ITRC has developed a series of fact sheets that summarizes the latest science, engineering, and technologies regarding environmental data management (EDM) best practices. This fact sheet describes:

- Defining overarching data quality questions for planning
- Understanding project needs and application to data
- Understanding cost of and budget considerations for data quality.

Additional information related to data quality is provided in the ITRC documents Using Data Quality Dimensions to Assess and Manage Data Quality; Considerations for Choosing an Analytical Laboratory; and Data Quality Review: Verification, Validation, and Usability.

1. INTRODUCTION

1.1 What Is Data Quality and Why Is It Important?

Data quality can mean different things in different projects or environmental situations. The requirements for data quality (often called data quality objectives or DQOs) depend upon the needs of each project or situation requiring data use. For example, environmental data quality requirements may be related to the determination of the presence/absence of a spilled material, or quantification of specific contaminants of concern within specific accuracy and precision limits. In a non-remediation scenario, a DQO may be counts of fish species by best visual observation within a given timeframe following a stream restoration project.

Data quality considerations may involve the use of appropriate spatial/temporal boundaries and spatial sampling density; methods used for field collection, laboratory analysis, and data validation methods; and consistency of data representation across multiple investigations. Data quality considerations for data storage and management may address topics such as the requirements of managing data securely for litigation purposes or how to convey information to the public about environmental conditions in their communities, with 24-hours-a-day/7-days-a-week (24/7) accessibility and availability of the data with no changes to the data allowed (read-only permission).

Understanding what is right for you, your project, and your data can determine and guide proper protocols and considerations that will support you and your project in the acquisition and use of environmental data. A clear understanding of the data needs and expectations is crucial. Establishing DQOs, or the quality requirements to be met before data are used in project decisions, evaluations, and conclusions, is essential to managing defensible data. For some projects, these DQOs are formally developed in various planning documents, such as quality assurance project plans (QAPPs), where data quality indicators of precision, accuracy/bias, representativeness, comparability, completeness, and sensitivity (PARCCS) are well defined. Less formally, DQOs might be established in various data management standard operating procedures (SOPs) or documented best practices.

The data quality documentation provided by this ITRC EDM subgroup is intended to help you start thinking about what data quality means and guide you to appropriate resources for your data and project. It is expected (and important) that data quality be monitored throughout the project and data lifecycles. Section 2 below begins the discussion on how these concurrent lifecycles work together. Additional details relevant to the data lifecycle can be found in the Data Lifecycle Fact Sheet. Throughout the project (and data) lifecycles, data quality will be visited and revisited with respect to project needs, considering project scale and complexity, data quality dimensions (see Section 4 below), and cost to the government agencies or private entities that will be using the data to make decisions. These resources are intended not to be prescriptive but to provide flexible overarching principles to inform progress, changes, or issues that may arise around environmental data quality for your project.

1.2 Why Ask the Overarching Data Quality Questions?

1.2.1 Get Started on the Right Path

When you start a new environmental data collection activity, it is important to get started on the right path by asking questions such as: What kind of project do I have? Why is the data important to my project? What are the intended uses of the data? and Who is my audience? (Figure 1).

Asking these overarching data quality questions is important because it helps you identify your project objectives to ensure

data quality. In other words, asking these questions helps you identify the data that are needed with the appropriate level of quality for their intended uses. Not asking these questions increases the risk of losing valuable time and resources collecting data that—while they may be of high quality—don't answer your project questions. The U.S. Environmental Protection Agency (USEPA) guidance on DQOs describes the risk well: "Unless some form of planning is conducted prior to investing the necessary time and resources to collect data, the chances can be unacceptably high that these data will not meet specific project needs." (USEPA 2006).

1.2.2 Take the Next Steps

Considering these questions will point you along the right path to identifying what the DQOs are for your project. But what are the next steps? The project lifecycle stages are identified in this document as "Plan" (project scoping and planning activities), "Execute" (implementation of those activities), and "Close" (completion, identification of lessons learned and finalization), and are related to the temporal nature of a project. Adaptive management, sometimes referred to as project monitoring, occurs continuously at each stage of the project lifecycle. The data lifecycle stages, which overlap or coincide with the project lifecycle stages, are equally important in our discussions and have been represented by the following categories: "Plan," "Acquire," "Process/Maintain," "Publish/Share," and "Retain". Data quality improvements occur continuously at each stage of the data lifecycle and may continue to occur for the lifetime of the data beyond the end of the program/project.

