant | Text | Acetone |
| Chemical Mixture ID # | Unique ID assigned to each Mixture | Text | MIX-1 |
| CASRN | Unique record for each component | Number | 67-64-1 |
| Numerical Qualifier | | Text | N/A |
| % of Mixture by Volume Minimum | % of component in blend | Percentage | N/A |
| Numeric Qualifier | Numeric Qualifier | Text | > |
| % of Mixture by Volume Maximum | % of component in blend | Percentage | N/A |
| Numeric Qualifier | Numeric Qualifier | Text | > |
| % of Mixture by Weight Minimum | % of component in blend | Percentage | 80 |
| Numeric Qualifier | Numeric Qualifier | Text | > |
| % of Mixture by Weight Maximum | % of component in blend | Percentage | 90 |
| Source | Database, MSDS, etc. | Text | CRC Safety Datasheet |
| Reference | Citation | Text or hyperlink | Brakleen® Non-Chlorinated Brake Parts Cleaner MSDS |

## DATA ENTRY

As part of the process of loading data into database the user will have a log in window complete with a user name and password assigned by CEC.

The data entry system will be a web-based form system with a wizard to guide the user through the data entry steps. The wizard will prompt the user where to enter the data. The user will be presented with a list of CAS numbers or a blend option in an effort to maintain data integrity. If the contaminant is identified as a mixture and does not have an associated CASRN then there will be a different protocol to follow. The user will be prompted with a warning system and/ or tips from the data wizard.

Next the user will be prompted to enter general information followed by chemical and physical properties. This wizard will allow the user to choose the appropriate fields from a drop down menu. The user will be prompted to use drop down/look up lists when possible. A sample of a draft user interface used to facilitate data entry.

1. **Contaminant**

**Select a contaminant:**

By CASRN:

123-45-6789 v

By Chemical Name:

Phenol v

*OR*

**Add a contaminant:**

| **CASRN:** | **Contaminant Name:** | (Add Chemical) |
|---|---|---|

2. **DOT Codes**

| **DOT Code** | **Field 2** | **Data Source** | **Reference** | |
|---|---|---|---|---|
| DOT Code 1 | ….. | ….. | ….. | (Delete DOT Code) |
| DOT Code 2 | ….. | ….. | ….. | (Delete DOT Code) |

(Add a DOT Code)

3. **UN/NA Numbers**

| **UN/NA Number** | **Field 2** | **Data Source** | **Reference** | |
|---|---|---|---|---|
| UN/NA Number 1 | ….. | ….. | ….. | (Delete UN/NA Number) |
| UN/NA Number 2 | ….. | ….. | ….. | (Delete UN/NA Number) |

(Add a UN/NA Number)

4. **Alternative Names**

| **Alternative Name** | **Field 2** | **Data Source** | **Reference** | |
|---|---|---|---|---|
| Alternative Name 1 | ….. | ….. | ….. | (Delete Alternative Name) |
| Alternative Name 2 | ….. | ….. | ….. | (Delete Alternative Name) |

(Add an Alternative Name)

5. **Purposes**

| **Purpose** | **Field 2** | **Data Source** | **Reference** | |
|---|---|---|---|---|
| Purpose 1 | ….. | ….. | ….. | (Delete Purpose) |
| Purpose 2 | ….. | ….. | ….. | (Delete Purpose) |

(Add a Purpose)

The user will be prompted with series of tables for each parameter listed in the SOP. The same type of user interface will be used during the QA/QC process.

## DATABASE MANAGEMENT

A primary necessity of managing data in a relational database is to ensure that each data record is unique and that the information contained within each field is consistent with conventions defined in other areas of the database. To ensure uniqueness, a key field, the CAS registry number, will be identified for each data record. In the case of blends/mixtures a unique mix ID will be assigned. Additional features include

- Primary data tables that store the contaminant data
- Dependent tables that store more details related to the data in the primary tables
- Look-up (domain) tables that store valid values to provide input to the primary tables
- Valid values are critical to any large relational database. Inconsistencies in naming conventions, spelling differences, and the use of nonstandard abbreviations can result in lost data and incorrect conclusions. Most tables and forms in the database will use look-up tables for acceptable valid values and will not allow the entry of data that do not conform.

## QUALITY ASSURANCE/QUALITY CONTROL

CEC will implement QA/QC measures to provide an accurate representation of all data collected and stored in WaterSuite. QA/QC procedures include restricting data import or entry to specific valid value lists that will not allow incorrect data to be included in the database. Exceptions will be made as disparate datasets may not fit in to the data confines, but will be reviewed by CEC staff members for accuracy. Remarks indicating deviations from the SOP will be documented the database in a designated field.

