EPA ID # 542-F-18-003

Section 1 - Introduction

This technical guide provides best practices for efficiently managing the large amount of data generated throughout the data life cycle. Thorough, up-front remedial investigation and feasibility study (RI/FS) planning and scoping combined with decision support tools and visualization can help reduce RI/FS cost and provide a more complete conceptual site model (CSM) earlier in the process. In addition, data management plays an important role in adaptive management application during the RI/FS and remedial design and action.

Why is EPA Issuing this technical guide?

The U.S. Environmental Protection Agency (EPA) developed this guide to support achievement of the July 2017 Superfund Task Force goals. Two additional companion technical guides should be used in conjunction with this data management technical guide:

- Smart Scoping for Environmental Investigations
- Strategic Sampling Approaches

This section defines the data life cycle approach and describes the benefits a comprehensive data life cycle management approach can accrue.

What is the "Data Life Cycle" Management Approach?

The Superfund program collects, reviews and works with large volumes of sampling, monitoring and environmental data that are used for decisions at different scales. For example, site-specific Superfund data developed by EPA, potentially responsible parties, states, tribes, federal agencies and others can include:

- Geologic and hydrogeologic data;
- Geospatial data (Geographic Information System [GIS] and location data);
- Chemical characteristics;
- Physical characteristics; and
- Monitoring and remediation system performance data.

In addition, EPA recognizes that regulatory information and other non-technical data are used to develop a CSM and support Superfund decisions. These data may include applicable or relevant and appropriate requirements (ARARs), future site use, population characteristics, site maps, models, exposure points, potential remedies and decision criteria. All these data are important to at least one Superfund process stage and, taken together, form the basis of an effective site management approach.

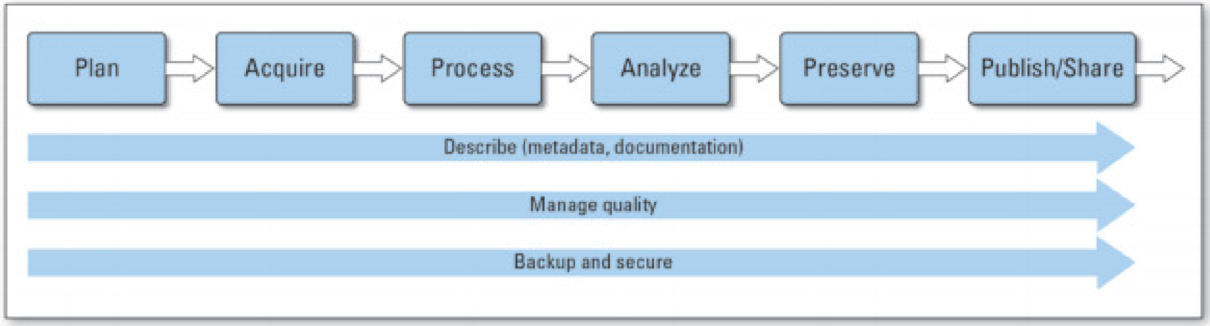
Therefore, approaching data collection and management in a deliberate and comprehensive fashion throughout the project "life cycle" should enhance the ultimate effectiveness, efficiency and defensibility of EPA's response action. Data life cycle management is potentially useful in any complex, data-intensive management process. For example, the U.S. Geological Survey (USGS) uses a data life cycle model across its programs, as illustrated in Highlight 1.[1] The model shows that throughout the data life cycle, cross-cutting program elements are necessary to ensure the data are usable for the intended purpose. First, the data must be described and documented in sufficient detail so that other data users can evaluate the results' validity and determine the data's usefulness and applicability for specific decision-making.

How is this Technical Guide Organized?

Section 1 introduces and defines the data life cycle and information flow, and describes the benefits of managing the data as part of comprehensive data management system.

Section 2 describes best practices for elements of the Superfund data life cycle. EPA recognizes that new tools and resources may be developed and has designed this document to allow for revision as one-page substitutions and additions.

Highlight 1. U.S. Geological Society Science Data Lifecycle Model



The second cross-cutting element addresses the importance of documenting data quality assurance measures at the project's inception and as data are generated. This element is particularly important when the data set contains qualitative or semi-quantitative data. The third element, data backup and security, is necessary to prevent physical data loss due to hardware or software failure, natural disasters or human error.