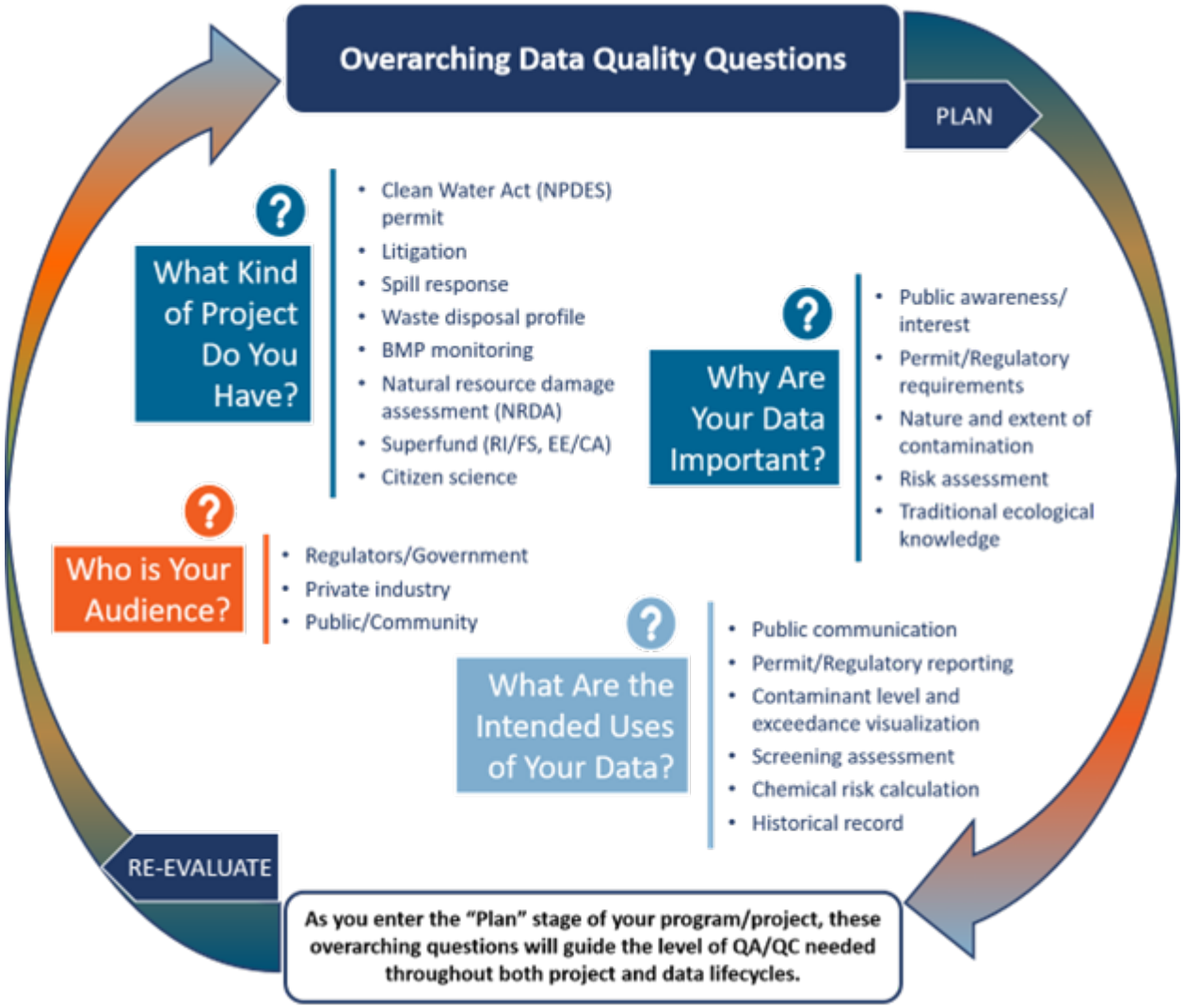


Figure 1: Overarching data quality questions to consider.