Each university will QA/QC 25% of the contaminants documented by another university. QA/QC will occur one week after data collection and may take a month to complete the QA/QC process. In addition, CEC will review 10% of the data entry from all universities.

QA/QC will entail clicking through listed data sources, checking numeric values, units etc. The user will be prompted to handle any necessary corrections during the QA/QC process. There will be a method of tracking QA/QC. As part of the QA/QC process each data entry will be date stamped. Errors will be corrected & documented at the same time.

BENZENE

**CAS:**
71-43-2

**Alternative Names:**
UN 1114
Annulene
Phene
Benceno
Pyrobenzol
1,3,5-Cyclohexatriene
Cyclohexatriene
Coal naptha
Benzole
Phenyl hydride
Benceno, puro
NSC 67315
Pyrobenzole

**NFPA Hazard Rating:**
Health: 2, Can cause temporary incapacitation or residual injury
Flammability: 3, Can be ignited under almost all ambient temperature conditions.
Instability: 0, Normally stable, even under fire conditions.
Special

## General Information

Purpose:

For benzene (EPA/OPP Pesticide Code: 008801) there are 0 labels match. SRP: Not registered for current use in the United States, but approved pesticide uses may change periodically and so federal, state and local authorities must be consulted for currently approved uses.

Manufacture of industrial chemicals such as polymers, detergents, pesticides pharmaceuticals, dyes, plastics, resins. Organic solvent for waxes, resins, oils, natural rubber, etc. Reference for quantitating compounds. Gasoline additive.

Used for printing and lithography, paint, rubber, dry cleaning, adhesives and coatings, detergents

Extraction and rectification; preparation and use of inks in the graphic arts industries; as a thinner for paints; as a degreasing agent

Chem intermediate for ethylbenzene, cumene, cyclohexane, nitrobenzene, maleic anhydride, chlorobenzenes, detergent alkylate, anthraquinone, benzene hexachloride, benzene sulfonic acid, biphenyl, hydroquinone, & resorcinol.

Used primarily as a raw material in the synthesis of styrene (polystyrene plastics and synthetic rubber), phenol (phenolic resins), cyclohexane (nylon), aniline, maleic anhydride (polyester resins), alkylbenzenes (detergents), chlorobenzenes, and other products used in the production of drugs, dyes, insecticides, and plastics.

MEDICATION (VET) (See also: Therapeutic Uses)

Manufacture of styrene, phenol, detergents, organic chemicals, pesticide, plastics and resins, synthetic rubber, aviation fuel, pharmaceuticals, dye, explosives, PCB gasoline, tanning, flavors and perfumes, paints and coatings; nylon intermediates; food processing; photographic chemicals.

Uses: Ethylbenzene/styrene 53%; cumene/phenol 22%; cyclohexane 12%; nitrobenzene/aniline 5%; detergent alkylate 3%; chlorobenzenes and other uses 5%

In the past benzene has been used in the shoe and garment industry as a solvent for natural rubber. Benzene has also found limited application in medicine for the treatment of certain blood disorders, such as polythemia and malignant lymphoma. Benzene, along with other light high octane aromatic hydrocarbons such as toluene and xylene, is used as a component of motor gasoline. Although this use has been largely reduced in the United States, benzene is still used extensively in many countries for the production of commercial gasoline.

| | | |
|---|---|---|
| Manufacturers: | Please see attached table | |
| Regulations: | CERCLA RQ* | 10 |
| | *Reportable Quantity | |
| EPCRA 313 TRI | 313* | |

*Indicates the chemical is subject to reporting under section 313 and section 6607 of the Pollution Prevention Act

| | |
|---|---|
| RCRA Code | U019 |

**Physical Properties**

| | |
|---|---|
| Color: | Clear, colorless liquid |
| Odor: | Aromatic odor |

Gasoline-like odor; rather pleasant aromatic odor
odor threshold 4.68 ppm

| | |
|---|---|
| Taste: | Taste threshold in water 0.5 - 4.5 mg/l |

**Fate & Transport**

| | |
|---|---|
| Specific gravity: | 0.8787 at 15 deg C |
| Solubility: | 1.79 x $10^{-3}$ mg/l at 25 deg C |
| Vapor Pressure: | 75 mm Hg at 20 deg C; 94.8 mm Hg at 25 deg C |
| Henry's Law constant: | 5.56 x $10^{-3}$ atm-cu m/mol at 25 deg C |
| Partition coefficient: | 2.13 |
| Decay constant: | *No data found in first tier search* |
| Reaction products: | *No data found in first tier search* |

Reactivity: Reactivity: 0. NFPA Hazard Classification 0. This degree includes materials that are normally stable, even under fire exposure conditions, and that do not react with water. Normal firefighting procedures may be used.