The concept of a data life cycle includes the individual actions, operations, or processes that must be undertaken at different stages to manage all data types, and help to ensure timely, comprehensive, and secure data management.

What is Active Data Management?

Active data management is part of a comprehensive approach that tries to minimize the time between when data requirements are set, data are collected, and when and how data are managed and made available. Active data management can improve information quality. The Superfund remedial program has traditionally used reports to exchange information. While such reports are necessary for project

---

[1] Faundeen, J.L., Burley, T.E., Carlino, J.A., Govoni, D.L., Henkel, H.S., Holl, S.L., Hutchison, V.B., Martín, Elizabeth, Montgomery, E.T., Ladino, C.C., Tessler, Steven, and Zolly, L.S., 2013, The United States Geological Survey Science Data Lifecycle Model: U.S. Geological Survey Open-File Report 2013–1265, 4 p., http://dx.doi.org/10.3133/ofr20131265

documentation, the data's usability can be diminished when managed and stored solely in the report format. For example, answering simple questions regarding a site requires additional labor costs and can require exhaustive literature searches. Also, key information can get lost in appendices and attachments. Active data management considers the "data" as the deliverable while reports serve to document data collection and, to some extent, interpret the data. The approach seeks to provide on-demand access to all site data in electronic format and reduce challenges associated with program data transition (such as removal to remedial, states to EPA, remedial project manager to remedial project manager, and potentially responsible party to EPA). This improved data interoperability can serve to limit project management costs when EPA, states, tribes, other federal agencies, and other stakeholders are reviewing and interpreting data. Management cost savings accrue because, unlike lengthy document development and comment and response, data interoperability gives all parties access to the same information, which, in turn, supports collaborative interpretation and use. Further, active data management can provide the ability to leverage nationally developed tools and provide economies of scale allowing project teams to forego re-collection of existing information or re-creation of a new data management approach at every site.

What are the Benefits of a Comprehensive Data Management Approach?

The benefits of managing the data life cycle in a comprehensive manner are:

1) Overall data quality improvement to support decision-making due to consistent content and a format that reduces data entry errors;
2) Clear data collection guidelines, processing and storage, which eliminates the cost of recollecting samples, and can preserve the integrity and availability of older information as inputs to the CSM;
3) A better understanding of data quality and any limitations when analyzing and making decisions; and
4) Improved accessibility to data in electronic format, which supports real-time interpretation and optimization of collaboratively collected data as well as use of decision support tools (such as statistical analysis, visualization, and modeling) while field crews are mobilized.

A comprehensive data management approach ensures the use of a common data platform and data consistency, accessibility, integration and versatility.

Common Data Platform. One tool to facilitate a comprehensive data management approach is a single, centralized data system. A common data transfer and storage platform provides for easy data transmission among data partners and users. A common data platform facilitates quicker decision-making because users can focus on the data content, not format. EPA regions may have regional data management plans that provide high-level minimum data requirements. Project teams are encouraged to develop data management plans at the project- and field-levels to address specialty data sets, such as those from direct sensing tools. Such plans ensure alignment of data management goals with data quality objectives (DQOs). Advanced visualization techniques requiring large data sets are often obtained from different data collection teams. A comprehensive approach to data management ensures that data collection produces a consistent data set to enhance understanding and communication of an evolving CSM. A consistent approach for collecting, processing and analyzing data facilitates the decision-making team's data transfer and integration, and allows for more effective sharing among data partners, users and project stakeholders.

Data consistency. Throughout the project life cycle, different contractors and EPA staff may be assigned to a project. A comprehensive data management approach helps ensure the team has consistent data throughout the project life cycle. Data may be generated in several phases and sometimes over many years. As additional data are generated and new interpretation tools are developed, previous site data should be readily available to the team for re-examination particularly in light of new developments or findings regarding the CSM. For example, groundwater quality data collected early in the remedial investigation (RI) can be used as a baseline to evaluate changes in mass flux or performance of a treatment system. However, to make the comparison, the older data need to be in a usable format with a clear understanding of the data's quality and usability. Changes in data elements, such as sampling methods, analytical methods, detection and reporting limits, and target analytes can be expected to change over the life of many projects. A consistent and well-documented approach to capturing, processing, storing and using data can significantly improve project teams' ability to use that data for decision-making and risk management.