This introduction has focused on preplanning, emphasizing the need to get started on the right path, and the concomitant data quality considerations. As you enter the planning stage of your project, Section 2 provides guidance and tools for considering the finer-grained aspects of data quality, or data quality dimensions, of your data as you move through the acquisition, processing, and data evaluation and sharing stages of the data lifecycle. The next section also presents a crosswalk between the project lifecycle and the key data quality considerations for the overlapping data lifecycle. As you

move through the different project and data lifecycle stages of your project, it is important to keep in mind that the planning process should be iterative, and re-evaluation of your overarching data quality questions is encouraged, and may be necessary, to stay on the right path.

2. THE IMPORTANCE OF PLANNING

"Fail to plan: plan to fail." This adage, based on a quote widely attributed to Benjamin Franklin, is particularly apt in the context of environmental data management. Planning appropriately for your project, program, and environmental data set is a key aspect of ensuring data quality. As shown in Figure 2, both project and data lifecycles begin with "Plan." Understanding your projects needs and successfully planning for them through planning documents such as a sampling and analysis plan (SAP), QAPP, or data management plan (DMP) can be used to provide key pieces of information for your project. Taking time at the beginning of your project or data collection effort to develop the appropriate planning document(s) for your project will set the stage and guide communications and workflow as you move to the "Execute" and "Close" stages of the project lifecycle. The crosswalk between the project and data lifecycles illustrated in Figure 2 shows that, although your project likely has a finite timescale, data quality considerations at the "Plan," "Acquire," "Process/Maintain," "Publish/Share," and "Retain" stages of the data lifecycle both overlap with the project lifecycle stages and extend throughout the lifetime of your environmental data set. Adaptive management (that is, adjusting your plan as necessary) is an important aspect of planning for both project and data lifecycles.

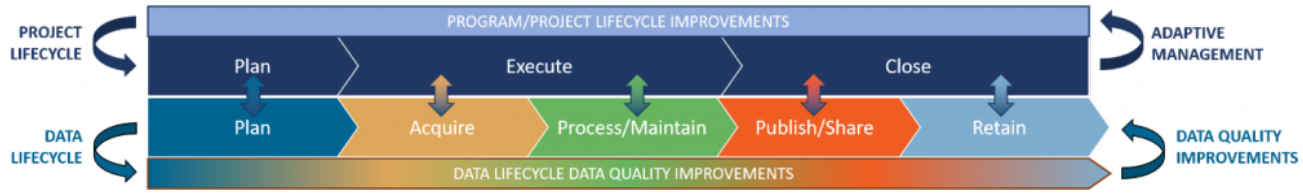


Figure 2. Crosswalk between project lifecycle and data lifecycle stages.

Crosswalk between Project and Data Lifecycles

Figure 2 depicts a crosswalk between program/project and data lifecycles, showing a generalized chronological alignment of individual lifecycle stages. The vertical arrows between lifecycle stages illustrate how feedback from the overarching project data quality questions informs data quality at each stage of the data lifecycle, and how the data collected at these stages inform the overall project questions. The circular arrows encompassing each lifecycle illustrate that both lifecycles are iterative; that is, adaptive management and data quality improvements may be continuously considered and performed for the project and data lifecycles, respectively. The project lifecycle is shown on top because it often precedes the data lifecycle. The overarching data quality questions considered in the "Plan" stage of the project lifecycle inform both the "Plan" stage of the data lifecycle and the actual data collection. Finally, project lifecycles generally have a finite duration that ends after the "Close" stage, whereas the data lifecycle exists for the lifetime of the data, which typically extends beyond the end of the project.

The ITRC EDM best practices apply to many aspects of EDM, which are addressed by documentation and training materials from the various ITRC EDM subgroups, including Data Management Planning, Field Data Collection, Data Exchange, Geospatial Data, Public Communications and Data Accessibility, and Traditional Ecological Knowledge. Data quality considerations at each stage of the program/project and data lifecycles are introduced in this overview document, but it is important to also make the connections between the Data Quality work products and the topics discussed in other groups of work products. To support an understanding of these connections, Figure 3 shows the project and data lifecycles in relation to data quality considerations. This figure also provides links to specific data quality work products for each step in the project and data lifecycles. Note that the matrix may not show all possible connections between the lifecycles and data quality, but rather is intended as a starting point showing some key connections between commonly encountered data quality considerations and at each lifecycle stage.