BENZENE reacts vigorously with allyl chloride or other alkyl halides even at -70° C in the presence of ethyl aluminum dichloride or ethyl aluminum sesquichloride. Explosions have been reported [NFPA 491M 1991]. Ignites in contact with powdered chromic anhydride [Mellor 11:235 1946-47]. Incompatible with oxidizing agents such as nitric acid. Mixtures with bromine trifluoride, bromine pentafluoride, iodine pentafluoride, iodine heptafluoride and other interhalogens can ignite upon heating [Bretherick 5th ed. 1995]. Benzene and cyanogen halides yield HCl as a byproduct (Hagedorn, F. H. Gelbke, and Federal Republic of Germany. 2002. Nitriles. In Ullmann's Encyclopedia of Industrial Chemistry. Wiley-VCH Verlag GmbH & Co. KGaA.). The reaction of benzene and trichloroacetonitrile evolves toxic chloroform and HCl gases. (Hagedorn, F., H.-P. Gelbke, and Federal Republic of Germany. 2002. Nitriles. In Ullmann's Encyclopedia of Industrial Chemistry. Wiley-VCH Verlag GmbH & Co. KGaA.).

## Detection and Treatability

Treatability:

The following processes used in drinking water treatment were found to be effective for the removal of benzene: aeration and air stripping (65 to greater than 99 percent), granular activated carbon (greater than 99 percent), ozone (99 percent), ultraviolet irradiation (35 to 99 percent with medium pressure lamp without $TiO_2$), ultraviolet irradiation + hydrogen peroxide (94 to 99 percent removal) and ultraviolet irradiation + ozone (99 percent). It should be noted that EPA has identified packed tower aeration and granular activated carbon as best available technologies capable of removing volatile organic compounds from water to their respective MCLs.

Processes found to be ineffective for the removal of benzene were ultraviolet irradiation (7 percent with low pressure lamp), and ultraviolet irradiation with $TiO_2$ (10 to 63 percent).

Other processes not typical to drinking water treatment were suggestive of benzene removal. They were: macroreticular resins and surfactant-modified zeolite clays, biological treatment, sonochemical oxidation, and electrochemical and radiation treatment.

| | | |
|---|---|---|
| Method detection limit: | US EPA 502.2 | 0.01 µg/L |
| | US EPA 8260/524.2 | 0.04 µg/L |
| | US EPA 8020 | 0.2 µg/L |

## HMIS hazard rating:

Health 3*, Fire 3, Reactivity 0 *Chronic health hazard (Citgo, MSDS)

## Health Effects and Medical Treatment

Health effects:

Benzene is irritant to skin.

A severe eye and moderate skin irritant.

Benzene is irritant to skin, & by defatting the keratin layer may cause erythema, vesiculation, & dry & scaly dermatitis.

Protective equipment:

Protective clothing consisting of coveralls or other full body clothing should be worn and changed at least twice weekly.

Where there is a possibility of benzene contact to eyes or skin, safety showers, eye-wash fountains, and cleansing facilities shall be installed and maintained.

Where high vapor concentrations are unavoidable, forced air masks should be used. Lifeline attended by ... person outside contaminated enclosure is mandatory. If skin contact is unavoidable, neoprene gloves must be worn.

Hydrocarbon vapor canister, supplied air or a hose mask; hydrocarbon insoluble rubber or plastic gloves; chemical goggles or face splash shield; hydrocarbon-insoluble apron such as neoprene.

PRECAUTIONS FOR "CARCINOGENS": ... Dispensers of liquid detergent should be available. ... Safety pipettes should be used for all pipetting. ... In animal laboratory, personnel should ... wear protective suits (preferably disposable, one-piece & close-fitting at ankles & wrists), gloves, hair covering & overshoes. ... In chemical laboratory, gloves & gowns should always be worn ... however, gloves should not be assumed to provide full protection. Carefully fitted masks or respirators may be necessary when working with particulates or gases, & disposable plastic aprons might provide additional protection. ... Gowns ... should be of distinctive color, this is a reminder that they are not to be worn outside the laboratory. Chemical Carcinogens

Performance data: For butyl rubber, natural rubber, neoprene, neoprene, neoprene/natural rubber, nitrile rubber, polyethylene, chlorinated polyethylene, polyurethane, and polyvinyl chloride give breakthrough times less (usually significantly less) than one hour reported by (normally) two or more testers. Vendor Recommendations: C or D ratings from three or more (apparently independent) vendors.

Wear appropriate personal protective clothing to prevent skin contact.

Wear appropriate eye protection to prevent eye contact.