Data accessibility. A comprehensive data management approach provides increased accessibility to project team members. Data are available in a consistent electronic format, often in near real time, allowing real-time interpretation and optimization of collaborative data collection, use of decision support tools (such as statistics, visualization and modeling) while crews are mobilized, and rapid evolution of the CSM to support dynamic field activities. The metadata that informs analytical results and spatial information is also captured, managed and available to support site decision-making.

Data Integration. A benefit of the comprehensive data management approach that includes regional-, site-, and field-level data management plans is that other secondary data sources, such as hydrogeologic features, precipitation, water quality, and population information, are easily integrated with the site-specific data. This data integration allows project teams to easily adjust the data assessment scale appropriately for risk management, remedial design, remedial action, community involvement or other project needs.

Data Versatility. During the post-record of decision (post-ROD) phase, site conditions may change and new information may emerge during the remedy's design, construction and evaluation, including sampling and analysis to confirm achievement of cleanup levels and remedial action objectives. To address these likely changes, EPA encourages the use of adaptive management, which provides a systematic process for planning for and responding to field conditions. Adapting the management approach and developing new solutions can require the examination of large volumes of existing data. Data life cycle management assures the data are readily available in a format that enhances project teams' and managers' ability to reliably adapt to changing site conditions.

Section 2 – Best Practices

This section describes best practices for data life cycle elements that project managers and site teams can consider during any phase of a project's life cycle. Section 2 is organized to provide best practices for each of the following data life cycle elements:

- Planning for Data Collection and Processing
- Collecting Data
- Processing Data
- Storing Data
- Making Decisions Using Data
- Communicating Data

Best Practices for Planning for Data Collection and Processing

Systematic project planning (SPP) is EPA's preferred process for building a consensus vision for conducting environmental investigation and remediation. It is a planning process that lays a scientifically defensible foundation for proposed project activities and usually includes key decision identification, CSM development in support of decision-making, and an evaluation of decision uncertainty along with approaches for managing that uncertainty in the context of the CSM.[2] The SPP is key to adaptive management. Applying SPP ensures that the project team will have adequate data to make decisions while avoiding generation of large volumes of data that do not enhance site understanding. An early SPP activity that is best conducted before data collection begins is preparation of a data management plan. A documented approach to data management, summarized in the data management plan, establishes the data management procedures throughout the data life cycle. The data planning process should follow a process that is transparent, objective and documented.

When conducting systematic planning, it is especially important to pay close attention to the following concepts:

1) Develop a comprehensive CSM so that the project team understands existing data and data needed to fill identified data gaps. A comprehensive CSM also helps prepare for the unexpected in the field. Understand analytical and spatial data needs but also plan to capture and manage important metadata electronically.
2) Engage stakeholders and end data users to ensure data collected will not only meet DQOs but will provide an appropriate data set for multiple end uses, such as risk assessment, risk management, feasibility analysis, remedy design, state/tribal review and communication with local officials and community members.
3) Exercise data tool outputs and field procedures with a data management plan and data management tools to fully understand the data that will be generated, including how that data

---

[2] Definition of systematic project planning. https://triadcentral.clu-in.org/gloss/dsp_glossterm.cfm?glossid=223.

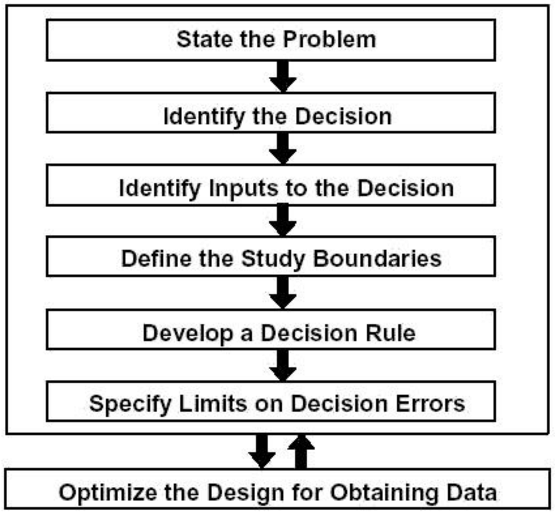
will be obtained; provided; and processed, stored and used. These factors have implications for data format, decision support tool inputs, processing procedures and more.