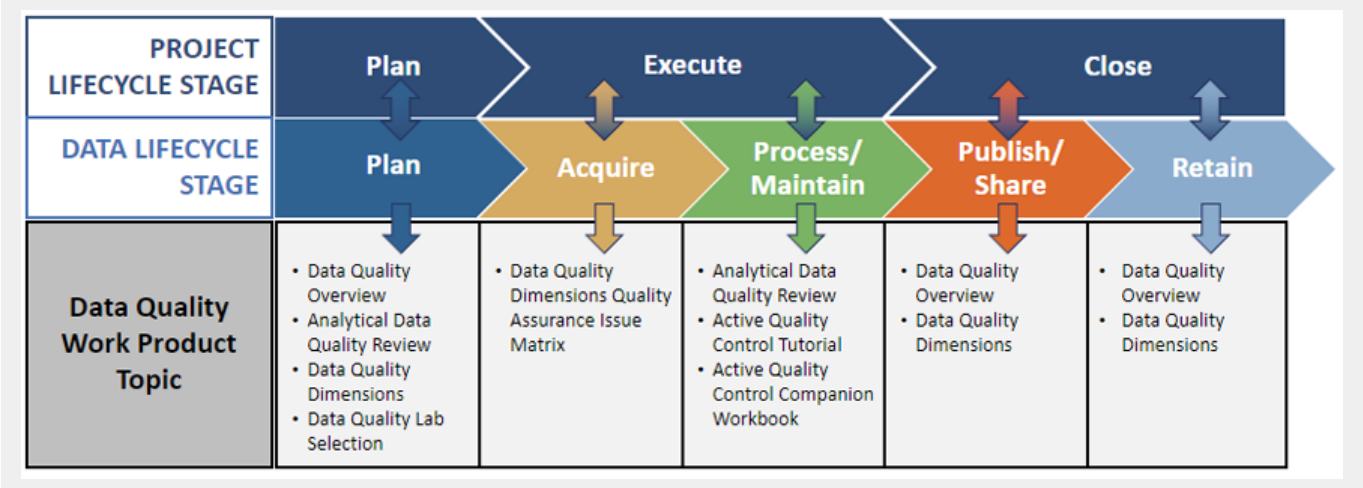


Figure 3. Cross reference of data quality work product contents relative to stages in the project and data lifecycles.

Lifecycle Connections to Data Quality Work Products

Figure 3 connects the project and data lifecycles crosswalk (Figure 2) with the work products prepared by the Data Quality subgroup. The intent of this figure is to help readers address specific needs for the different aspects of environmental data management at the various stages of the lifecycles. For each stage of the lifecycles shown on the crosswalk, Table 1 provides links to key data quality guidance documents that were compiled by the ITRC Environmental Data Management Best Practices team.

During the project planning stage, it is important to revisit the overarching questions, in particular “What kind of project do you have?” and “What are the intended uses of your data?” Depending on the answers to these questions, your planning document may be a simple field form or a full QAPP following USEPA guidance. Whatever the extent, planning documents provide key pieces of information for the development of data quality for many projects. A planning document sets forth details regarding project management and tasks, data generation and acquisition, data assessment and oversight, and data validation and usability. Planning for an analytical data program (for example, requirements for analytical methodologies and data reporting, including numerical reporting limits and data formats) may also be required for your project. In these cases, it is crucial to understand the needs of the project and key questions to ask when aligning with an analytical laboratory. Further guidance for an analytical data program is provided in the Considerations for Choosing an Analytical Laboratory subtopic sheet. Ideally, the planning document will be the result of careful discussion of the overarching data quality questions among the project team. Regardless of scope or detail included in a planning document, such documents are the road map to acquiring the data of the type and quality necessary to meet project needs.

Federal, state, or regional project managers can provide guidance on which type of planning document is needed for your project. DQOs are usually defined in a QAPP. The QAPP can be a stand-alone document or part of a larger, comprehensive work plan. Some regulatory programs require that a QAPP be developed on an individual project basis; however, some states have established program-wide DQOs (such as for underground storage tank remediation) and have their own program QAPPs. In these cases, it is possible that only a SAP for a specific task or sampling event needs to be written in compliance with the program’s QAPP.

Commonly used federal, regional, and state quality assurance planning guidance documents related to data quality are listed in Table 1.

