

# **Tools Used for Visualizing Sampling and Analysis Data During Response to a Contamination Incident**

## **DISCLAIMER**

The U.S. Environmental Protection Agency, through its Office of Research and Development, funded and managed the research described here under Contract EP-C-16-015 to Eastern Research Group, Inc. It has been subjected to the Agency's review and has been approved for publication. Note that approval does not signify that the contents necessarily reflect the views of the Agency. Mention of trade names, products, or services does not convey official EPA approval, endorsement, or recommendation.

Questions concerning this document, or its application should be addressed to:

Erin Silvestri  
U.S. Environmental Protection Agency  
Office of Research and Development  
Center for Environmental Solutions and Emergency Response  
26 West Martin Luther King Drive (NG16)  
Cincinnati, OH 45268  
Phone 513.569.7619

## FOREWORD

The U.S. Environmental Protection Agency (EPA) is charged by Congress with protecting the Nation's land, air, and water resources. Under a mandate of national environmental laws, the Agency strives to formulate and implement actions leading to a compatible balance between human activities and the ability of natural systems to support and nurture life. To meet this mandate, EPA's research program is providing data and technical support for solving environmental problems today and building a science knowledge base necessary to manage our ecological resources wisely, understand how pollutants affect our health, and prevent or reduce environmental risks in the future.

The Center for Environmental Solutions and Emergency Response (CESER) within the Office of Research and Development (ORD) conducts applied, stakeholder-driven research and provides responsive technical support to help solve the Nation's environmental challenges. The Center's research focuses on innovative approaches to address environmental challenges associated with the built environment. We develop technologies and decision-support tools to help safeguard public water systems and groundwater, guide sustainable materials management, remediate sites from traditional contamination sources and emerging environmental stressors, and address potential threats from terrorism and natural disasters. CESER collaborates with both public and private sector partners to foster technologies that improve the effectiveness and reduce the cost of compliance, while anticipating emerging problems. We provide technical support to EPA regions and programs, states, tribal nations, and federal partners, and serve as the interagency liaison for EPA in homeland security research and technology. The Center is a leader in providing scientific solutions to protect human health and the environment.

This report describes research that was conducted to identify commonalities, efficiencies, lessons learned, and knowledge gaps among data visualization and statistical analysis tools currently used throughout EPA and the federal government. Research emphasized tools that are used to visualize sampling and analysis data collected in support of remediation after an intentional or unintentional contamination incident to streamline and improve the capabilities of United States Coast Guard (USCG) and EPA responders. This research effort built upon a previous effort to identify and recommend user-friendly tools that more easily facilitate the acquisition of field sampling data and subsequent management of sampling data following a wide-area incident. Recommended tools identified through this research will be exercised during the Department of Homeland Security (DHS)/EPA-sponsored Analysis for Coastal Operational Resiliency (AnCOR) Project.

Gregory Sayles, Director  
Center for Environmental Solutions and Emergency Response

# TABLE OF CONTENTS

Disclaimer .....	ii
Foreword .....	iii
List of Tables and Figures.....	v
Abbreviations .....	vi
Acknowledgments.....	vii
Executive Summary .....	ES-1
1 Introduction .....	1
2 Quality Assurance/Quality Control .....	3
3 Literature Review Results .....	3
3.1 Dash by Plotly .....	6
3.2 Domo.....	6
3.3 Esri Suite.....	7
3.3.1 Esri ArcGIS Online (EPA’s GeoPlatform) .....	7
3.3.2 Esri ArcGIS Dashboards .....	7
3.3.3 Esri ArcGIS Insights .....	8
3.3.4 Esri ArcGIS Story Maps .....	9
3.4 Electronic Data Exchange and Evaluation System (EXES) .....	9
3.5 GeoDa .....	10
3.6 Google Data Studio .....	10
3.7 Highcharts .....	11
3.8 IBM Cloud Pak for Data .....	11
3.9 IBM SPSS Modeler/Statistics .....	12
3.10 Looker .....	13
3.11 Microsoft Power BI.....	14
3.12 MicroStrategy .....	14
3.13 Oracle Platforms .....	15
3.14 Panel.....	16
3.15 Qlik Sense .....	16
3.16 RStudio/Shiny .....	17
3.17 Statistical Analysis Software (SAS) .....	18
3.18 Sisense.....	19
3.19 STATA.....	19
3.20 Tableau.....	20

3.21	Voilà.....	21
4	Operational Expert Feedback .....	21
5	Final Recommendation.....	26
6	References .....	30

APPENDIX A. Literature Search Source Criteria and Keywords

APPENDIX B. Literature Review Questionnaire and Scoring Criteria

## LIST OF TABLES AND FIGURES

Table 1.	Tool/Software Overview.....	5
Figure 1.	General sampling phases and activities.....	1
Figure 2.	Search terms .....	3
Figure 3.	Biological sampling activities: framework and tools relationship in a wide-area biological incident.....	28

## ABBREVIATIONS

<b>Acronym</b>	<b>Definition</b>
AI	artificial intelligence
AnCOR	Analysis for Coastal Operational Resiliency
API	application programming interface
BI	business intelligence
CBRN	chemical, biological, radiological, or nuclear
CESER	Center for Environmental Solutions and Emergency Response (EPA)
CEMM	Center for Environmental Measurement and Modeling (EPA)
CPHEA	Center for Public Health and Environmental Assessment (EPA)
COVID-19	coronavirus disease 19
CSV	comma-separated values
DHS	Department of Homeland Security
DMAP	Data Management/Analytics Platform
EPA	U.S. Environmental Protection Agency
ER	emergency response
ERG	Eastern Research Group, Inc.
EXES	Electronic Data Exchange and Evaluation System
GIS	Geographic Information System
HSMMD	Homeland Security and Materials Management Division (EPA)
HTML	Hypertext Markup Language
IBM	International Business Machines Corporation
ID	identification
IT	information technology
JPG	Joint Photographic Group
LAN	Local Area Network
ML	machine learning
MQO	method quality objective
OMS	Office of Mission Support (EPA)
ORD	Office of Research and Development (EPA)
PDF	portable document format
PESD	Pacific Ecological Systems Division (EPA)
QAPP	quality assurance project plan
SAS	Statistical Analysis Software
SPSS	Statistical Package for the Social Sciences
SQL	Structured Query Language
STATA	statistics and data
SVG	Scalable Vector Graphics
USCG	U.S. Coast Guard
WECD	Watershed and Ecosystem Characterization Division
XML	Extensible Markup Language

# ACKNOWLEDGMENTS

Contributions of the following individuals and organizations to this report are acknowledged:

## **US Environmental Protection Agency (EPA) Project Team**

Erin Silvestri (Principal Investigator, EPA/ORD\*/CESER/HSMMD)

Timothy Boe (EPA/ORD/CESER/HSMMD)

Jamie Falik (EPA/ORD/CESER/HSMMD)

## **US EPA Technical Reviewers of Report**

Michael McManus (EPA/ORD/CEMM/WECD)

Marc Weber (EPA/ORD/CPHEA/PESD)

## **US EPA Quality Assurance**

Ramona Sherman (EPA/ORD/CESER/HSMMD)

## **Eastern Research Group, Inc. (ERG)**

Molly Rodgers

Amanda Speciale

\*ORD, Office of Research and Development

CESER, Center for Environmental Solutions and Emergency Response

HSMMD, Homeland Security and Materials Management Division

CEMM, Center for Environmental Measurement and Modeling

WECD, Watershed and Ecosystem Characterization Division

CPHEA, Center for Public Health and Environmental Assessment

PESD, Pacific Ecological Systems Division

## EXECUTIVE SUMMARY

In the event of a chemical, biological, radiological, or nuclear (CBRN) wide-area incident, the U.S. Environmental Protection Agency (EPA) is responsible for clearance, waste disposal processes, and data collection and quality checks to advise decision-making. The U.S. Coast Guard (USCG) shares this responsibility for certain incidents in the maritime domain. This research effort sought to identify data visualization and statistical analysis tools that are in use throughout EPA, the federal government, and commercial and/or academic settings, as well as identify efficiencies, lessons learned and knowledge gaps identified by the emergency response community. Research focused on tools that are used to visualize sampling and analysis data collected in support of remediation after an intentional or unintentional contamination incident to streamline and improve the capabilities of USCG and EPA responders. This research built upon work to identify and recommend user-friendly tools that more easily facilitate the acquisition of field sampling data and subsequent management of sampling data following a wide-area incident. Recommended tools identified through this research will be exercised during the Department of Homeland Security (DHS)/EPA-sponsored Analysis for Coastal Operational Resiliency (AnCOR) project.

EPA first conducted a literature review and market research to identify and describe data visualization and statistical analysis platforms (i.e., tools, applications, and programs) currently in use throughout EPA, the federal government, and commercial and/or academic settings. Over 30 sources were identified as having information relevant to this project's research objectives, and these sources were subsequently reviewed and summarized. This discovery research ultimately identified 21 data visualization and/or statistical analysis tools, and this report provides a brief description of each tool's relevant features.

To supplement the literature review and market research, operational expert feedback was solicited from the response and research community to understand what visualization tools are currently being used for presenting and analyzing data collected during a contamination incident response, as well as to identify efficiencies, lessons learned, and knowledge gaps based on their experience. Four important capabilities were consistently shared by operational experts who were interviewed:

1. Tool and data access and flexibility are paramount,
2. Geospatial context is critical,
3. Data should be centrally and separately managed from visualization and analysis tools to enable EPA and its partners to easily adapt to advances in technology, and
4. EPA should prioritize efforts to establish a process for creating data management plans with other agency partners to standardize and communicate data flows, identify what kind of data are generated, what formats were used, where data are sourced, and how data can be accessed.