4) Automate data management activities where appropriate and check data quality at the point of generation.

5) Designate a data management professional for your site, project and field effort as part of your team.

6) Leverage existing tools to gain efficiency and economies of scale. Many have sufficient flexibility to accommodate field, site and regional data management needs.

Highlight 2. DQO Process.



When considering the data needs and data collection activities to support decision-making, project teams use the DQO process to align the data approach with the intended project decisions (Highlight 2). The EPA's "Systematic Planning Using the Data Quality Objectives Process" (EPA/240/B-06/001*)* explains the DQO process at an agency-wide, cross-program level. Since the nature of data and decision-making varies greatly among EPA program offices, agency-level DQO guidance is necessarily non-specific. The Intergovernmental Data Quality Task Force prepared a template quality assurance project plan (QAPP) that applies the DQO process specifically to cleanup investigations. The "Uniform Federal Policy for Quality Assurance Project Plans" template is a series of worksheets that improve data quality and project outcomes by prompting the user to develop the information and quality control procedures that fulfill the DQO process. Housed on Superfund's Federal Facilities website, the template is supported by a manual, training materials and other tools that assist in the DQO process' application at hazardous waste sites. In addition to aiding QAPP writers, the standardized worksheet format speeds QAPP approval staff's review.[3]

The intent of data collection should be more than informing the project team of next steps. Systematic project planning defines the project's direction, DQOs help to answer how the project team might arrive there, and data management ensures information can be used to make those decisions.

---

1) [3] U.S. EPA (2012) Intergovernmental Data Quality Task Force, Uniform Federal Policy for Implementing Environmental Quality Systems: Evaluating, Assessing and Documenting Environmental Data Collection/Use and Technology Programs, March. https://www.epa.gov/fedfac/assuring-quality-federal-cleanups and https://www.epa.gov/sites/production/files/documents/ufp_qapp_worksheets.pdf

Best Practices for Collecting Data

During the RI/FS stage, many different data types are collected. It is beyond this guides scope’ to address the best practices for numerous techniques and tools that could be used. When considering sampling strategies to address CSM data gaps, project teams should develop an understanding of the data generated by each, and consider the following best practices:

1) **Identify the levels of quality needed.** The levels of quality (such as precision and accuracy) needed for each technique should be established before data collection. Identify the measurement units and develop measurement metrics to ensure the right types of data are collected. Using collaborative data sets and multiple lines of evidence can improve the site team’s understanding of conditions even when the quality of a single instrument or data collection method alone may not be sufficient. Consider quality in the context of all collaborative data streams and lines of evidence; assess where variability is coming from (such as sample design, sample processing, extraction, or analytical). Improve quality by directly addressing the variability source this assessment identifies. For example, under EPA’s recommended incremental sampling design, potential variability is addressed through the collection of triplicate samples and seven additional analyses.

2) **Assess reliability of data sources.** All data sources’ reliability, including direct instrument measurement, should be assessed. Reliability can relate to the ruggedness of the physical instruments used to collect the data and the ability to perform under all anticipated field conditions, or, the consistency of the readings generated by the technique. For example, cone penetrometer testing logs present soil lithology based on unbiased physical measurements, whereas borehole logs may be subject to the logger’s interpretation and experience. If permeable zone correlation is the primary driver for collecting lithology information, then hole-to-hole consistency in soil type interpretation is critical, and interpretation of permeability by different loggers can be problematic. Standardizing core descriptions of non-aqueous phased liquid in planning documents, taking physical core measurements such as grain size and permeability, and logging boreholes using consistent methods such as the unified soil classification system are techniques that help to limit variability associated with multiple professionals providing bore log descriptions.

3) **Consider data quality.** Transcription and electronic recording and download errors can affect data quality. Manual data collection is subject to random transcription errors during collection when recording readings from instruments. Scale factors, correction factors, calibration, instrument stabilization, and field conditions may generate systematic error in electronic files. Mislabeling data files for download is another source of systematic error. Taking clear and detailed field notes of data transfer activities should help identify and correct these errors. In addition, creating valid values lists and using them to automatically flag errors and performing data audits are ways to check and ensure data quality. Conducting a demonstration of method applicability for field techniques not only provides an opportunity to understand sampling design, sample preparation, and instrument performance for a given site matrix, but it also allows project teams to optimize all the procedures that may impact data quality.