# APPENDIX F
# DATA GAPS

## Table F.1 Contaminant database gaps

*CAS – Chemical Abstract Service; LD50 – Lethal Dose 50%; RFD – Reference Dose; NOAEL – No Observed Adverse Effect Level; HAZ – Hazard Category; UN – United Nations number; DOT – US Department of Transportation number; MDL – Method Detection Limit; X – Data Found; A – Data Unavailable; B – Data Not Applicable; C – Data Not Found*

| CAS | CONTAMINANT | Alternative Names | Color | Odor | Taste | Specific Gravity | Vapor Pressure | Solubility | Henry's Law Constant | LD50 | RFD | NOAEL | HAZ | Health Effects | Medical Treatment | Protective Equipment | Purpose | Manufacturer | UN | DOT | Partition Coefficient | Reactivity | MDL | Treatability |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 71-43-2 | Benzene | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X |
| 108-95-2 | Phenol | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X |
| 107-21-1 | Ethylene Glycol | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | B | X | X | X | X | X |
| 108-88-3 | Toluene | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X |
| 1634-04-4 | Methyl Tert-butyl Ether | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X |
| 68334-30-5 | Diesel Fuel (Ultra Low Sulfur) | X | X | X | B | X | X | X | X | X | X | X | X | X | X | X | X | B | X | X | X | X | X | X |
| 68476-34-6 | Diesel Fuel #2 | X | X | X | B | X | X | X | X | X | X | X | X | X | X | X | X | B | X | X | X | X | X | X |
| 64742-47-8 | Petroleum Distillate Blend (2-26) | X | X | X | B | C | X | X | X | X | X | X | X | X | X | X | X | B | X | X | X | X | X | X |
| 34885-03-5 | 4-Methylcyclohexanemethanol | X | X | X | B | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X |
| 770-35-4 | Propylene glycol phenyl ether | X | X | X | A | X | X | X | X | X | X | X | X | X | X | X | X | X | C | C | X | X | X | X |
| 51730-94-0 | Dipropylene glycol phenyl ether | X | X | X | A | X | X | X | X | X | A | A | X | X | X | X | X | X | C | C | X | X | A | X |
| 100-41-4 | Ethylbenzene | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X |
| 110-54-3 | N-hexane | X | X | X | A | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X |
| 95-63-6 | 1,2,4-Trimethylbenzene | X | X | X | B | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X |
| 1330-20-7 | Xylene (Mixed Isomers) | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X |
| 107-19-7 | Propargyl alcohol | X | X | X | A | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X |
| 67-56-1 | Methanol | X | X | X | A | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X |

| CAS | CONTAMINANT | Alternative Names | Color | Odor | Taste | Specific Gravity | Vapor Pressure | Solubility | Henry's Law Constant | LD50 | RFD | NOAEL | HAZ | Health Effects | Medical Treatment | Protective Equipment | Purpose | Manufacturer | UN | DOT | Partition Coefficient | Reactivity | MDL | Treatability |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 12125-02-9 | Ammonium chloride | X | X | X | X | X | X | X | A | X | X | X | X | X | X | X | X | X | C | A | X | X | X | X |
| 10222-01-2 | 2,2-Dibromo-3-nitrilopropionamide | X | X | X | A | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X |
| 3252-43-5 | Dibromoacetonitrile | X | X | X | A | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X |
| 25322-68-3 | Polyethylene glycol | X | X | X | B | X | X | X | X | X | A | X | X | X | X | X | X | X | C | X | X | X | X | X |
| 7647-01-0 | Hydrochloric acid | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X |
| 7775-27-1 | Sodium persulfate | X | X | X | X | X | X | X | X | X | A | X | X | X | X | X | X | C | X | X | A | X | X | X |
| 64742-47-8 | Hydrotreated Light petroleum distillate | X | X | X | B | C | X | X | X | X | X | X | X | X | X | X | X | C | X | X | X | X | X | X |
| 81741-28-8 | Tributyl tetradecyl phosphonium chloride | X | X | X | A | X | X | X | B | X | A | A | X | X | X | X | X | X | X | C | A | X | A | X |
| 104-76-7 | 2-ethyl hexanol | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X |
| 57-55-6 | Propylene glycol | X | X | X | X | X | X | X | A | X | X | X | X | X | X | X | X | X | C | X | X | X | X | X |
| 542-88-1 | Bis(chloromethyl)ether | X | X | X | A | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X |
| 117-81-7 | Bis(2-ethylhexyl)phthalate | X | X | C | A | X | X | X | X | X | X | X | X | X | X | X | X | X | C | A | X | X | X | X |
| 7440-36-0 | Antimony | X | X | X | X | X | X | X | B | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X |
| 7440-47-3 | Chromium | X | X | X | X | X | X | X | B | X | X | X | X | X | X | X | X | X | X | X | A | X | X | X |
| 7440-02-0 | Nickel | X | X | X | X | X | X | X | B | X | X | X | X | X | X | X | X | X | C | A | A | X | X | X |
| 335-67-1 | Perfluorooctanoic acid | X | X | X | A | X | X | X | X | X | X | X | X | X | X | X | X | X | X | C | X | X | X | X |
| 1763-23-1 | Perfluorooctane sulfonic acid | X | X | A | A | A | X | X | A | X | X | X | X | X | X | X | X | C | C | C | X | X | X | X |
| 1336-21-6 | Ammonium hydroxide | X | X | X | X | X | X | X | A | X | A | A | X | X | X | X | X | X | X | X | A | X | X | X |
| 7664-93-9 | Sulfuric acid | X | X | X | X | X | X | X | X | X | A | A | X | X | X | X | X | X | X | X | X | X | X | X |
| 111-76-2 | 2-Butoxyethanol | X | X | X | A | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X |