Table 1: Links to data quality planning documentation

Government Organization	Reference
Federal Guidance	
US Department of Defense (DoD)	Uniform Federal Policy for Quality Assurance Project Plans (QAPPs)
US Department of the Navy (DoN)	DoN Systematic Planning

US Environmental Protection Agency (USEPA)	How EPA Manages the Quality of its Environmental Information | US EPA Guidance for Quality Assurance Project Plans, EPA QA/G-5 | US EPA Guidance on Systematic Planning Using the Data Quality Objectives Process, EPA QA/G-4 | US EPA Agency-wide Quality Program Documents | US EPA
Regional Landing Pages for Quality Assurance	
US EPA Region 1	Managing the Quality of Environmental Data at EPA Region 1 | US EPA
US EPA Region 2	US EPA Region 2 Guidance for the Development of Quality Assurance Project Plans for Environmental Monitoring Projects | US EPA
US EPA Region 3	EPA Region 3 Quality Assurance Project Plans | US EPA
US EPA Region 4	Region 4 Quality Assurance Project Plan (QAPP) Tool Box | US EPA
US EPA Region 5	Managing the Quality of Environmental Data at EPA Region 5 | US EPA
US EPA Region 6	Managing the Quality of Environmental Data at EPA Region 6 | US EPA
US EPA Region 7	Managing the Quality of Environmental Data at EPA Region 7 | US EPA
US EPA Region 8	Managing the Quality of Environmental Data at EPA Region 8 | US EPA
US EPA Region 9	Region 9 Wetlands Quality Assurance Project Plan Guidance and Template | US EPA
US EPA Region 10	Quality Assurance Project Plans for Tribes in Region 10 | US EPA
State Guidance	
Alaska (ADEC)	Division of Water Quality Assurance Division of Air Quality Monitoring and Quality Assurance Division of Spill Prevention and Response Contaminated Sites Remediation Program
Connecticut (DEEP)	Quality Assurance (ct.gov)
Illinois (IEPA)	Guidance—Resource Assessments (illinois.gov)
Indiana (IDEM)	IDEM: Nonpoint Source: Quality Assurance Project Plan Template Instructions
Massachusetts (DEP)	Quality Assurance Project Plans
Minnesota (MPCA)	MPCA Quality Assurance Project Plan Guidance (p-aeo2-13) Underground Storage Tank and Petroleum Remediation Quality Assurance Program Plan (p-eao2-04) Toxic Substances Control Act Polychlorinated Biphenyl Inspection Program QAPP (p-eao2-08) \
Missouri (DNR)	Missouri Department of Natural Resources Data Quality Management Brownfields Quality Assurance Project Plan Templates Data Quality Management | Missouri Department of Natural Resources (mo.gov)
New Jersey (DEP)	Quality Assurance Project Plan Technical Guidance
New York State (NYSDEC)	Quality Assurance and Quality Control of Water Quality Data
South Carolina (DHEC)	Guidance Document for Preparing Quality Assurance Project Plans (QAPPs) for Environmental Monitoring Projects/Studies Underground Storage Tank (UST) Quality Assurance Program Plan (QAPP)
Texas (TCEQ)	Quality Assurance Project Plans for Nonpoint Source Projects Quality Assurance and Monitoring Procedures for Surface Water Quality Monitoring Air Quality Research: Quality Assurance Texas Risk Reduction Program
Washington State (Ecology)	Guidelines for Preparing Quality Assurance Project Plans for Environmental Studies Washington State Department of Ecology Quality Management Plan 2020

3. UNDERSTANDING PROJECT NEEDS AND APPLICABILITY TO DATA

To be able to meet the data quality goals of your project, program, or data collection activity, you need to understand the "end goal" and project needs. Depending on the project needs and the intended use of the data being collected, the requirements and scale of complexity will vary. For example, a project that is collecting data for purely informational purposes will have less stringent requirements than one that has intended data usage for regulatory compliance or litigation purposes. Identifying the data needs and intended usage helps to define the criteria surrounding acceptability and measurement metrics for assessment associated with the key data quality dimensions (Section 4). To support this understanding and help the data user identify requirements of each data gathering component, an inventory of the types of data to be obtained and activities to be conducted should take place. Overall systematic planning discussions should be used to outline the objectives of the project or program and therefore support an understanding of the project and data needs and requirements. The ITRC EDM Data Quality subgroup has outlined a set of key data quality dimensions that can be used to define performance and acceptance criteria based upon the objectives of the project. The Using Data Quality Dimensions to Assess and Manage Data Quality subtopic sheet details how these dimensions can be defined and carried throughout the data and project lifecycles. Additionally, an overview can be found in Section 4 of this document, and Table 2 lists data quality considerations and links related to the ITRC EDM topics.