4) **Verify each manual data entry and transfer**. Quality control begins with ensuring the initial recording of a data point accurately reflects the measurement or condition. Ensure that field teams

are trained, and systems (valid values lists, data audits) are in place to verify initial data collection quality.

5) **Establish standard systems for identifying locations and sample media types**. A unique numbering system for sample locations and sample types should reduce the likelihood of mislabeling and improve data review and management efficiency. Consider the following practical tips:

    - Check existing databases and the CSM information to see what sample or well identification descriptions (ID) have already been used so as not to duplicate an existing numbering system. When multiple contractors, parties and regulatory agencies are working on a site, it can be easy to duplicate sample or well IDs.
    - Be careful as to how much information a sample ID contains. For data sorting and filtering, it is better to add fields to the database that describe individual sample points, such as depth, rather than to capture this information with the sample ID. In addition, the use of too many characters in a sample ID increases the potential for transcription errors.
    - Limit the amount of interpretation field crews must make with regard to sample IDs. The following example illustrates why field crew interpretation should be limited:
        - At a recent field effort using incremental sampling in combination with x-ray fluorescence, the QAPP for the demonstration of method applicability (DMA) required the field crews to collect composite and incremental soil samples at four depth horizons:
            - Depth 1: 0-1 inches (bare soil) or 0-2 inches (vegetated soil)
            - Depth 2: 1-6 inches (bare soil) or 2-6 inches (vegetated soil)
            - Depth 3: 6-12 inches
            - Depth 4: 12-18 inches
        - The field crews struggled with determining whether a sample was from “bare” or “vegetated” soil and which depth horizon to use. In addition, the sample IDs included this depth-specific nuance, and the top two intervals’ sample IDs were mislabeled and had transcription errors. In this case, the DMA was a valuable tool in correcting these problems before the full sampling effort’s initiation. The improved methodology will consider the top interval to be 0-1 inches regardless of vegetation, and the QAPP will specify and stress to the field crew that, for lawns or vegetated soil, the 0-1-inch interval starts at the root mass base.

6) **Use electronic data forms.** Using consistent data formats and software from project initiation to completion improves data collection efficiency and consistency. Scribe is an EPA software tool used to collect and manage environmental data. It can import electronic data deliverable (EDD) files, including analytical laboratory EDD files and locational data EDD files, such as GPS data. Scribe outputs include labels for collected samples, electronic and hardcopy chain-of-custody generation, and analytical laboratory result data reports. Scribe users may manage, query and view data, and

can export electronic data for use with GIS tools and in reports. The EPA strongly prefers Scribe's use for Superfund data collection and management, but the Agency also supports the concurrent use of other commercially available software that can enhance data integration and visualization. Using standard software for data collection and management: ensures all pertinent data are collected and recorded in a consistent and repeatable manner; assists in seamlessly transferring information among stakeholders; and minimizes the likelihood of transcription errors.

7) **Track metadata**. Metadata includes information on a data resource's content, such as the data source, limitations, access and use restrictions, data quality, and contact information. These descriptive fields help a user decide if a data set is appropriate for their proposed use. Many electronic data systems contain metadata fields where these data can be entered but few data management strategies plan for how that data will be managed. Field logbooks can capture some of this information but are generally not formatted to capture all pertinent metadata in a consistent format; further, retrieving and interpreting that information months or years after a field effort can be challenging.

8) **Require accurate geospatial information.** Accurate geospatial location information is essential to site data interpretation. Collect GIS-compatible data, when appropriate. A GIS platform provides a standard base for communicating, transferring and interpreting all data types. Ensure that the coordinate system used for the GIS is geo-referenced to the site, not a stand-alone coordinate system. Check the accuracy of the site attributes to make sure the data represent the most current site configuration. Distribute GIS files to data partners to facilitate data transfer and interpretation. A GIS system generally relies on high accuracy x and y coordinates to locate a sample on the earth's surface. Data interpretation, however, is critically affected by the accuracy and maintenance of the depth or z coordinate. Accurate geospatial information must therefore include high accuracy surface elevation, sample depth, well screen depth, depth within a well screen, and other critical vertical information. Depth information should be in separate fields within the database. For additional resources on ensuring accuracy of geospatial information please see the following website: https://www.epa.gov/geospatial/geospatial-policies-and-standards