| CAS | CONTAMINANT | Alternative Names | Color | Odor | Taste | Specific Gravity | Vapor Pressure | Solubility | Henry's Law Constant | LD50 | RFD | NOAEL | HAZ | Health Effects | Medical Treatment | Protective Equipment | Purpose | Manufacturer | UN | DOT | Partition Coefficient | Reactivity | MDL | Treatability |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 67-64-1 | Acetone | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X |
| 10043-01-3 | Aluminum sulfate | X | X | X | X | X | X | X | A | X | X | X | X | X | X | X | X | X | X | X | A | X | X | X |
| 156-60-5 | trans-1,2-dichloroethylene | X | X | X | A | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X |
| 7705-08-0 | Ferric chloride | X | X | X | A | X | X | X | A | X | A | A | X | X | X | X | X | X | X | X | A | X | A | X |
| 7758-94-3 | Ferrous chloride | X | X | X | A | X | X | X | A | X | A | A | X | X | X | X | X | X | X | X | A | X | X | X |
| 142-82-5 | Heptane | X | X | X | X | X | X | X | X | X | A | X | X | X | X | X | X | X | X | X | X | X | X | X |
| 7681-52-9 | Sodium hypochlorite | X | X | X | X | X | X | X | B | X | A | A | X | X | X | X | X | X | X | X | A | X | X | X |
| 8052-41-3 | Stoddard solvent | X | X | X | X | X | X | X | X | X | A | A | X | X | X | X | X | C | X | X | X | X | X | X |
| 127-18-4 | Tetrachloroethylene | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X |
| 79-01-6 | Trichloroethylene | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X |
| 67-66-3 | Chloroform | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X |
| 124-48-1 | Dibromochloromethane | X | X | X | A | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X |
| 75-27-4 | Bromodichloromethane | X | X | X | A | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X |
| 7439-92-1 | Lead | X | X | X | X | X | X | X | B | A | X | B | X | X | X | X | X | X | C | A | X | X | X | X |
| 10043-52-4 | Liquid calcium chloride | X | X | X | X | X | A | X | A | X | A | X | X | X | X | X | X | X | C | A | A | X | X | X |
| 7782-44-7 | Liquid Oxygen | X | X | X | X | X | X | X | X | A | A | A | X | X | X | X | X | X | X | X | X | X | X | X |
| 124-38-9 | Carbon dioxide/ refrigerated liquid | X | X | X | X | X | X | X | X | X | B | B | X | X | X | X | X | X | X | X | X | X | X | X |
| 74-98-6 | Propane | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X |
| 68425-31-0 | Pipeline-Natural Gas Liquids | X | X | X | A | X | X | X | X | X | X | X | X | X | X | X | C | B | X | X | X | X | X | X |
| 7727-37-9 | Nitrogen Liquefied | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X |
| 8008-20-6 | Kerosene | X | X | X | A | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X |