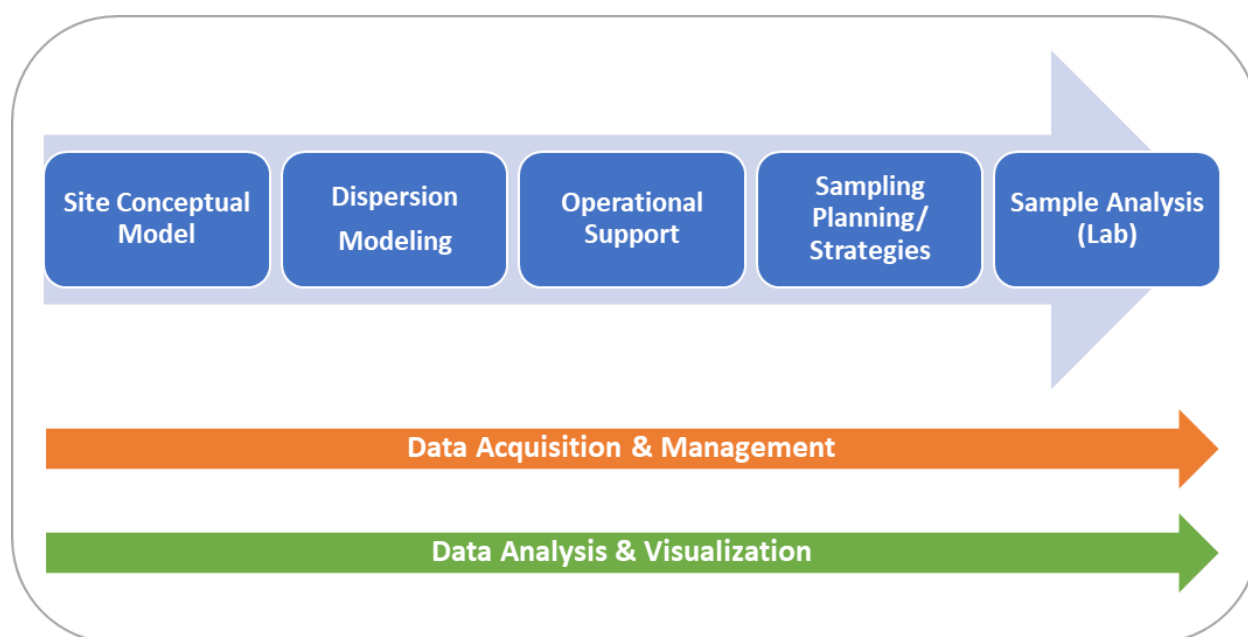
Throughout research conducted and interviews held, Esri's suite of products was consistently cited. The Esri product suite is widely adopted among the response community and is meeting the needs of response teams. The suite of Esri products has been used by the EPA regions in



support of the wildfire responses and by the USCG during incident response such as oil spills and hazardous materials, and also in support of natural disasters such as hurricanes. In support of the upcoming AnCOR field exercise, the project team recommends leveraging the work of EPA regional response teams, adopting a similar workflow that makes use of data stored in the EPA Emergency Response (ER) Cloud, visualizing data through maps on the GeoPlatform and Esri Operational dashboards, and implementing Esri Insights for analysis needs. The Esri suite of tools has the most features that meet the largest number of needs, are supported at the enterprise level, seamlessly integrate with other field data capture tools, are familiar to responders, and are generally seen as easy to customize and tailor to meet the specific needs of the operation.

# 1 INTRODUCTION

The U.S. Environmental Protection Agency (EPA) is designated as a coordinating Agency, under the National Response Framework, to prepare for, to respond to, and to support the recovery from a threat to public health, welfare, or the environment caused by actual or potential oil and hazardous materials incidents. Hazardous materials may include chemical, biological, and radiological or nuclear (CBRN) substances, whether accidentally or intentionally released. EPA can also have responsibilities to address debris and waste through decontamination, removal, and disposal operations. The U.S. Coast Guard (USCG) shares this responsibility for certain incidents in the maritime domain. As shown in Figure 1 below, EPA will play a role in several activities and phases in support of response and recovery efforts.



**Figure 1. General sampling phases and activities.**

Emergency response occurs over many phases, from the initial characterization sampling to evaluate the contamination event through clearance sampling and waste disposal processes. During all phases of the response to a wide-area CBRN incident, a substantial amount of data will need to be analyzed and visualized to support decision-making and communication needs. Much of the data that will be generated will contain a geospatial component (e.g., sampling location); therefore, it will be important to provide response personnel with the ability to analyze and view data within a geospatial context. Data visualization tools are necessary to effectively organize, document, quality assure, and communicate data during a wide-area CBRN incident. Understanding how these processes and tools are connected and work together are critical to advancing EPA's and Department of Homeland Security's (DHS's) data analysis and visualization capabilities.

This EPA project supports the Analysis for Coastal Operational Resiliency (AnCOR) project. AnCOR is a multi-agency program with the purpose of developing and demonstrating capabilities and strategic guidelines to prepare the U.S. for a wide-area release of a biological agent, including mitigating impacts to USCG facilities and assets [1]. A comprehensive screening of a range of available tools was needed to identify applicable features and current uses to identify the most efficient and compatible tools for analyzing and visualizing field sampling and laboratory data. Through this project, a resulting recommendation for how to improve and/or integrate the tools to better meet the needs of EPA during a response to a contamination event was made. Opportunities where tools may be better integrated with other web-based platforms currently being used for managing, statistically analyzing, sharing, or viewing data were identified.

This project sought to identify data visualization and statistical analysis tools that are in use throughout EPA, the federal government, and commercial and/or academic settings, as well as identify efficiencies, lessons learned, and knowledge gaps identified by the emergency response community. Research focused on tools that are used to visualize and analyze sampling and analysis data collected in support of remediation after an intentional or natural contamination incident to streamline and improve data visualization tools to better fit the needs of stakeholders within the DHS, including the USCG and EPA response community. This project extends research that was conducted in support of a related project, also in support of the AnCOR program, to identify and recommend user-friendly tools that more easily facilitate the acquisition of field sampling data and subsequent management of sampling data following a wide-area incident [2].

This project had three primary objectives:

1. **Conduct a literature review/market research** to identify and describe open source or commercial off-the-shelf data visualization and statistical analysis platforms (i.e., tools, applications, and programs) currently in use throughout EPA, the federal government, and commercial and/or academic settings,
2. **Solicit operational feedback** from stakeholders within the EPA response community regarding visualization tools currently being used for presenting and analyzing data collected during a response to a contamination incident, and
3. **Develop a final summary and recommendation report** describing recommendations for adopting and integrating improved statistical analysis and data visualization tools to enhance EPA's capabilities in support of managing data generated throughout all phases of the incident response cycle to inform decision-making during a response to a contamination incident.

This report is structured as follows:

- Chapter 2 discusses quality assurance/quality control activities,
- Chapter 3 summarizes the results of the literature review and market research,
- Chapter 4 summarizes the key takeaways from the interviews held with operational experts, and

- Chapter 5 discusses the recommendations reached from the research conducted.

## 2 QUALITY ASSURANCE/QUALITY CONTROL

The purpose of this study was to synthesize existing knowledge and conduct research to identify a range of available tools in use to support data visualization for sampling and analysis data related to a wide-area contamination incident. The work and conclusions presented as part of this study were empirical and observational – no scientific experiments were performed. Technical area leads evaluated the quality of the information collected by this effort (i.e., secondary data) and, based on their expert opinion, determined if the information should be documented within the literature review. Collected literature was evaluated using target search terms (Appendix A) and assessed according to the “Literature Review Scoring Criteria” as shown in Appendix B. All supporting documentation of the secondary data considered worthy for inclusion were cited. However, no experimental confirmation of secondary data (e.g., accuracy, precision, representativeness, completeness, and comparability) was conducted as part of this study.

## 3 LITERATURE REVIEW RESULTS

A literature review was conducted to provide an up-to-date picture of available open source or commercial off-the-shelf products that are available to support data visualization and statistical analysis needs, and applicable initiatives by regional and state partners. Figure 2 presents search terms that guided research, along with a “needs” statement. While not explicitly included, results pertaining to spatial statistical analysis or spatial data analysis were captured using the broader search terms, and products were subsequently screened for spatial capabilities.

Articles, market/vendor data, reports, guidance documents, case studies, after action reports and other pertinent information such as EPA enterprise-wide guidelines related to relevant platforms, applications, and programs were evaluated to identify tools that provide features to support data visualization and statistical analyses. Ongoing initiatives by EPA’s regional on-scene coordinators were also evaluated to leverage work underway that shares common objectives with this project. In addition to market/vendor information, other resources evaluated included:

- EPA’s Data Management/Analytics Platform (DMAP) Initiative,
- EPA’s GeoPlatform,
- EPA’s Enterprise Data Visualization Platform (Qlik),
- EPA’s Newly Formed Data Science Community of Practice,

- Statistical Analysis Tools/ Platforms
- Data Visualization Tools/ Platforms
- Geospatial Visualization
- Visualization Dashboards
- Data Analytics
- COVID-19 Data Visualization

**Needs Statement:**

*Identify the most efficient and compatible data visualization and statistical analysis tools for response to a contamination.*

**Figure 2. Search terms.**

- EPA’s Data Analytics and Visualization Community of Practice,
- Related EPA Response Data Management (e.g., MicroSAP Tool, Data Capture/Management), and
- EPA Regional Initiatives (e.g., Region 8’s Full Data Management Lifecycle and Emergency Response Team Portal).

Each information source was read, assessed, and documented based on several criteria. To standardize this process, a standardized Literature Assessment Form (Appendix B) was used to document the overall quality of information source. Upon completion of entry via the form, the project team literature reviewer’s evaluation was stored in a spreadsheet to document the assessment. The resulting spreadsheet was used to summarize key research findings.

Relevant sources were defined as those related to collecting, managing, synthesizing (i.e., reporting, analyzing, visualizing) field sampling data following a wide-area event. The following criteria were considered by the reviewer in the Literature Assessment Form (Appendix B): utility, clarity and completeness, uncertainty and variability, soundness, evaluation and review, focus, and verity. All the information sources reviewed were deemed at least moderately relevant by the reviewer based on the evaluation criteria that were summarized, and the relevant information is included in this report.

Twenty-one tools or technologies that have features relevant to EPA’s needs were identified. Table 1 presents an overview of the tools that were identified and several key attributes. The sections that follow provide a brief overview of each tool and describe any related ongoing initiatives/projects for which the tools are used.

**Table 1. Tool/Software Overview**

Tool/Software	Fee Structure	Ease of Use and Configurability <sup>1</sup>	Has Data Aggregation and Visualization Features	EPA Enterprise Offering	Data Easily Refreshed	Supports Geospatial Data?	Online Collaboration Capabilities?
Dash by Plotly	Subscription	Medium	Yes	No	Yes	Yes	Yes
Domo	Subscription	High	Yes	No	Yes	Yes	Yes
Esri Suite	Subscription	High	Yes	Yes	Yes	Yes	Yes
EXES	Free/Open Source	High	No	Yes	Yes	No	No
GeoDa	Free/Open Source	Medium	Yes	No	Yes	Yes	No
Google Data Studio	Free/Open Source	High	Yes	No	Yes	Yes	Yes
Highcharts	Subscription	High	Yes	Yes	Yes	Yes	Yes
IBM Cloud Pak for Data	Subscription	Low	Yes	No	Yes	Yes	Yes
IBM SPSS Modeler/ Statistics	Subscription	Low	Yes	No	Yes	Yes	No
Looker	Subscription	Medium	Yes	No	Yes	Yes	Yes
Microsoft Power BI	Subscription	High	Yes	No	Yes	Yes	Yes
MicroStrategy	Subscription	Low	Yes	No	Yes	Yes	Yes
Oracle Platforms	Subscription	Low	Yes	No	Yes	Yes	Yes
Panel	Free/Open Source	Medium	Yes	No	Yes	Yes	Yes
Qlik Sense	Subscription	High	Yes	Yes	Yes	Yes	Yes
R-Studio/R-Shiny	Free/Open Source	High	Yes	Yes	Yes	Yes	Yes
Statistical Analysis Software (SAS)	Subscription	Low	Yes	No	Yes	Yes	Yes
Sisense	Subscription	High	Yes	No	Yes	Yes	Yes
STATA	Subscription	Medium	Yes	No	Yes	No	No
Tableau	Subscription	High	Yes	No	Yes	Yes	Yes
Voilà	Free/Open Source	High	Yes	No	Yes	Yes	Yes

<sup>1</sup> High = Easy to download/access and customize; Medium = Requires some coordination for acquisition, but otherwise easy to implement; Low = Agreements Required

### 3.1 Dash by Plotly

Dash is an open-source framework for creating interactive web analytic applications based on code written in Python, R or Julia [3]. Plotly's plotting tools and Dash's dashboard features are combined for enhanced interactive user dashboard experiences. Dash apps can be quickly developed and linked to underlying data and made visible within a standard web browser [4].