9) **Use and verify electronic data delivery of laboratory data.** Superfund projects often utilize multiple analytical laboratories to analyze field samples and report data. The site team should follow EPA requirements to ensure the analytical laboratory has expertise in the requested analysis, and can provide quality data in the required EDD format to support project decisions. Analytical laboratories should have access to relevant portions of the project QAPP, and may be asked to aid in developing DQOs for the project. Upfront communication with the analytical laboratories is key to ensuring the laboratories can deliver the required data in a consistent and compatible format. The EPA strongly recommends the use of Scribe-compatible EDD file formats. The staged electronic data deliverable (SEDD) is a uniform, Scribe-compatible format developed by the federal government for electronic delivery of analytical data, which can improve the efficiency of analytical laboratory data delivery, review, storage, and retrieval. Using a consistent EDD format, such as SEDD, can make the data review and evaluation most cost-effective and efficient by reducing transcription errors and automating portions of the data review process. At minimum, analytical laboratories should deliver data in a Scribe-compatible format that is consistent with historical data requirements and other

software being used for site data management. Several EPA regions have adopted additional requirements for analytical laboratory EDD submissions to ensure laboratory data are consistently reported and can be assessed against EPA data quality requirements.

For more information on SEDD, please see the following website:
https://www.epa.gov/clp/staged-electronic-data-deliverable-sedd

For more information on EPA's EDD submission process in Superfund please see the following website:
https://www.epa.gov/superfund/epa-superfund-electronic-data-submission-multi-regions-edd

Best Practices for Processing Data

Many data types require processing prior to use to ensure they are ready for integration and analysis. According to the USGS Science Lifecycle Model, processing includes various activities associated with preparation of new or previously collected data inputs. Data processing should entail definition of data elements; integration of disparate datasets; extraction, transformation, and load operations; and application of calibrations to prepare the data for analysis. The most common example of these practices are fixed laboratory chemical analysis data. The laboratory follows rigorous quality control procedures and provides users with information to assess data quality; however, the user is responsible for independently determining the quality of the data set through a formal data validation process. Data from direct reading instruments and field methods can also be subjected to review through data quality checks developed by the project quality assurance team. Some data may need to have correction factors applied or converted to standard units prior to use. Environmental data sets are often disparate and processing includes considerations for CSM integration. For example, direct sensing data from an electrical conductivity meter should be integrated with both relative hydraulic conductivity data from hydraulic profiling and lithologic logs from soil borings to provide cross-checks on each other and to determine if these multiple lines of evidence converge. In addition, collaborative data sets (such as x-ray fluorescence [XRF] data and laboratory data measuring the same contaminant) can be used to improve spatial information even if statistical correlations between the two data sets are poor. Processing involves not only preparing data for integration and analysis but also determining how disparate data sets will be used to inform the CSM.

As data are exchanged and transferred from their initial source to databases or other intermediate platforms, errors and incomplete exchanges can occur. A best practice is to verify the data transfer or import to ensure the original data's integrity. Determine if spot checks or 100 percent checking of the data are necessary based on the limits for decision errors identified in the DQO analysis.

Best Practices for Storing Data

The application of dynamic work strategies in Superfund investigations involves collecting a large volume of data, interpreting the data in real time, and making real-time decisions. A robust system for data storage and long-term preservation is necessary to ensure the data are available, complete, and accurate during the project life cycle. A data storage strategy is part of the SPP and includes a detailed discussion of data management procedures, equipment (software and hardware), lines of communication, reporting formats, and time frames for implementing data storage activities. Data planning assists the project team to adequately assess costs and resource needs associated with data management.

Many electronic database systems are available to store data on a temporary (field) or permanent basis. Scribe and Scribe.net are data management systems developed by EPA and used to provide site-specific, flexible data collection, management, and exchange. Additional information on field-oriented data management using these and other tools is available in EPA Triad Central Technology Bulletin Management and Interpretation of Data Under a Triad Approach (EPA 542-F-07-001), https://www.epa.gov/remedytech/management-and-interpretation-data-under-triad-approach-technology-bulletin. Ruggedness, acceptable data content, and ease of use are important considerations for selecting field data storage systems. Database systems should be scalable (able to accommodate both small and large amounts of data) and transportable (easily moved).