| CAS | CONTAMINANT | Alternative Names | Color | Odor | Taste | Specific Gravity | Vapor Pressure | Solubility | Henry's Law Constant | LD50 | RFD | NOAEL | HAZ | Health Effects | Medical Treatment | Protective Equipment | Purpose | Manufacturer | UN | DOT | Partition Coefficient | Reactivity | MDL | Treatability |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 68476-85-7 | Liquefied Natural gas | X | X | X | A | X | X | X | B | X | X | X | X | X | X | X | C | B | X | X | A | X | X | X |
| 1317-65-3 | Limestone | X | X | X | X | X | A | X | B | X | C | X | X | X | X | X | C | C | C | A | A | X | X | X |
| 67-63-0 | Isopropyl alcohol | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X |
| 107-83-5 | Isohexane | X | X | X | A | X | X | X | X | X | A | A | X | X | X | X | X | X | X | X | X | A | X | X |
| 67254-74-4 | Napthenic oil | C | X | X | A | C | A | X | A | A | A | A | X | X | X | X | C | B | A | A | A | X | X | X |
| 74-84-0 | Ethane | X | X | X | A | X | X | X | X | X | A | A | X | X | X | X | X | X | X | X | X | X | X | X |
| 75-28-5 | Isobutane | X | X | X | A | X | X | X | X | X | A | X | X | X | X | X | X | X | X | X | X | X | X | X |
| 106-97-8 | Normal butane | X | X | X | A | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X |
| 110-82-7 | Cyclohexane | X | X | X | A | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X |
| 68476-30-2 | Fuel oil number 2 | C | X | X | A | X | X | X | A | X | X | X | X | X | X | X | C | B | X | X | X | X | X | X |
| 8006-61-9 | Unleaded gasoline | X | X | X | X | X | X | X | X | X | A | X | X | X | X | X | X | X | X | X | A | X | X | X |
| 68649-42-3 | Phosphorodithioic acid, O,O-di-C1-14-alkyl esters, zinc salts | X | X | X | A | C | A | X | A | A | A | A | X | X | X | A | A | X | C | C | A | X | A | X |
| 811-97-2 | 1,1,1,2-tetrafluoroethane | X | X | X | X | A | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X |
| 431-89-0 | 1,1,1,2,3,3,3-Heptafluoropropane | X | X | X | X | X | X | X | X | X | A | A | X | X | X | X | X | X | X | X | X | X | X | X |
| 1306-19-0 | Cadmium Oxide | X | X | X | X | X | X | X | A | X | X | A | X | X | X | X | X | X | X | X | A | X | A | X |
| 7782-50-5 | Chlorine | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X |
| 2139900 | Potassium Phosphate Dibasic | X | X | X | A | X | A | X | A | X | A | A | X | X | X | X | X | A | C | C | A | X | X | X |
| 64475-85-0 | Solvent 140 | X | X | X | A | X | X | X | B | X | A | A | X | X | X | X | C | B | C | C | X | X | A | X |
| 57-13-6 | Urea | X | X | X | X | X | X | X | X | X | A | X | X | X | X | X | X | X | C | A | X | X | X | X |
| 7779-90-0 | Zinc phosphate | X | X | X | A | X | A | X | B | X | X | X | X | X | X | X | X | B | C | C | A | X | X | X |

| CAS | CONTAMINANT | Alternative Names | Color | Odor | Taste | Specific Gravity | Vapor Pressure | Solubility | Henry's Law Constant | LD50 | RFD | NOAEL | HAZ | Health Effects | Medical Treatment | Protective Equipment | Purpose | Manufacturer | UN | DOT | Partition Coefficient | Reactivity | MDL | Treatability |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 8002-05-9 | Crude Oil | X | X | X | X | X | A | X | A | X | X | X | X | X | X | X | X | B | X | X | X | X | X | X |
| 70892-10-3 | Fuel Oil Number 1 | X | X | X | A | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X |
| 101043-37-2 | Microcystin-LR | X | X | X | X | A | X | X | A | X | X | X | C | X | X | X | C | X | B | C | X | X | X | X |
| 001310-73-2 | Sodium hydroxide | X | X | X | X | X | X | B | C | X | X | X | X | X | X | X | X | X | X | X | X | X | C | X |
| 7732-18-5 | Water | X | X | X | X | X | X | X | A | X | C | B | B | X | X | X | X | B | C | X | B | X | C | X |
| 7647-14-5 | Sodium chloride | X | X | X | X | X | X | X | C | X | C | X | X | X | X | X | X | X | C | X | A | X | C | X |
| 7697-37-2 | Nitric acid | X | X | X | X | X | X | X | X | C | X | X | X | X | X | X | X | X | X | X | X | X | C | X |
| 7664-38-2 | Phosphoric acid | X | X | X | X | X | X | X | C | X | X | X | X | X | X | X | X | X | X | X | A | X | C | X |
| 91-20-3 | Naphthalene | X | X | X | A | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | C | X |
| 7446-70-0 | Aluminum chloride | X | X | X | A | X | X | X | C | X | X | A | X | X | X | X | X | X | X | X | A | X | C | X |
| 16961-83-4 | Hydrofluorosilicic acid | X | X | X | A | X | X | X | C | X | C | A | X | X | X | X | X | X | X | X | A | X | C | X |

# APPENDIX G
# SAMPLE LETTER TO LEPC

Date

*[Address 1]*
*[City, State, zip]*

Dear *[LEPC Coordinator],*
*[Water Utility]* is completing a source water vulnerability assessment for *[plant name]* Treatment Plant. As part of this assessment, we are collecting information about storage tank locations and contents as well as the locations and volumes of toxic and hazardous materials stored, carried and used in the upstream watershed. This information will be entered into a password-protected geographic information system that links the locations of specific contaminants with chemical profiles that will help *[Water Utility]* to plan for and respond to future spill or other contamination events.