While free for individual data

exploration, Plotly offers a subscription-based service to host, manage, serve, and scale dashboard applications via Dash Enterprise [3]. Dash is scalable and Dash apps can handle hundreds of simultaneous users, making it a highly collaborative tool [4].

<b>Fee Structure:</b>	Subscription
<b>Ease of Use and Configurability:</b>	Medium (requires some coordination for acquisition, but otherwise easy to implement)
<b>Has Data Aggregation and Visualization Features?</b>	Yes
<b>EPA Enterprise Offering?</b>	No
<b>Data Easily Refreshed?</b>	Yes
<b>Supports Geospatial Data?</b>	Yes
<b>Online Collaboration Capabilities?</b>	Yes

### 3.2 Domo

Domo is an enterprise action platform focused on enabling users to quickly build custom applications. The platform facilitates combining several data sets into one using the Domo Business Cloud. Additionally, users can create data visualizations from raw data using the Analyzer tool [5].

The Analyzer tool includes data visualization screens, including many different chart types, map options, and filters.

Domo emphasizes collaboration during the development process, and therefore has produced features such as annotations for commentary, governance tools for data access, and identical views across devices such as tablets or desktops. Additionally, Domo has a large catalog of layout templates available for developer use. Many of the features are in a template format; however, users can implement their own customizations [5].

<b>Fee Structure:</b>	Subscription
<b>Ease of Use and Configurability:</b>	High (easy to download/access and customize)
<b>Has Data Aggregation and Visualization Features?</b>	Yes
<b>EPA Enterprise Offering?</b>	No
<b>Data Easily Refreshed?</b>	Yes
<b>Supports Geospatial Data?</b>	Yes
<b>Online Collaboration Capabilities?</b>	Yes

As a part of Domo's data visualization offerings, users can create "Stories." Domo describes their "Stories" as pages of custom data visualizations that integrate logical display transitions, and "Stories" are differentiated from standard pages due to additional real-time, collaborative development features for privileged users. Story pages can be exported in portable document format (PDF) [5].

### 3.3 Esri Suite

Esri provides a large suite of tools to support geospatially driven data visualization. EPA provides an enterprise-level offering for Esri tools. Features include maps, operational dashboards, analytics, and enhanced communication components such as Story Maps to display geospatial data. The sections below describe several Esri product offerings that have relevance to this project and that were also cited by operational experts who were interviewed during the project.

<b>Fee Structure:</b>	Subscription
<b>Ease of Use and Configurability:</b>	High (easy to download/access and customize)
<b>Has Data Aggregation and Visualization Features?</b>	Yes
<b>EPA Enterprise Offering?</b>	Yes
<b>Data Easily Refreshed?</b>	Yes
<b>Supports Geospatial Data?</b>	Yes
<b>Online Collaboration Capabilities?</b>	Yes

#### 3.3.1 Esri ArcGIS Online (EPA's GeoPlatform)

EPA's offering of Esri's ArcGIS Online is commonly referred to as the EPA GeoPlatform. The EPA GeoPlatform is available to all EPA staff and is widely used by EPA to support geospatially referenced data collection, analysis, viewing, and reporting through maps and dashboards on both desktop and mobile products. ArcGIS Online is a cloud-based software-as-a-service that integrates geospatial tools and supports web mapping applications in a collaborative environment [6].

Geographic information system (GIS) data and related assets that are created can be stored, consumed (e.g., via a web or feature service), edited, and/or shared within the EPA GeoPlatform to facilitate collaboration. Other Esri data collection and data management products such as Survey 123, Collector<sup>1</sup>, Field Maps, and QuickCapture can be integrated with ease. These Esri tools are form-based applications used to collect, manage, and store data while incorporating geospatial abilities. Access to the EPA GeoPlatform requires account approval; however, users do not need an EPA Local Area Network (LAN ID) for access [6].

Esri's ArcGIS Online is widely used across EPA's emergency response community and is an integrated component to EPA's emergency response (ER) data management framework [6].

#### 3.3.2 Esri ArcGIS Dashboards

Esri ArcGIS Dashboards provide a framework to combine data visualization components and location-based analytics into a common view to aid communicating data-driven information. By combining location-based GIS data and services with interactive dashboards and visualizations, users can analyze trends, view the operational status of key elements, and easily refresh and

---

<sup>1</sup> ArcGIS Collector functionality is available in the new ArcGIS Field Maps



monitor data in real-time [7]. The dashboards are easily customized by including charts, maps, lists, tables, and gauges. There are many types of dashboard templates that are offered, including:

- Strategic dashboards to track success of organization goals,
- Tactical dashboards for historical data and trends,
- Operational dashboards for real-time data monitoring, and
- Informational dashboards for community outreach [7].

The ArcGIS Dashboards enhance situational awareness and are fully integrated with ArcGIS. Directly integrating real-time data in a geospatial context is especially important for emergency management and response. For example, the Emergency Operations Center in Raleigh, North Carolina, found it challenging to coordinate response among different groups (such as the fire department, first responders, etc.) since they did not have consolidated real-time data filtering. By working with ArcGIS Dashboards, they could successfully develop many customized dashboards to create an app for the display of real-time data and analytics [8]. More recently, Esri Dashboards has been shown to be a popular data analytics tool for the COVID-19 (coronavirus disease 2019) response and tracking. State governments and other academic agencies are using Esri to create maps and charts that can help identify COVID-19 hotspots and trends over time.

ArcGIS Dashboards was recently used in support of the Region 10 wildfire response. EPA emergency response teams used dashboards to track the location and decontamination status of impacted homes in the region. Dashboards were viewed as a valuable tool, and many stakeholders agreed they would continue to utilize this tool in future responses.

ArcGIS Dashboards has a large user community and many readily available resources for developers to utilize. For example, potential customers can access tutorials and lessons that show example use cases of how to monitor real-time emergencies or assess hurricane damage. There are also articles and documentation on how to best customize your data visualizations to make them most effective and user-friendly for informed decision-making [9].

### **3.3.3 Esri ArcGIS Insights**

ArcGIS Insights is a newer product offering of interest to several EPA operational experts interviewed as part of the project to meet standard data analysis needs. ArcGIS Insights provides analysis software that connects location analytics, data, and business intelligence (BI) into an easy-to-use workflow. ArcGIS Insights supports conducting advanced spatial, statistical, and predictive analyses that support decision-making within a geospatial perspective. Insights can connect directly to data through ArcGIS, relational databases, or spreadsheets, making it easy to utilize data generated from a variety of sources. Common nonspatial analysis methods are available, as well as support for popular scripting languages such as Python and R [10]. EPA has a limited pool of licenses that are available upon request to EPA's Office of Mission Support to further evaluate the applicability and utility of this product.

### 3.3.4 Esri ArcGIS Story Maps

Esri ArcGIS Story Maps allow users to customize maps and visualizations to support communication activities by directly combining data and contextual narratives to emphasize spatial relationships or correlations and to promote awareness of a situation. Users can easily customize ArcGIS Story Maps to make them more interactive by adding text, multimedia images or videos, or other enhancements to increase visual appeal to the audience. Templates are also included to guide users. Story Maps are commonly used by communications or public affairs offices to present data and information to a more general audience [11].

## 3.4 Electronic Data Exchange and Evaluation System (EXES)

EPA's Electronic Data Exchange and Evaluation System (EXES) was referenced by operational experts who were interviewed during this project. EXES provides data assessment and management for analytical laboratory data. This tool was created by EPA for efficient processing, assessment, and distribution of laboratory data for EPA's Contract Laboratory Program; however, its flexible design can support all laboratory data assessment needs across a variety of sectors [12].

<b>Fee Structure:</b>	Free/Open Source
<b>Ease of Use and Configurability:</b>	High (easy to download/access and customize)
<b>Has Data Aggregation and Visualization Features?</b>	No
<b>EPA Enterprise Offering?</b>	Yes
<b>Data Easily Refreshed?</b>	Yes
<b>Supports Geospatial Data?</b>	No
<b>Online Collaboration Capabilities?</b>	No

Users can customize evaluation parameters and specify their analytical methods in the EXES interface. In addition, flexible import and export formats are provided to meet user needs. For example, data imports may be formatted in XML, CSV, Excel, etc. Outputs, or Electronic Data Deliverables, can be generated in multiple formats, and are compatible with other EPA databases such as Scribe. Once an analytical method is specified, users can run automated tests in EXES for method quality objectives and quality assurance project plan (QAPP) requirements. EXES was designed to accommodate evolving processes and requirements users often face when analyzing laboratory data [12].

### 3.5 GeoDa

GeoDa is a software tool that incorporates spatial data analysis in a data visualization interface to present statistical results in a user-friendly way. GeoDa is compatible with vector spatial data in a variety of different formats (e.g., shapefiles, GeoJSON) and can support converting formats through its interface (such as .csv to a shapefile). Additionally, GeoDa supports multi-layers to enhance data visualizations, and users can choose the layers on which to conduct analyses [13].

<b>Fee Structure:</b>	Free/Open Source
<b>Ease of Use and Configurability:</b>	Medium (requires some coordination for acquisition, but otherwise easy to implement)
<b>Has Data Aggregation and Visualization Features?</b>	Yes
<b>EPA Enterprise Offering?</b>	No
<b>Data Easily Refreshed?</b>	Yes
<b>Supports Geospatial Data?</b>	Yes
<b>Online Collaboration Capabilities?</b>	No

Users can explore statistical results through tests and models, as well as analyze spatial and temporal patterns across linked views. GeoDa also supports identifying statistical relationships and spatial clusters, comparing averages, and finding relationships/trends over time and space. Data visualization screens include maps; statistical charts such as box and line; and legends. GeoDa is a free and open-source software and available for public use [13].

### 3.6 Google Data Studio

Google Data Studio is a visualization and reporting tool to aid users in decision-making and unlocking marketing insights. Data Studio allows users to create user-friendly custom reports and dashboards that can be easily shared. Data Studio includes a large library of reusable report and dashboard templates that integrate dynamic and interactive controls. These interactive controls include time periods, geography/maps, and other dimensions. Additionally, Data Studio has a number of visualization screens such as time series, bar charts, pie charts, tables, heat maps, geo maps, scorecards, scatter charts, bullet charts, and area charts [14].

<b>Fee Structure:</b>	Free/Open Source
<b>Ease of Use and Configurability:</b>	High (easy to download/access and customize)
<b>Has Data Aggregation and Visualization Features?</b>	Yes
<b>EPA Enterprise Offering?</b>	No
<b>Data Easily Refreshed?</b>	Yes
<b>Supports Geospatial Data?</b>	Yes
<b>Online Collaboration Capabilities?</b>	Yes

Data Studio is only compatible with Google platforms. Like other collaborative Google tools, built-in collaboration components allow individuals and teams to work on a dashboard or to report simultaneously with real-time updates and changes. Data Studio is a free and open-source platform for the public [14].

### 3.7 Highcharts

Highcharts is a Scalable Vector Graphics (SVG)-based JavaScript charting library that allows users to create interactive charts for web and mobile projects. Data can be imported in any form and can be easily changed or refreshed. Highcharts includes TypeScript which auto-completes code for easier developer use [15].