The agency's regional offices have adopted a variety of data management systems for storing site-level data. Two common platforms used for data storage in the regions. One, WQX/STORET is EPA's water quality storage system, and the second, EQuIS, is a commercial data management and decision support system for soil, water, air, geotechnical, and other environmental data types. Some data storage solutions have electronic data checker tools to assess data quality and manage submission and export formats. In addition, some have modules to simplify field data collection and manage large-volume sensor data collection activities. It is important to work with regional data management specialists and the appropriate EPA program office to identify applicable policy and requirements during the project planning phase.

The advantages of comprehensive data storage systems are that data can be organized for easy retrieval and use, and the data are in a single, secure location. Some data storage systems use proprietary or license-based software, such as ArcGIS and EQuIS, and team access may require software ownership. Many data systems have free software to allow stakeholders to view, but not manipulate data. Ensuring a data management system's long-term integrity is part of the planning process and includes methods for securing databases and managing users' rights to upload or change data. For additional information on maintaining data security from electronic tampering, preventing loss of electronic data quality while in electronic storage, and unauthorized release of electronic data or personally identifiable information, please see EPA's privacy policy found on this website: https://www.epa.gov/irmpoli8/epas-privacy-policy-personally-identifiable-information-and-privacy-act-information

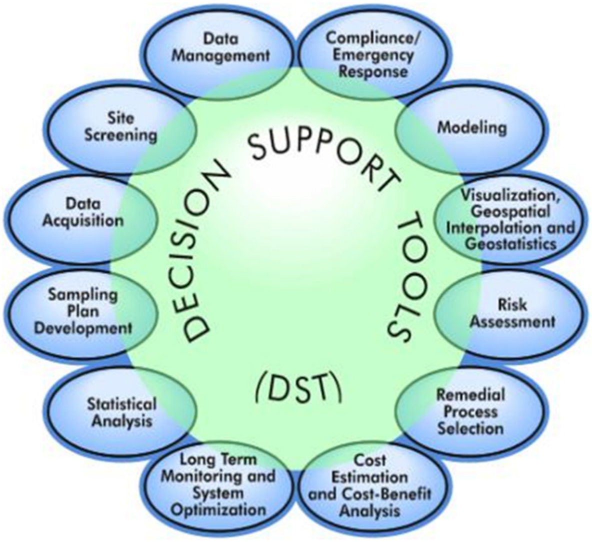
Best Practices for Making Decisions Using Data

Decision-making for contaminated sites usually involves integration of different data sets from many technical areas. Systematic project planning incorporates decision logic flow diagrams to guide the field decision-making process. Using collaborative data sets and multiple lines of evidence further strengthens data interpretation while providing increased confidence in CSM development. Where collaborative data support each other and multiple lines of evidence converge, the project team has increased confidence in interpreting that data or CSM component. Conversely, instances where collaborative data sets or lines of evidence diverge or lead to different conclusions may indicate the presence of data gaps, inconsistent spatial scales, or the need to update a CSM component to account for differences.

Decision support tools (DSTs) are interactive software tools that use data. These tools can be used by decision-makers throughout the project life cycle to help answer questions, solve problems, and support or refute conclusions. They can be incorporated into a structured decision-making process for environmental site cleanup (Highlight 3). Individual tools may integrate data from many different technologies including GIS, global positioning system (GPS), databases, and visualization tools. Using DSTs is a best practice to provide a transparent, standardized, reproducible approach to data analysis that can incorporate and quantify uncertainty in the data sets and decisions.

The DSTs can be used to support specific project tasks such as statistical data evaluation, sampling design (visual sampling plan) or groundwater sampling optimization (monitoring and remediation optimization system software), or multiple functions required for data acquisition, spatial data management, contaminant modeling, and cost estimating (spatial analysis and decision assistance software). Detailed analysis of several DSTs' attributes and applications can be found in the DST matrix on the Federal Remediation Technology Roundtable website: http://www.frtr.gov/decisionsupport/index.htm.

Highlight 3. Examples of Decision Support Tools.