Our focus area includes *[county name]* County. In particular, we have limited the scope of our request to the area as shown in the map included in this request [attach map of the Zone of Concern]. To build the database required for our assessment, we would like to request the following materials:

Tier II Emergency and Hazardous Chemical Inventory report data for facilities within [County Name], including detailed descriptions of the locations, quantities and type of storage facilities for each toxic or hazardous material, in particular those that are stored in liquid form in above ground storage tanks.

Information on any pipelines within the above area, including pipe diameter, operating pressure, product transported, and potential impact radius if available.

Any commodity flow studies of hazardous materials hauled over streets and highways, or any hazardous route designations in your county or region.

Hazardous Material Commodity Flow or Density Study reports for all freight rail lines with track within the above area. According to the Federal Rail Agency map, the following rail companies haul freight through St. Charles County:
[list of rail lines]

Electronic Tier II data are preferred, as we will be adding them directly to our database; where data are available electronically they may be emailed to [contact name, email & phone]. If any data are only available as hard copy, these may be mailed to [contact name], at the address below.

Thank you for your assistance in this matter. Please do not hesitate to contact me if you have any questions, and feel free to forward this request to any other appropriate contact at [county name].

Sincerely,
[signature]
[name]

# REFERENCES


American Chemical Society. (2015, July 27). *ACS Publications*. Retrieved from http://pubs.acs.org/

American Water Works Association. 2014. AWWA Management Standard, Source Water Protection (ANSI/AWWA G300-14). USA: American Water Works Association.

Camp, J., LeBoeuf, E., & Abkowitz, M. (2010). Application of an enhanced spill management information system to inland waterways. *Journal of Hazardous Materials, 175*, 583-592.

Cao, W., Lia, J., Joksimovica, D., Yuana, A., & Banting, D. (2013). Probabilistic spill occurrence simulation for chemical spills management. *Journal of Hazardous Materials, 262*, 517-526.

EPA. 1974 (amended 1986 and 1996). Safe Drinking Water Act. USC Title 42 (Public Health and Welfare), Chapter 6A (Public Health Service), Subchapter XII (Safety of Public Water Systems). Web: https://www.gpo.gov/fdsys/pkg/USCODE-2010-title42/pdf/USCODE-2010-title42-chap6A-subchapXII.pdf and https://www.epa.gov/sites/production/files/2015-04/documents/epa816f04030.pdf.

EPA. (2014, February 11). *Web Services*. Retrieved from EPA United States Environmental Protection Agency: http://water.epa.gov/scitech/datait/tools/waters/services/index.cfm

EPA. (2015, June 4). *Rest Services*. Retrieved from USGS Science for a Changing World: http://waterservices.usgs.gov/rest/

EPA. (2015, July 24). *Water Contamination Information Tool (WCIT): A Robust Tool for the Water Sector*. Retrieved from http://water.epa.gov/scitech/datait/databases/wcit/upload/WCIT_Factsheet_FINAL-2.pdf

Foster, J., & McDonald, A. (2000). Assessing pollution risks to water supply intakes using geographical information systems (GIS). *Enviornmental Modelling and Software, 15*, 225-234.

Hunt, Charles D., 1979, National Atlas of the United States of America--Surficial Geology: U.S. Geological Survey, NAC-P-0204-75M-O, [map]. (Accessed as part of NHD dataset: https://www.sciencebase.gov/catalog/item/5703f830e4b0328dcb826dd7).

Jobson HE. 1989. User’s manual for an open-channel streamflow model based on the diffusion analogy. Report 89-4133. Available at: http://pubs.er.usgs.gov/publication/wri894133.

Jobson, H. (1996). Prediction of traveltime and longitudinal dispersion in rivers and streams. *U.S. Geological Survey Water Resources Investigations Report 96-4013*, p. 69. Retrieved from http://pubs.usgs.gov/of/1996/4013

Martin, P., LeBoeuf, E., Danieal, E., Dobbins, J., & Abkowitz, M. (2004). Development of a GIS-based spill management information system. *Journal of Hazardous Materials, B112*, 239-252.

Reed, J. C, and C. A Bush. 2005. Generalized geologic map of the conterminous United States. The University of Chicago Press, Denver, Colorado. (Available from: http://pubs.usgs.gov/atlas/geologic/48States/, accessed as part of NHD dataset: https://www.sciencebase.gov/catalog/item/5703f6b5e4b0328dcb826d06).