One of Highcharts' most highlighted features is its mobile and touch capabilities, including a seamless transition from desktop to mobile devices. Extensive editing features are also available for end-users including annotations and toolbars to provide visualization context. The visualization features include maps, charts, tables, color axis, bubbles, and tiles [15].

Highcharts has the capacity to handle large datasets with millions of data points. Users can also export or print visualizations in a variety of formats including JPG or PDF [15]. EPA has routinely used Highcharts to support data visualization projects and Highcharts is listed as an EPA-approved charting library [16]. Because Highcharts is a JavaScript library, additional developer skillsets are needed to support, create, and deploy data visualizations created.

<b>Fee Structure:</b>	Subscription
<b>Ease of Use and Configurability:</b>	High (easy to download/access and customize)
<b>Has Data Aggregation and Visualization Features?</b>	Yes
<b>EPA Enterprise Offering?</b>	Yes
<b>Data Easily Refreshed?</b>	Yes
<b>Supports Geospatial Data?</b>	Yes
<b>Online Collaboration Capabilities?</b>	Yes

### 3.8 IBM Cloud Pak for Data

IBM Cloud Pak for Data is a collection of services offered through IBM that supports managing data, statistics, visualization services, artificial intelligence (AI), and machine learning (ML). IBM Cloud Pak for Data is designed to handle large quantities of data and provides scalable, flexible tools to analyze and draw conclusions from the data [17].

IBM Cloud Pak for Data includes a data warehousing solution, Db2 Warehouse on Cloud, where data are stored to leverage Cloud Pak services. The database service itself contains a flexible user-interface and includes features such as widgets, tools, tables, lists, graphs, etc. in a centralized hub for task and data management [18].

<b>Fee Structure:</b>	Subscription
<b>Ease of Use and Configurability:</b>	Low (agreements required)
<b>Has Data Aggregation and Visualization Features?</b>	Yes
<b>EPA Enterprise Offering?</b>	No
<b>Data Easily Refreshed?</b>	Yes
<b>Supports Geospatial Data?</b>	Yes
<b>Online Collaboration Capabilities?</b>	Yes

The IBM Cognos Analytics product allows users to create their own dashboards and narrative driven stories, utilizing the capabilities of mass data storage offered by IBM's other services. Dashboards can contain interactive visualizations such as maps, ArcGIS, charts, graphs, and

tables. Dashboards are customizable as both a tool for viewing results and for collecting data. Data modeling features in the overall Cognos Analytics service can be processed and presented through the dashboard service, allowing users to obtain real-time feedback based on updates to statistical models and forecasts [19].

### 3.9 IBM SPSS Modeler/Statistics

The IBM SPSS suite of products contains SPSS Modeler – a software product for building analytical and predictive models – and SPSS Statistics – a tier-based software product that offers more advanced and technical statistical tools [20].

For SPSS Modeler, a task-based workflow is used to guide users, and the user interface is optimized for users with non-programming backgrounds. A

strength of this software is the collection of “nodes,” which act as customizable packages to do many common data analysis tasks. Often meticulous tasks such as data filtering, data exploration, data quality assurance and quality control, and model fitting are simplified by SPSS Modeler’s step-by-step approach to model-building. When working with data housed within the IBM-integrated databases, SPSS Modeler can process millions of records reliably, while ensuring compliance of data security protocols [20].

SPSS Modeler has the flexibility to work in tandem with open-source solutions such as R and Python to utilize their packages/libraries on SPSS-formatted data structures. The software supports geographical data types and can support creating non-interactive visualizations (in the forms of charts, graphs, and heatmaps) that can be exported out and displayed through other services (such as IBM Cognos Dashboard) [20].

SPSS Statistics is geared towards data analysts with a fundamental understanding of both statistics and statistical modeling, and for data-heavy projects that require advanced statistical methods. SPSS Statistics licensing is tier-based, with their Premium version offering statistical tools, including some that rely on ML algorithms, such as:

- In-depth Sampling Assessment and Testing,
- Forecasting,
- Geospatial Analytics,
- Spectral Analysis, and
- Temporal Causal Modeling.

The IBM SPSS suite of products is a desktop-only software, lacking integration with mobile devices. However, data that are generated from analyses conducted within SPSS Modeler or Statistics can be accessed and visualized through other services, either IBM-based (Cognos Analytics) or third-party (e.g., R, Python, Tableau) [21].

<b>Fee Structure:</b>	Subscription
<b>Ease of Use and Configurability:</b>	Low (agreements required)
<b>Has Data Aggregation and Visualization Features?</b>	Yes
<b>EPA Enterprise Offering?</b>	No
<b>Data Easily Refreshed?</b>	Yes
<b>Supports Geospatial Data?</b>	Yes
<b>Online Collaboration Capabilities?</b>	No

### 3.10 Looker

Looker is a business intelligence and data visualization product that utilizes Google Cloud to manage, analyze, and display a variety of data. Looker can be deployed on-premises or through other cloud providers. Looker has plans to provide a fully integrated mobile solution [22].

Looker's architecture takes advantage of modern cloud databases and allows the user to choose the most appropriate cloud provider for their needs.

Because Looker is browser-based and designed as a multi-cloud solution, it is simple and straightforward to change where Looker is deployed and its underlying source databases, while not affecting performance for the end-user [22].

Looker's standout feature is LookML, a SQL-based modeling language for defining business rules and data structures. Looker's approach separates the content of the data from the structure of the data, and LookML allows users to create generalized query structures that can then be used by non-technical users to easily access the data they need [23].

General functions are conducted through Looker's Action Hub. This hub provides for a customizable interface where users can conduct tasks to manage their data. Looker provides the ability to create customizable, and automatically updated, interactive dashboards and visualizations. A public library of charts, maps, widgets, and graphs is available to support developing advanced graphics [24]. Visualizations such as maps, GIS, charts, plots, graphs, and tables can be projected through the dashboard. Looker's analytics services allow for rudimentary data analysis (i.e., summary statistics and linear regression modeling), but lack the sophistication to conduct complex statistical analysis that can be achieved through more-dedicated statistical software packages. Looker dashboards can be deployed on-premises or through their cloud offerings [25].

<b>Fee Structure:</b>	Subscription
<b>Ease of Use and Configurability:</b>	Medium (requires some coordination for acquisition, but otherwise easy to implement)
<b>Has Data Aggregation and Visualization Features?</b>	Yes
<b>EPA Enterprise Offering?</b>	No
<b>Data Easily Refreshed?</b>	Yes
<b>Supports Geospatial Data?</b>	Yes
<b>Online Collaboration Capabilities?</b>	Yes

### 3.11 Microsoft Power BI

Microsoft Power BI is a platform that supports decision-making by optimizing business intelligence. The platform facilitates easy data connections and provides various data visualization features that include custom dashboards, data aggregation for storytelling, data analysis for decision-making, and the use of real-time data to increase collaboration [26].

<b>Fee Structure:</b>	Subscription
<b>Ease of Use and Configurability:</b>	High (easy to download/access and customize)
<b>Has Data Aggregation and Visualization Features?</b>	Yes
<b>EPA Enterprise Offering?</b>	No <sup>2</sup>
<b>Data Easily Refreshed?</b>	Yes
<b>Supports Geospatial Data?</b>	Yes
<b>Online Collaboration Capabilities?</b>	Yes

Microsoft Power BI can handle extremely large datasets and projects. Large datasets are shown using visualization screens such as dashboards and charts. Additionally, the platform can assist in data cleanup and modeling. While this platform can integrate data on-premises or on the cloud, it integrates more seamlessly with other Microsoft and Windows products. Several states have used Microsoft Power BI for COVID-19 dashboards to show charts that highlight trends and organize the data for state health decision-making [26]. Additionally, Microsoft Power BI works in conjunction with Esri ArcGIS maps to incorporate spatial data and analysis and enhance map visualization features [27].

### 3.12 MicroStrategy

MicroStrategy is a business intelligence and data visualization platform that can be deployed both on-premises or through the cloud. MicroStrategy offers mobile integration via their MicroStrategy Mobile application.

MicroStrategy promotes its “HyperIntelligence,” which is described as allowing for augmentation of any enterprise application to include relevant data accessible by MicroStrategy. HyperCards are an example form of HyperIntelligence, where MicroStrategy will parse through the data presented in an application, and tie in any relevant data that MicroStrategy can access (e.g., from other

<b>Fee Structure:</b>	Subscription
<b>Ease of Use and Configurability:</b>	Low (agreements required)
<b>Has Data Aggregation and Visualization Features?</b>	Yes
<b>EPA Enterprise Offering?</b>	No
<b>Data Easily Refreshed?</b>	Yes
<b>Data Easily Refreshed?</b>	Yes
<b>Supports Geospatial Data?</b>	Yes
<b>Online Collaboration Capabilities?</b>	Yes

---

<sup>2</sup> EPA’s Office of Mission Support confirmed that limited licenses are available and can be obtained by submitting a request through the Working Capital Fund to access through the Agency’s Office 365 subscription; however, staff must present clear justification for the need since Qlik is the Agency’s preferred/funded platform.



applications) into a small dashboard within the application. HyperCards are customizable, and free online tutorials to support their use are available [28].

MicroStrategy provides built-in data connectors to many popular database providers. MicroStrategy also provides integration for R, Microsoft Office, Qlik, Tableau, and Jupyter Notebooks where data are passed through an integrated application supporting an external analysis and the platform sends those results to the MicroStrategy environment [29].

Users can create and deploy self-service dashboards that have real-time data refreshing and updating. Dashboards can project spatial data through GIS and offer standard statistical analysis tools. The catalogue of statistical analysis tools offered through MicroStrategy is comparable to other popular BI platforms, but less robust and less customizable compared to software specifically designed for more powerful statistical analyses [30].

### 3.13 Oracle Platforms

Oracle Analytics offers both a desktop product, known as Oracle Analytics Desktop [31] and a cloud-based platform, known as Oracle Analytics Cloud [32]. Oracle Analytics Desktop serves as a standalone application that is used to analyze and visualize data, whereas Oracle Analytics Cloud is a platform that contains a number of services for data management, analysis, and visualization.

<b>Fee Structure:</b>	Subscription
<b>Ease of Use and Configurability:</b>	Low (agreements required)
<b>Has Data Aggregation and Visualization Features?</b>	Yes
<b>EPA Enterprise Offering?</b>	No
<b>Data Easily Refreshed?</b>	Yes
<b>Supports Geospatial Data?</b>	Yes
<b>Online Collaboration Capabilities?</b>	Yes

The analytics and visualization services offered through Oracle Analytics Cloud are more advanced than those offered in Oracle Analytics Desktop, utilize machine learning and AI, and additionally integrate cloud-based deployment. The main benefit to Oracle Analytics Desktop is that it is explicitly targeted as an analytics software, rather than a broader data management platform [32].

The Oracle Analytics Cloud Enterprise Edition contains the Oracle Business Intelligence Cloud Service, where users can manage all data and tasks through a centralized hub. Administrators can manage user permissions so that only verified users may access certain functions within the Business Intelligence Cloud, such as SQL tasks, modeling, dashboard creation and access, and more [33].

In addition to the Business Intelligence Cloud Service, Oracle Analytics Cloud also includes the Oracle Data Visualization Cloud service and a limited number of licenses for Oracle Data Visualization Desktop. Oracle Data Visualization Cloud Service is a web-based tool specifically designed for visually exploring data without needing advanced technical skills. The user interface supports drag-and-drop features for tables, datasets, graphs, etc. to display in the



viewing window. The desktop application allows for the same visualization services and can also work offline [34].