The EPA has found that utilizing high-resolution site characterization tools and strategies along with visualizing those results can lead to a better understanding of the CSM and more effective and targeted remedial actions. Further, these tools and approaches can be combined to expedite field investigations and drive dynamic work strategies as well as facilitate timely and collaborative data by stakeholders. Recent Superfund pilot projects and institutionalization of EPA's optimization program have shown these approaches can expedite project schedules; reduce transaction costs for data sharing among stakeholders; and lead to a transparent, fast, and collaborative approach to site decision-making.

Visualization technology (for example, three-dimensional visualization and analysis or 3DVA) is a valuable decision support mechanism that integrates geology, hydrogeology, and contaminant chemistry data into a single spatially correct visual model. Geostatistical algorithms within these software packages can further help teams to interpret data or interpolate between data points. It is important to note however, that these geostatistical interpolations must be performed by qualified professionals and are subject to further refinement based on collaborative data, other lines of evidence and professional judgment of key technical team members. For example, geologic interpretations can be supplemented with environmental sequence stratigraphy as described in Best Practices for Environmental Site Management: A Practical Guide for Applying Environmental Sequence Stratigraphy to Improve Conceptual Site Models published in EPA's groundwater issue from September 2017, EPA /600/R-17/293, https://semspub.epa.gov/work/HQ/100001009.pdf. Visualization capabilities in

DSTs allow the user to rapidly assimilate new field data and display information to support dynamic field decision-making.

The use of DSTs and visualization tools are best practices and support implementing adaptive management throughout the project life cycle. By applying the appropriate DSTs and viewing the most current data and interpretations, stakeholders can evaluate options and develop data collection contingencies and identify logical sequencing of field investigation tasks. Stakeholders can make decisions in real-time and adapt sampling strategies to reflect the most current CSM, potentially saving time and resources.

Best Practices for Communicating Data

The EPA encourages the use of dynamic work strategies and real-time data collection; however, these approaches require project teams to evaluate and respond to data quickly. For many sites data collection teams, decision-makers, and stakeholders are geographically dispersed and timely data sharing can be a challenge. Several collaboration tools are available to communicate data among teams. Many of these tools can also be used as portals for teams to store and access information over the project life cycle. The EPA has found these communication tools particularly useful for sharing data visualization and CSM products. Examples of data communication tools include:

- Project- and site-oriented websites where team members can quickly and securely share information, such as EPA on-scene coordinators’ website and SharePoint sites.
- Custom, project-specific websites and databases, developed by the regions for storing data, visualizing and exchanging information with stakeholders.
- Virtual meeting tools or commercial web conferencing (Adobe Connect, Skype, Go To Meeting, Meeting Place) allow teams to review and discuss information as if they were in the same physical space.

More data evaluation and storage tools are becoming available, and project teams can use SPP to identify the data communication tools and procedures to be used throughout the data life cycle. In general, dynamic and expedited field efforts require timely stakeholder data evaluation and decision-making along with a resource commitment to meet timely data review demands and real-time decision making. In many cases geographically dispersed project teams are at an advantage for this project type. For example, during a recent Superfund preliminary design investigation, the project team used direct sensing tools to rapidly characterize the contaminant distribution and subsurface geology and hydrogeology at a West Coast site. At the end of each day, the field data were uploaded to a project team website and an East Coast visualization expert would download the data and update the CSM thereby assuring the project team’s ability to view the data the next morning. With the updated information the technical team would meet remotely to discuss and plan future activities. The result was a completion of field activities, interpretation of data, generation of a report, and move to remedy design in weeks versus months.

Conclusion

Approaching data collection and management in a deliberate and comprehensive fashion throughout the project “life cycle” should enhance the ultimate effectiveness, efficiency, and defensibility of EPA’s response action. Applying best practices to data management activities increases the data’s usefulness and allows the use of new data interpretation tools and programs.

Disclaimer

The use of these best management practices may require site-specific decisions to be made with input from state, tribal, and/or local regulators and other oversight bodies. The document is neither a substitute for regulations or policies, nor is it a regulation or EPA guidance document itself. In the event of a conflict between the discussion in this document and any statute, regulation or policy, this document would not be controlling and cannot be relied on to contradict or argue against any EPA position taken administratively or in court. It does not impose legally binding requirements on the EPA or the regulated community and might not apply to a particular situation based on the specific circumstances. This document does not modify or supersede any existing EPA guidance document or affect the Agency’s enforcement discretion in any way.