Ries, K.G., III, 2006, The National Streamflow Statistics Program: A computer program for estimating streamflow statistics for ungaged sites: U.S. Geological Techniques and Methods Report TM Book 5, Chapter A6, 45 p.

Ries III, K. (2007). The national streamflow statistics program: A computer program for estimating streamflow statistics for ungaged sites. *U.S. Geological Survey Techniques* *and Methods 4-A6*, p. 37. Retrieved from http://pubs.usgs.gov/tm/2006/tm4a6/pdf/tm4a6.pdf

Royal Society of Chemistry. (2015, July 27). *Publishing: Journals, books and databases* . Retrieved from http://pubs.rsc.org/

Ryan, D. (2007, January). ICWater: A New Software Tool for Protecting the Nation's Drinking Water. *Stream Notes*, pp. 1-4. Retrieved 7 24, 2015, from http://www.stream.fs.fed.us/news/streamnt/pdf/SN_01_07.pdf

Ryan, D. a. (nd). *Incident Command Tool for Drinking Water Protection.* Retrieved July 24, 2015, from Horizon Systems: http://www.horizon-systems.com/nhdplus/applications/ICWater_Factsheet.pdf

Samuels, R. (2015, July 24). *Questions on Riverspill/ICWater - National Hydrography Dataset.* Retrieved from nhd.usgs.gov: nhd.usgs.gov/documents/Questions%20and%20Responses%20from%20Dr%20Samuels.docx

Samuels, W. B. (2001, December). RiverSpill: A GIS-based Real Time Transport Model for Source Water Protection. *Proceedings of the Water Environment Federation*. doi:10.2175/193864702785665328

Sham, C., Gullick, R., Long, S., & Kenel, P. (2010). *Operational Guide to AWWA Standard G300.* Denver: American Water Works Association.

USGS Water Services. Available at: http://waterservices.usgs.gov/ [Accessed continuously throughout 2015]

UT Dallas. (2015, July 24). *EPlan, Frequently Asked Questions (FAQ)*. Retrieved from EPlan News: http://eplannews.utdallas.edu/FAQs.htm

# ABBREVIATIONS

| | |
|---|---|
| ATSDR | Agency for Toxic Substances and Disease Registry |
| AWWA | American Water Works Association |
| | |
| CAS | Chemical Abstract Service |
| CASRN | Chemical Abstract Service Registry Number |
| CDR | Chemical Data Reporting Rule |
| CEC | Corona Environmental Consulting |
| CERCLIS | Comprehensive Environmental Response, Compensation and Liability Information System |
| DEM | Digital Elevation Model |
| DEP | Department of Environmental Protection |
| DHS | Department of Homeland Security |
| DOH | Department of Health |
| DOT | Department of Transportation |
| DTB | Drinking Water Treatability Database |
| | |
| EIA | Energy Information Administration |
| EPA | United States Environmental Protection Agency |
| ESRI | Environmental Systems Research Institute |
| | |
| FRS | Facility Registry System |
| | |
| GIS | Geographic Information System |
| GCWW | Greater Cincinnati Water Works |
| | |
| IRIS | Integrated Risk Information System |
| | |
| LD50 | Lethal Dose, 50% - the amount of toxic agent sufficient to kill 50% of a population of animals within a certain time period |
| LEPC | Local Emergency Planning Committee |
| LIDAR | Light Detection and Ranging |
| | |
| MSDS | Material Safety Data Sheet |
| | |
| NAICS | North American Industry Classification System |
| NHD | National Hydrography Dataset |
| NOAEL | No Observed Adverse Effect Level |
| NPDES | National Pollutant Discharge Elimination System |
| | |
| ORD | Office of Research and Development (EPA) |
| ORP | Oxidation Reduction Potential |
| | |
| PAC | Powdered Activated Carbon |
| POI | Point of Interest |

| | |
|---|---|
| POTW | Publicly Owned Treatment Works |
| PSC | Potential Source of Contamination |
| PWS | Public Water System |
| | |
| RfD | Reference Dose |
| RCRA | Resource Conservation and Recovery Act |
| | |
| SCADA | Supervisory Control and Data Acquisition |
| SDWA | Safe Drinking Water Act |
| SERC | State Emergency Response Commission |
| SOP | Standard Operating Procedure |
| SWA | Source Water Assessment |
| SWP | Source Water Protection |
| | |
| TOXNET | Toxicology Data Network |
| TRI | Toxics Release Inventory |
| | |
| USACE | United States Army Corps of Engineers |
| USGS | United States Geological Survey |
| | |
| WRF | Water Research Foundation |
| | |
| ZCC | Zone of Critical Concern |