### 3.14 Panel

Panel is an open-source python library that supports creating web applications and dashboards. Panel supports almost all plotting libraries and supports Python and static HTML/JavaScript applications without having to connect domain-specific code to specific web tools [35]. Panel provides users with options to host their own dashboards. The website provides documentation in the form of tutorials and blogs that can be referenced to support managing and maintaining dashboards in a preferred environment; however, the implementation and underlying architecture fall to the user to establish and maintain. A noted desirable feature of Panel is its compatibility with HoloViews/GeoViews feature, which supports gridded and geospatial data [36].

<b>Fee Structure:</b>	Free/Open Source
<b>Ease of Use and Configurability:</b>	Medium (requires some coordination for acquisition, but otherwise easy to implement)
<b>Has Data Aggregation and Visualization Features?</b>	Yes
<b>EPA Enterprise Offering?</b>	No
<b>Data Easily Refreshed?</b>	Yes
<b>Supports Geospatial Data?</b>	Yes
<b>Online Collaboration Capabilities?</b>	Yes

### 3.15 Qlik Sense

Qlik Sense is a data analytics software used to create custom dashboards and visualizations by automatically connecting data relationships across several data sources. Qlik aims to “close the gaps between data, insights, and actions and automate the process from raw to complete analytics” [37]. Qlik Sense can also support integrating data from other popular data analytics platforms such as Tableau and Power BI [38].

<b>Fee Structure:</b>	Subscription
<b>Ease of Use and Configurability:</b>	High (easy to download/access and customize)
<b>Has Data Aggregation and Visualization Features?</b>	Yes
<b>EPA Enterprise Offering?</b>	Yes
<b>Data Easily Refreshed?</b>	Yes
<b>Supports Geospatial Data?</b>	Yes
<b>Online Collaboration Capabilities?</b>	Yes

Users can explore interactive dashboards that inform data-driven decisions. Developers can create dashboards and functions using natural language processing. From there, charts can be auto generated and customized using open application programming interfaces (APIs). Predictive calculations may also be processed through several data tools such as R and Python. Qlik has mobile capabilities on iOS and Android, as well as offline analysis capabilities using the Enterprise Mobility Management platforms. Qlik offers capabilities to integrate geo-analytics with location-based mapping and geodata lookup services [39].

EPA offers an enterprise-wide Qlik solution that was successfully used for multiple EPA projects requiring visualization dashboards. For example, these projects include EPA’s efforts to manage site visits for the Northern California wildfires and EPA’s assessment of Superfund sites and hazardous debris from Hurricane Maria (2017) damage in Puerto Rico [40]. Qlik requires developers to support the creation of dashboards, and developers must have proper credentials to access to EPA’s internal development environment. EPA has an active user community and personnel that are available to support program office dashboard development needs. EPA’s Qlik instance is also compatible with EPA’s Scribe database, and data can be easily refreshed within these systems.

### 3.16 RStudio/Shiny

RStudio is a free software package that utilizes the power of the open-source R programming language to conduct robust statistical analyses [41]. As an open-source technology, RStudio allows for the community to easily share solutions and user-created packages between each other [41].

<b>Fee Structure:</b>	Free/Open Source
<b>Ease of Use and Configurability:</b>	High (easy to download/access and customize)
<b>Has Data Aggregation and Visualization Features?</b>	Yes
<b>EPA Enterprise Offering?</b>	Yes
<b>Data Easily Refreshed?</b>	Yes
<b>Supports Geospatial Data?</b>	Yes
<b>Online Collaboration Capabilities?</b>	Yes

RStudio is capable of handling large quantities of data; however, data management issues are common for users. Other data analysis software, especially when combined within a larger data management solution, may assist in resource allocation to alleviate processing loads. RStudio requires that users have a fundamental understanding of data storage and R-data types to effectively manage large quantities of data without running into processing issues.

RStudio can handle data from other popular data management sources, such as PowerBI, Tableau, Qlik, and IBM. It has packages that support importing shapefiles and building maps to support visualization and advanced spatial analysis and statistics. RStudio also supports creating and exporting data in a variety of formats to use within other applications and platforms [41]. Additionally, RStudio Connect facilitates more effective data sharing where content creators can easily publish and share products including Shiny apps, dashboards, reports, and Jupyter notebooks. RStudio Connect offers the ability to automatically update and distribute changes to published products through a single portal, as well as control security requirements for users [42].

RStudio, as a company, developed Shiny as a tool for building interactive dashboards and applications that are created within the RStudio software. Shiny is a set of packages used within RStudio to create interactive dashboards that contain charts, tables, graphs, analysis results and narratives that can be hosted either locally or through RStudio services. Some capabilities of the dashboards include integration of statistical analysis, ArcGIS, basic and advanced statistical charts, forecasting, images, tables, significance testing, and heat maps [43].

Shiny is an open-source R package for building interactive web applications in R but leveraging HTML, CSS, and JavaScript under the hood. Designing dashboards is less intuitive compared to other solutions that offer drag-and-drop functionality because Shiny dashboards are entirely based on underlying R code. Dashboards created through Shiny can be deployed either on-premises or online at shinyapps.io. These dashboards can be accessed via mobile devices [43].

### 3.17 Statistical Analysis Software (SAS)

Statistical Analysis Software (SAS) Enterprise BI Server is a software offering that comprises a suite of products for all business intelligence needs, including SAS Visual Analytics, which supports the development of data relationships in dynamic reports and dashboards. SAS Visual Analytics also includes geographical features so users can develop maps, as well as tables and charts to interact with the data both on and offline. Add-ons are available to support more in-depth statistical packages, and a machine-learning/data mining package is offered to extend the capabilities of the SAS Visual Analytics product [44].

<b>Fee Structure:</b>	Subscription
<b>Ease of Use and Configurability:</b>	Low (agreements required)
<b>Has Data Aggregation and Visualization Features?</b>	Yes
<b>EPA Enterprise Offering?</b>	No
<b>Data Easily Refreshed?</b>	Yes
<b>Supports Geospatial Data?</b>	Yes
<b>Online Collaboration Capabilities?</b>	Yes

Data are managed through the SAS Information Delivery Portal, which provides users a straightforward, customizable interface to organize a desktop for viewing relevant data, alerts, charts, maps, etc. The SAS Information Delivery Portal serves as the overall hub for connecting data managed through SAS. Reports created through SAS reports, shareable processing scripts designed by SAS Stored Processes, and data maps built within SAS Information Maps can all be effectively managed within SAS Information Delivery Portal. The hub is made up of pages, and “portlets,” which are small windows that pull data and visualizations directly from other SAS products and services [45].

SAS BI Dashboard is the dashboard and visualization service offered by SAS. Dashboards can be presented either as a standalone dashboard (which can be deployed on-premises), or as a portlet within the SAS Information Delivery Portal. Dashboards can display geographical data, and interactive charts, graphs, and tables. Dashboards can integrate the statistical tools found in the SAS Visual Analytics service to include modeling techniques (such as forecasting) into the dashboards. Dashboards can be viewed via mobile devices [46].

### 3.18 Sisense

Sisense is a business intelligence platform that supports data management, analysis, and visualization. Data visualization features include charts, graphs, and map/GIS integration [46].

Pre-built data connectors allow users to connect and pull data from several different platforms to centrally manage data, permitting data changes or refreshes to be easily completed by the developer. Sisense emphasizes that no previous experience with coding or business intelligence platforms is required to understand the platform's functionalities. A large suite of APIs, software development kits and developer tools are available, and the platform provides flexible deployment options, multiple levels of security and governance roles, and flexible data engines to manage complex data. There are additional packages available to extend capabilities including Sisense for Cloud Driven Teams, Sisense for BI and Analytics Teams, and Sisense for Product Teams [46].

<b>Fee Structure:</b>	Subscription
<b>Ease of Use and Configurability:</b>	High (easy to download/access and customize)
<b>Has Data Aggregation and Visualization Features?</b>	Yes
<b>EPA Enterprise Offering?</b>	No
<b>Data Easily Refreshed?</b>	Yes
<b>Supports Geospatial Data?</b>	Yes
<b>Online Collaboration Capabilities?</b>	Yes

### 3.19 STATA

STATA is a desktop-only statistical analysis software that contains a wide range of built-in tools to conduct statistical analyses and generate statistical models. STATA offers many types of advanced and specialized data analysis tools, including survival analysis, Bayesian analysis, extended regression models, and generalized linear models. As of STATA's most recent update, the software fully integrates Python. Users can use any Python package directly within the STATA interface [48]. In addition, STATA supports the import of both SAS and SPSS data types, along with other common data types (e.g., .csv, .xlsx) [49].

<b>Fee Structure:</b>	Subscription
<b>Ease of Use and Configurability:</b>	Medium (requires some coordination for acquisition, but otherwise easy to implement)
<b>Has Data Aggregation and Visualization Features?</b>	Yes
<b>EPA Enterprise Offering?</b>	No
<b>Data Easily Refreshed?</b>	Yes
<b>Supports Geospatial Data?</b>	No
<b>Online Collaboration Capabilities?</b>	No

STATA, however, is only compatible with nominal and numerical data (not spatial data). STATA supports creating visualizations in the form of basic and advanced charts, graphs, and plots using STATA's built-in reporting feature. Additional data transformations would likely be required to integrate data generated in STATA into external dashboards hosting via other platforms [50].

### 3.20 Tableau

Tableau is a data analytics platform that supports simplifying data into a user-friendly and understandable format. Tableau's features include integration with other applications, collaborative dashboard sharing, analytics, user role and permissions management, support for managing datasets and data connections, and flexible deployment formats [51].

<b>Fee Structure:</b>	Subscription
<b>Ease of Use and Configurability:</b>	High (easy to download/access and customize)
<b>Has Data Aggregation and Visualization Features?</b>	Yes
<b>EPA Enterprise Offering?</b>	No
<b>Data Easily Refreshed?</b>	Yes
<b>Supports Geospatial Data?</b>	Yes
<b>Online Collaboration Capabilities?</b>	Yes

Data can be embedded into any application, including mobile devices. One of Tableau's most popular features is turning data into visualization dashboards that help support data-driven storytelling. Additionally, these dashboards can be shared among users for easy collaboration, and users can be separated into different governing roles when working with a mix of stakeholders. Data can be connected to Tableau dashboards through the cloud or an on-premise database, and deployment is compatible and flexible with an existing data infrastructure [51].

Tableau's visualization screens include charts, tables, graphs, maps, infographics, and dashboards [52]. The visualization screens are highly customizable, can be easily downloaded, and are widely used by a variety of sectors, including federal, state, academic, and commercial entities.

More recently, Tableau is shown to be a popular data analytics tool for the COVID-19 response and tracking. State governments and other academic agencies are using Tableau to create maps and charts that can help identify COVID-19 hotspots and trends over time [53]. Tableau's platform facilitates easily aggregating and refreshing data from state and local governments and from health departments. Because Tableau has such a large user community, there are many forums, user stories, and reference materials to support researching and implementing desired features for creators that do not have extensive technology backgrounds.

### 3.21 Voilà

Voilà is a free service, open-source solution that allows users to create dashboards using the underlying programming architecture of Jupyter Notebooks. Dashboards are customizable, and a series of templates are available to support data presentation options. Voilà can be used as a standalone application, or as a Jupyter server extension [54].

Voilà creates interactive dashboards by converting the input notebook and resulting outputs to HTML. The newly converted HTML is then hosted either via the Jupyter server or as a Tornado application (web framework for Python). The widgets on the page are interactive, as they have access to the underlying Jupyter processes. A notebook directory contains the list of Jupyter Notebooks that can be rendered using Voilà. These notebooks include the widgets and controls that allow for reader interaction with published dashboards [55].

<b>Fee Structure:</b>	Free/Open Source
<b>Ease of Use and Configurability:</b>	High (easy to download/access and customize)
<b>Has Data Aggregation and Visualization Features?</b>	Yes
<b>EPA Enterprise Offering?</b>	No
<b>Data Easily Refreshed?</b>	Yes
<b>Supports Geospatial Data?</b>	Yes
<b>Online Collaboration Capabilities?</b>	Yes

Dashboards created using Voilà can be accessed via mobile devices and can be built to be interactive. Creators can place scalable GIS maps, interactive charts and graphs, or 3D models into these dashboards for users to explore. Additionally, there is an active community of users to troubleshoot issues, and public galleries to explore other user-created visualizations and templates [55].

## 4 OPERATIONAL EXPERT FEEDBACK

To supplement information obtained from the literature review and market research, operational expert feedback was solicited from the response and research community to understand what visualization tools are currently being used for presenting and analyzing data collected during a contamination incident response, as well as to identify efficiencies, lessons learned, and knowledge gaps based on their experience. The project team held a series of meetings over several months with operational experts across the response community. Understanding the user experience and what is important to a user is critical to ensuring the right solutions are selected. Conducting user interviews is a method that can be used to gain a more thorough understanding of tasks users must accomplish, an appreciation of any challenges or constraints, and their top priorities.

Eleven operational experts were identified, including EPA on-scene coordinators, regional leads, and other operational experts within EPA, as well as experts outside of EPA (e.g., FBI). Individuals were selected based on their relevant expertise in this research area. Qualifications considered included:

- Experience collecting and/or analyzing microbial sampling data from an incident response, and
- Experience with data needs related to measurement, detection limits, characterization, exposure determinations, and site clearance.

Discussions were centered on, but not limited to, the following topics:

- Types of data visualization/statistical analysis tools/platforms most frequently used or accessed,
- Specific features/capabilities that make them most useful,
- The frequency by which data visualization platforms/tools are used,
- The role data visualization plays in executing job duties,
- Features/capabilities that are the highest priority,
- Any challenges and/or constraints that impact how job tasks are conducted, and
- Identification of the top data visualization priorities.

Four important capabilities were consistently shared by operational experts who were interviewed:

1. Tool and data access and flexibility are paramount,
2. Geospatial context is critical,
3. Data should be centrally and separately managed from visualization and analysis tools to enable EPA and its partners to easily adapt to advances in technology, and
4. EPA should prioritize work to establish a process for creating data management plans with other agency partners to standardize and communicate data flows, identify what kind of data are generated, what formats are being used, where data are sourced, and how data can be accessed.

Below is a summary of the key takeaways that emerged from the interviews. The considerations below will inform future strategies to enhance EPA's capabilities to streamline and improve data visualization tools to meet the needs of EPA responders and other agency partners.

### ***Tool Access***

- EPA's Office of Mission Support (OMS) confirmed that EPA can use the full suite of Esri products. Several products may not have Agency-wide licensing available; therefore, obtaining a license will likely require a specific request.
- OMS suggested creating several use cases to explore and validate candidate solutions, including testing processing limitations and data transfer/storage constraints, to ensure a robust solution is selected.
- Solutions to improve current limitations related to data storage and access should be evaluated.
  - Stakeholders noted that the amount and granularity of data that will be needed for cleanup is often underestimated.
- Emergency response (ER) activities involve time-sensitive activities, requiring expedited capabilities related to:

- Transferring data
- Sharing data
- Analyzing data
- ER Team data are managed in a separate cloud environment that is not part of the standard OMS infrastructure offerings.
  - Specific procedures/processes (or widespread awareness) for rapidly scaling storage and processing capacity do not appear to exist.
- Stakeholders suggested that maintaining emergency response data infrastructure separate from EPA's core IT services is important, including tools to support data collection, data management and visualization.
- Stakeholders emphasized the importance of the ER Team maintaining control over, and centralizing emergency response data. Doing so will facilitate expedited access to data, streamline data services and ensure EPA data are only stored on EPA servers.
- Timely access and flexibility to configure data infrastructure to quickly adapt to response needs is crucial.
- EPA Superfund Technical Assessment and Response Team contractors, most of whom cannot access assets inside EPA's firewall, provide a significant level of data management and GIS support for response activities.
- A stakeholder noted that it would be beneficial to explore other data storage options and to consider an environment with an architecture that can be accessed both behind and in front of the firewall.
- There was consensus that where data are stored is less important so long as data connections can be easily established.

### ***Response Stakeholder Collaboration***

- Because there is a broad range of stakeholders that require access to data, establishing proper access and efficiently distributing data can be challenging.
  - Different levels of information are shared with different stakeholders, including EPA credentialed staff, stakeholders, and the public.
  - Managing data access decisions for multiple stakeholders introduces additional complexities.
- Visually representing large datasets requires adequate infrastructure for storage, processing, distribution, and analysis among ALL stakeholders (e.g., on-scene coordinators, states, other partners).
- Distributed collaboration among partners and other agencies is a significant challenge.
- Integrating different platforms would be challenging in circumstances where a common operating picture is needed, but agencies are using different tools.
- Stakeholders noted that EPA efforts to establish a process for creating a data management plan with other agency partners could be beneficial where data flows identify what kind of data are generated, what formats are used, where data are sourced, and how to access the data.
- Ultimately, data access and compatibility are key drivers of successful collaboration.
  - Understanding various agency objectives and the impact on EPA's mission can support developing an intentional process, data flow, and key business rules.



- Stakeholders emphasized the importance of identifying specific objectives to help identify and focus on the data that are truly important.
- Variability in the use of different identifiers and flags among stakeholders often requires cross-referencing and resolution prior to performing required data translation/transformations when leveraging external data.
- Quality assurance routines can be established to elevate questionable records that can be communicated during daily check-in calls where data updates and issues are discussed and resolved.

### ***Tools***

- Stakeholders emphasized that using Qlik to support analytics and visualization is limited by platform access constraints (i.e., requires a local area network (LAN) ID for anything other than public facing, published views).
- Overall, among those interviewed, there appeared to be less enthusiasm toward Qlik adoption in the on-scene coordinator community.
- ER teams are beginning to explore Esri's new ArcGIS Insights software to support geospatially oriented analytics and visualizations.
- EPA aims to leverage enterprise-wide solutions, and most stakeholders agreed that the enterprise solution that best meets the needs identified is Esri's suite of products.
- Stakeholders expressed a preference to standardize tools employed by the emergency response community based on a single suite of compatible tools (e.g., Esri-based products). Standardization could leverage existing familiarity, minimize resistance to adoption (i.e., perceptions of ever-changing tools), focus training needs, build skills capacity, and minimize the number of different accounts that users need to track.
- There is wide adoption of Esri products and broad use of EPA's GeoPlatform by the EPA regions and by ER teams.
- Stakeholders indicated that Esri's ArcGIS operational dashboards provide flexibility to create different views from common source data, as well as control access to different views. Other products referenced include ArcGIS Insights and ArcGIS QuickCapture<sup>3</sup>.
- While the Esri suite is widely seen as fulfilling ER needs, stakeholders stressed the importance of also independently maintaining source data from ArcGIS Online to facilitate adapting to and leveraging other tools.
  - Currently, some ER teams store data in the cloud in a MS MySQL database.
  - Data can be sourced to other tools (e.g., Highcharts, Qlik).
  - It is important to maintain control over the data.
- Operational plans should be driven by data and results.
  - The ability to drill-down, apply filters (e.g., based on a status code) and adjust maps would be helpful.
- Stakeholders discussed the use of EPA's Scribe tool in response efforts.

---

<sup>3</sup> ArcGIS QuickCapture, <https://www.esri.com/en-us/arcgis/products/arcgis-quickcapture/overview>, last accessed: March 5, 2021. QuickCapture is a tool to support expediting field data collection simple observations and minimal clicks through an interface that is optimizing for use in the field.

- Scribe stores sampling, observational, and monitoring field data, and outputs include labels for collected samples, chain of custody generation and analytical lab result data reports.
- Scribe supports exporting electronic data for use with other tools such as ArcGIS Survey123 and spreadsheets to facilitate further analyzing data and incorporating data into communication products.
- Scribe data can be imported into Esri dashboards via web services.
- At present, of those interviewed, there was limited staff that have the knowledge to support Scribe data management, transformation, storage, etc.
- Some stakeholders indicated that the Scribe process is not seamless and that data transformations are often manually completed.
- Suggestions were made to document/illustrate the current state and workflow, and the anticipated future state and workflow. Having additional training and demonstrations provided would also be helpful to minimize the learning curve.
- A stakeholder mentioned a web-based data assessment and management tool, the Electronic Data Exchange and Evaluation System (EXES), that was developed to efficiently evaluate analytical laboratory data.
  - EXES output can be imported to Scribe.
- Microsoft (MS) Teams channels and wikis were used to support “virtual collaboration.”

### ***Other General Observations***

- Contingencies to mitigate the impact of communication outages (i.e., internet outages) should be considered.
- Offline capabilities with subsequent synchronization to a centralized platform is an important feature. Indication that data are successfully transferred to avoid and ensure that no data are lost would be ideal.
- Stakeholders agreed that storing large amounts of data anticipated from a wide-area response is a formidable challenge.
- Diminished laboratory capacity for biological sampling is a concern.
- EPA’s Emergency Management Information Technology Workgroup is working to identify solutions to address data management challenges.
  - Concerted efforts should be made to elevate and empower the data management role.
  - Recognize the importance of data in all response activities and the need for intentional coordination activities from the beginning.
- Stakeholders emphasized the important role of data management and the data team, as well as having a flexible environment to accommodate changes in mission assignments and requests for additional metrics to assess data management tool capabilities.
- It is important to have the ability to easily access data that enhances situational awareness and understand, in advance, what other externally curated data are needed.

## 5 FINAL RECOMMENDATION

EPA identified a clear need to identify and better understand the universe of currently available software platforms and tools in use throughout EPA, the federal government, and commercial and/or academic settings to support statistical analyses and data visualization. This project had three primary objectives to address this need:

1. **Conduct a literature review/market research** to identify and describe data visualization and statistical analysis platforms (i.e., tools, applications, and programs) currently in use throughout EPA, the federal government, and commercial and/or academic settings,
2. **Solicit operational feedback** from stakeholders within the EPA response community regarding visualization tools currently being used for presenting and analyzing data collected during a response to a contamination incident, and
3. **Develop a final summary and recommendation report** describing recommendations for adopting and integrating improved statistical analysis and data visualization tools to enhance EPA's capabilities in support of managing data generated throughout all phases of the incident response to inform decision-making during a response to a contamination incident.

Through this project, EPA hoped to gain a better understanding of the technology options that exist, which options are currently in use, and the emergency response community's perspective on important considerations related to tool access, flexible customization, geospatial context, and requirements that are unique to response events. Centrally storing and separately managing data independent of visualization and analysis tools was an important data management consideration that was emphasized by operational experts. Doing so affords maximum control over data that can be made available to any number of platforms in the future to enable EPA and its partners to easily adapt to advances in technology.

As can be seen from the screening of available commercial products, many products contain very similar offerings. Therefore, in addition to specific desirable features, other practical considerations become a factor in selection preferences, including licensing fees, enterprise-wide availability, ease of user access and configurability<sup>4</sup>, and workforce skillsets.

Supported by both research conducted and operational experts' input, the project team recommends that EPA and DHS/USCG should focus on a suite of products that:

- Are compatible within the overall data management workflow as shown in Figure 3 below,
- Are easily accessible by key stakeholders,
- Can be easily configured to share and collaborate among stakeholders,

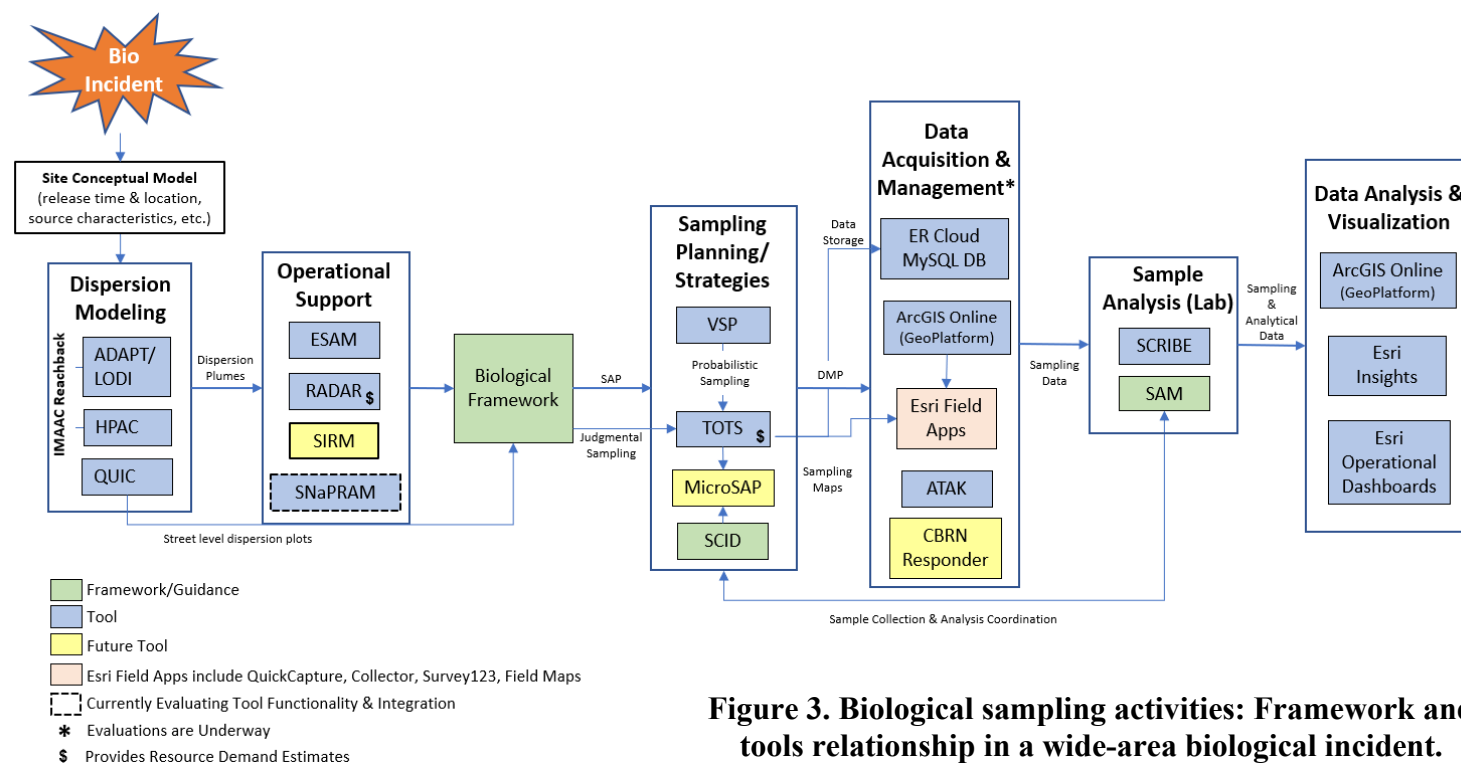
---

<sup>4</sup> Ease of user access and configurability in the context of this project relate acquiring software within the organization and the ability to configure software to meet the needs of a specific response event.

- Can be readily acquired and broadly licensed for use,
- Are compatible with external databases, and
- Are generally accepted by potential users to aid in adoption and regular use.

The project team recommends adopting and integrating Esri's suite of products to support statistical analysis and data visualization needs to support the upcoming AnCOR field exercise and further enhance EPA's capabilities to better manage all data generated over the course of a response to inform decision-making during a response to a contamination incident. The project team recommends leveraging the work of EPA's regional response teams and adopt a similar workflow that makes use of Esri's ArcGIS Online (EPA GeoPlatform) web maps and Esri Operational Dashboards to support data visualization. Should EPA need to conduct additional data analyses, the project team recommends exploring and exercising Esri Insights for integrated analysis needs. These products can be tightly integrated with other software packages and data libraries that leverage Python [56] and R [57] to support data-driven analyses. The Esri suite of tools also supports many add-ons and packages to enhance and support additional capabilities as data visualization needs change or evolve. This suite of tools has the most features that meet the largest number of needs, is familiar to and accepted by target stakeholders, and is generally viewed as easy to customize and tailor to meet the specific needs of the operation.

Throughout research conducted and interviews held, Esri's suite of products was consistently cited as offering many of the required capabilities and meeting the expressed needs of the response community. The Esri product suite is widely adopted among the response community and has been used by the USCG in support of various missions including search and rescue, pollution response, and to natural disasters. Several Esri Field Apps are routinely used to support field data collection and would seamlessly integrate to support the next phase of the overall data workflow as shown in Figure 3 [58].



**Figure 3. Biological sampling activities: Framework and tools relationship in a wide-area biological incident.**

#### ACRONYMS

ADAPT - Atmospheric Data Assimilation and Parameterization Tool  
 ATA - Android Team Awareness Kit  
 CBRN - Chemical, Biological, Radiological, or Nuclear  
 CIT - Critical Infrastructure Tool  
 Decon ST - Decontamination Strategy and Technology Selection Tool  
 ESAM - Environmental Sampling and Analytical Methods  
 HPAC - Hazard Prediction and Assessment Capability  
 IMAAC - Interagency Modeling and Atmospheric Assessment Center  
 I-WASTE - Incident Waste Decision Support Tool  
 LODI - Lagrangian Operational Dispersion Integrator  
 MicroSAP - Microbiological Sampling and Analysis Plan  
 QUIC - Quick Urban and Industrial Complex

RADAR - Remediation Data Repository  
 RAP - Remedial Action Plan  
 SAM - Selected Analytical Methods  
 SAP - Sampling and Analysis Plan  
 SCID - Sample Collection Information Document  
 SCRIBE - Specification Change Review, Implementation, and Baseline Evaluation Board  
 SNaPRAM - Sampler Network Performance for Resuspended Aerosols Model  
 TOTS - Trade-off Tool for Sampling  
 VSP - Visual Sample Plan  
 WEST - Waste Estimation Support Tool  
 WMPT - Waste Management Planning Tool

Regarding data management strategies for the ANCOR field exercise, the project team recommends working with EPA's National Response Team to acquire shared space in the EPA Emergency Response (ER) Cloud to store and manage field study data. According to the *Data Handbook for On-Scene Coordinators*, the ER Cloud is regularly used to house data in a Microsoft SQL Server relational database for projects that may require more advanced workflows or data analysis [6]. Subsets of applicable data can be shared with other Esri products (e.g., an Operational Dashboard or Map) to support the field exercise.

In addition to recommending a data visualization and analysis tool to exercise during the AnCOR study, the EPA and DHS/USCG team should also consider other important insights that were conveyed by operational experts, including:

- ER activities involve time-sensitive activities, requiring expedited capabilities related to:
  - Transferring data,
  - Sharing data, and
  - Analyzing data.
- Maintaining emergency response data infrastructure separate from EPA's core IT services is important, including tools to support data collection, data management and visualization.
- Operational plans should be driven by data and results.
- Specific decision objectives should be identified to inform and target what data are needed.
- EPA should establish a process for creating a data management plan with other federal agency partners to document data flows, identify what kind of data are generated, specify data formats, identify where data are sourced, and how to access the data.
- Granting access to platforms for key stakeholders should be easily accomplished.
- EPA should emphasize and elevate the important role of data management and the data team.

The output obtained and presented as a result of this project will inform the selection and implementation of a solution to address data visualization and analysis needs for the AnCOR field exercise. Lessons learned from the future exercise, as well as the important insights shared from the response community during this project, will further improve coordination and preparedness among EPA staff and DHS/USCG staff.

## 6 REFERENCES

*All hyperlinks last accessed 12 July 2021.*

1. *Analysis for Coastal Operational Resiliency*, 2021. United States Environmental Protection Agency. <https://www.epa.gov/emergency-response-research/analysis-coastal-operational-resiliency>
2. U.S. Environmental Protection Agency. 2021 (in press). *Data Management for Wide-area Responses: Literature Review and Operational Expert Feedback*. U.S. Environmental Protection Agency, Washington, DC, EPA/600/R-21/095, 2021.
3. *Dash*, 2021. Plotly. [plotly.com/dash](https://plotly.com/dash)
4. *Introduction to Dash*, 2021. Plotly. <https://dash.plotly.com/introduction>
5. *The Domo Business Cloud*, 2021. Domo. <https://www.domo.com/platform#visualize-and-analyzeGeoPlatform>
6. *Data Handbook for On-Scene Coordinators*. United States Environmental Protection Agency. 2020.
7. *ArcGIS Dashboards*, 2021. Esri. <https://www.esri.com/en-us/arcgis/products/arcgis-dashboards/overview>
8. *Raleigh Enhances Situational Awareness for Emergency Response with Operational Dashboards*, 2021. Esri. <https://www.esri.com/en-us/landing-page/product/2020/raleigh-nc-case-study>
9. *Try ArcGIS Dashboards*, 2021. Esri. <https://learn.arcgis.com/en/paths/try-operations-dashboard-for-arcgis/>
10. *ArcGIS Insights*, 2021. Esri. <https://www.esri.com/en-us/arcgis/products/arcgis-insights/overview>
11. *ArcGIS Story Maps*, 2021. Esri. <https://www.esri.com/en-us/arcgis/products/arcgis-storymaps/overview>
12. *EXES Fact Sheet*, 2018. United States Environmental Protection Agency. [https://www.epa.gov/sites/production/files/2018-09/documents/exes\\_fact\\_sheet\\_v6.pdf](https://www.epa.gov/sites/production/files/2018-09/documents/exes_fact_sheet_v6.pdf)
13. *Introducing GeoDa 1.18*, 2021. GeoDa. <https://geodacenter.github.io/>
14. *Google Data Studio*, 2018. Google. [https://services.google.com/fh/files/misc/data\\_studio\\_product\\_overview.pdf](https://services.google.com/fh/files/misc/data_studio_product_overview.pdf)
15. *Highcharts JavaScript Charting Library*, 2021. Highcharts. <https://www.highcharts.com/blog/products/highcharts/>
16. U.S. EPA Web Guide. *JavaScript: Data Visualizations*, 2016.
17. *IBM Cloud Pak for Data as a Service*, 2021. IBM. <https://www.ibm.com/products/cloud-pak-for-data/as-a-service>
18. *IBM Db2 Warehouse on the Cloud*, 2021. IBM. <https://www.ibm.com/cloud/db2-warehouse-on-cloud>
19. *IBM InfoSphere Information Server*, 2021. IBM. <https://www.ibm.com/analytics/information-server>
20. *IBM SPSS Modeler*, 2021. IBM. <https://www.ibm.com/products/spss-modeler>
21. *IBM SPSS Software*, 2021. IBM. <https://www.ibm.com/analytics/spss-statistics-software>
22. *Platform Overview*, 2021. Looker. <https://looker.com/platform/overview>

23. *What is LookML?* 2021. Looker. <https://docs.looker.com/data-modeling/learning-lookml/what-is-lookml>
24. *Directory*, 2021. Looker. <https://looker.com/platform/directory#viz>
25. *Building Dashboards*, 2021. Looker. <https://docs.looker.com/dashboards>
26. *What is PowerBI?*, 2021. Microsoft. <https://powerbi.microsoft.com/en-us/power-bi-pro>
27. *Power BI and Esri ArcGIS*, 2021. Microsoft. <https://powerbi.microsoft.com/en-us/power-bi-esri-arcgis/>
28. *HyperIntelligence*, 2021. MicroStrategy. <https://www.microstrategy.com/en/hyperintelligence>
29. *MicroStrategy Outlook*, 2021. MicroStrategy. <https://demo.microstrategy.com/Outlook/>
30. *Drivers and Connectors*, 2021. MicroStrategy. <https://www.microstrategy.com/en/support/drivers-and-connectors>
31. *Oracle Analytics Desktop*, 2021. Oracle. <https://www.oracle.com/solutions/business-analytics/analytics-desktop/oracle-analytics-desktop.html>
32. *Oracle Analytics Cloud Overview*, 2021. Oracle. <https://www.oracle.com/middleware/technologies/oracle-analytics-cloud.html#:~:text=Oracle%20Analytics%20Cloud%20provides%20the,analytics%20that%20deliver%20proactive%20insights>
33. *Oracle Business Intelligence Cloud Service*, 2021. Oracle. <https://www.oracle.com/middleware/technologies/business-intelligence-cloud-service.html>
34. *Oracle Data Visualization – Capabilities*, 2021. Oracle. <https://www.oracle.com/business-analytics/data-visualization/capabilities.html>
35. *Panel – A High Level App and Dashboarding Solution for Python*, 2021. Panel. <https://panel.holoviz.org/>
36. *Jupyter Dashboarding – Some Thoughts on Voila, Panel, and Dash*, 2019. Met Office Informatics Lab. <https://medium.com/informatics-lab/jupyter-dashboarding-some-thoughts-on-voila-panel-and-dash-b84df9c9482f>
37. *Why Qlik is Different*, 2021. Qlik. <https://www.qlik.com/us/products/why-qlik-is-different>
38. *Qlik Sense*, 2021. Qlik. <https://www.qlik.com/us/products/qlik-sense>
39. *Qlik Sense Data Analytics Platform*, 2021. Qlik. <https://www.qlik.com/us/products/qlik-sense#tabsPanels>
40. Allen, R. *Data with a Mission: Data Visualization at US EPA*, 2018. U.S. Environmental Protection Agency. [https://cfpub.epa.gov/si/si\\_public\\_record\\_report.cfm?dirEntryId=340893](https://cfpub.epa.gov/si/si_public_record_report.cfm?dirEntryId=340893)
41. *RStudio*, 2021. RStudio. <https://rstudio.com/products/rstudio/>
42. *RStudio Connect*, 2021. RStudio. <https://www.rstudio.com/products/connect/>
43. *Shiny*, 2021. RStudio. <https://rstudio.com/products/shiny/>
44. *SAS Visual Analytics*, 2021. SAS. [https://www.sas.com/en\\_us/software/visual-analytics.html](https://www.sas.com/en_us/software/visual-analytics.html)
45. *Overview of the SAS Information Delivery Portal*, 2021. SAS. <http://documentation.sas.com/?cdcId=idpcdc&cdcVersion=4.4&docsetId=idp&docsetTarget=about.htm&locale=en>



46. *SAS BI Dashboard 4.41: User's Guide*, 2020. SAS.  
<http://documentation.sas.com/?cdcId=bidbrdc&cdcVersion=4.41&docsetId=bidbrdug&docsetTarget=titlepage.htm&locale=en>
47. *Sisense Fusion Platform*, 2021. Sisense. <https://www.sisense.com/product/>
48. *Stata Features*, 2021. Stata. <https://www.stata.com/features/>
49. Grant, R. *Making interactive online graphics from Stata, using Stata2leaflet and Stata2D3*. Kingston University and St George's University of London.  
[https://www.stata.com/meeting/uk14/abstracts/materials/uk14\\_grant.pdf](https://www.stata.com/meeting/uk14/abstracts/materials/uk14_grant.pdf)
50. *New in Stata 16*, 2021. Stata. <https://www.stata.com/new-in-stata/>
51. *About*, 2021. Tableau. <https://www.tableau.com/about>
52. *Our Platform*, 2021. Tableau. <https://www.tableau.com/products/our-platform>
53. *Welcome to the COVID-19 Data Hub*, 2021. Tableau.  
<https://www.tableau.com/covid-19-coronavirus-data-resources>
54. *Voila Notebook*, 2021. GitHub. <https://github.com/voila-dashboards/voila/tree/master/notebooks>
55. *Voila Gallery*, 2021. Voila. <https://voila-gallery.org/>
56. *ArcGIS API for Python*, 2021. Esri. <https://developers.arcgis.com/python/>
57. *R-ArcGIS Bridge*, 2021. Esri. <https://www.esri.com/en-us/arcgis/products/r-arcgis-bridge/overview>
58. *Biological Response Tools: Tools and Frameworks Available to Support Remediation Activities Associated with Wide-Area Sampling in a Biological Incident*, 2021. U.S. Environmental Protection Agency. <https://www.epa.gov/emergency-response-research/biological-response-tools>

# APPENDIX A. LITERATURE SEARCH SOURCE CRITERIA AND KEYWORDS

## 1. Sources

Sources that were prioritized were those sources expected to contain the most relevant information and meet established quality standards, including:

- Information from sources that are considered recognized, reputable, and credible;
- Information sources from nationally and internationally recognized scientific, technical, or response organizations;
- Information from written text, publications, reports, subject-matter experts, and internet sites;
- Information sources included:
  - Peer-reviewed journals, scientific manuals, and other scientific publications;
  - Federal, state, and local agency web sites or publications;
  - University web sites or publications;
  - Professional society and organization web sites or publications;
  - Recognized international scientific/environmental organizations;
  - International government web sites and publications;
  - Military web sites and publications;
  - Industry providers of tools and technologies (i.e., vendors); and
  - Conference proceedings.

Other relevant sources included articles, reports, guidance documents, case studies, national exercise materials and conclusions, after action reports, and EPA web sources that have sought to compile response and recovery guidance.

## 2. Search Criteria

The following search terms used to guide our efforts, along with a preliminary “needs” statement.

- Statistical Analysis Tools/ Platforms
- Data Visualization Tools/ Platforms
- Geospatial Visualization
- Visualization Dashboards
- Data Analytics
- COVID-19 Data Visualization

### **Needs Statement:**

*Identify the most efficient and compatible data visualization and statistical analysis tools for response to a contamination.*

## APPENDIX B. LITERATURE REVIEW SCORING CRITERIA

To standardize the review process, a *Literature Assessment Questionnaire* was used to document the overall quality of literature. The *Literature Assessment Questionnaire* was developed using Microsoft Forms, a secure online tool for publishing and conducting surveys. After the project team literature reviewer completed the form, the reviewer's evaluation was stored in a spreadsheet to document the assessment. The resulting spreadsheet was used to summarize key research findings.

### 1. General Observations

To effectively document the topic at hand, the reviewers observed the relevancy to addressing key data management needs, such as those listed here, when assessing and summarizing articles:

- Tool or Application Description
- Operational Platform
- Operating Procedures
- Example Use Cases
- Visualization Screens
- Compatibility with Data Management and Statistical Analysis Applications, and Other EPA Systems
- Fee Structure
- Licensing Requirements
- Ease of Use and Configurability
- Data Aggregation/Visualization Features
- Applicability for Incident Response
- Utilization in Prior Incidents
- Limitations
- Innovative Statistical Analysis/Data Visualization Applications or Platforms Not in Use in the Federal Sector
- Available Data Visualization Applications and/or Suites that Leverage EPA's Enterprise Investments

## 2. Literature Assessment

Relevant articles were defined as those crucial to answering research questions pertaining to handling recycling and disposal of vehicles following a wide-area event. Each article was evaluated and scored using a Likert scale (i.e., [1] Poor – [5] Excellent) based on the following seven criteria: applicability and utility, clarity and completeness, uncertainty and variability, soundness, evaluation and review, focus, and verity:

- ***Applicability and Utility:*** The extent to which the information is relevant for the intended use.
- ***Clarity and Completeness:*** The degree of clarity and completeness with which the data, assumptions, methods, QA, and analyses employed to generate the information are documented.
- ***Uncertainty and Variability:*** The extent to which variability and uncertainty (quantitative and qualitative) related to results, procedures, measures, methods, or models are evaluated and characterized.
- ***Soundness:*** The extent to which the scientific and technical procedures, measures, methods, or models employed to generate the information is reasonable for, and consistent with, the intended application.
- ***Evaluation and Review:*** The extent of independent verification, validation, and peer review of the information or of the procedures, measures, methods, or models.
- ***Focus:*** The extent to which the work not only addresses the area of inquiry under consideration but also contributes to its understanding; it is germane to the issue at hand.
- ***Verity:*** The extent to which data are consistent with accepted knowledge in the field or, if not, the new or varying data are explained within the work. The degree to which data fit within the context of the literature and are intellectually honest and authentic.

Table B-1 shows the rubric for tallying articles.

**Table B-1. Rubric for Tallying Articles**

<b>Overall Rating</b>	<b>Description</b>
35	High quality article. Article shall be recorded and summarized accordingly.
25—34	Moderately high-quality article. Article shall be recorded and summarized accordingly.
15—24	Lower quality article but with some useful information. Article shall be recorded and summarized accordingly.
<15	Unacceptable/Do not use

Articles that scored higher or equal to 15 were deemed at least moderately relevant and were recorded and summarized accordingly; however, articles scoring less than 15 were discarded from the list of relevant articles.



PRESORTED STANDARD  
POSTAGE & FEES PAID  
EPA  
PERMIT NO. G-35

Office of Research and Development (8101R)  
Washington, DC 20460

Official Business  
Penalty for Private Use  
\$300