Table 2. Considerations when identifying project needs

Topic	Key considerations when identifying project needs
Data Governance and Planning	Develop project and sample planning documents; ask: What is required and to what extent?; define communications guidelines and document data management procedures.
Field Data Collection	Ensure that sample plan details and SOPs are clear and well-understood; establish necessary data types; define sample nomenclature, quality control (QC) sample collection types and frequency, locational details including units, field documentation needs or mobile collection device training and use clear, laboratory needs and shipping requirements and communication procedures
Geospatial Data	Understand standard formats and nomenclature/metadata; clarify coordinates and horizontal and vertical datums used, as well as resolution requirements, storage and data management/version control; ensure understanding of project collection plans and SOPs
Traditional Ecological Knowledge	Understand and align goals and formats; review and vet stories for use; define organizational structure for unstructured data types as applicable, version control
Data Exchange	Determine analytical data needs, electronic data deliverable (EDD) formatting requirements (including input schema definition, valid values, etc.) and metadata documentation procedures
Public Communications	Identify formatting requirements, data sharing rules and data visualization needs

4. DATA QUALITY DIMENSIONS AND USABILITY

The term "data quality dimension" has been widely used to describe the measure of the quality of data. However, even among data quality professionals, the key data quality dimensions are not universally agreed upon. Data management professionals use the term "data quality dimension" to describe a characteristic or attribute of data that can be measured or assessed against defined criteria to determine the quality of data (DAMA UK Working Group 2013)

The ITRC EDM Best Practices Team defined a set of key data dimensions that direct and address a specific aspect of data quality. These data dimensions are shown in Figure 4.



Figure 4. Categories of key data dimensions.

Understanding how each of these data dimensions affects your data activities and data management will allow for proper planning of your project activities and a framework in which to measure, assess, and improve your data. Identifying how these data dimensions can be applied to the data in your project, from collection of field samples, to geospatial data, to your analytical program and data review and usability, understanding the data dimension is a powerful tool in obtaining data quality. To that end, the ITRC EDM Best Practices Team developed the subtopic sheet Using Data Quality Dimensions to Assess and Manage Data Quality, as well as an associated Data Dimensions Matrix. These provide further detail in understanding and applying these data dimension attributes to your environmental data. Additionally, understanding the impact of your analytical data and assessing its usability is discussed further in the Analytical Data Quality Review: Verification, Validation, and Usability Fact Sheet.

5. THE COST OF DATA QUALITY

According to *Forbes Magazine,* "Organizations believe poor data quality to be responsible for an average of $15 million per year in losses. Gartner[1] also found that nearly 60% of those surveyed didn't know how much bad data costs their businesses because they do not measure it in the first place" (Bansal 2021).

Organizations often understand the correlation between business process, or lack thereof, and overall organizational quality. This concept also holds true in the specific case of environmental data quality. The Gartner, Inc., quote above highlights the important cost implications associated with data quality and the importance of preparation to address it. The following text provides examples of considerations for cost and budget that aim to make conversations on this topic easier.

[1] Gartner, Inc., is a technological research and consulting firm that conducts research on technology and shares this research through private consulting as well as executive programs and conferences.

5.1 Cost Considerations

Data quality costs can be identified by two main aspects: the costs incurred in association with assuring good data quality and the costs associated with poor data quality. Both are equally important to understand. The cost of good data quality is shown by incurred costs related to appropriate data quality being met and implemented. These are beneficial costs typically incurred in the scoping and planning stages. In Section 4 of this document, we presented a set of data dimensions to help with identifying areas of data quality to assess and manage in your data set. Understanding these data quality dimensions (see the Using Data Quality Dimensions to Assess and Manage Data Quality subtopic sheet) will help to ensure data quality and minimize costs associated with poor data quality planning. Although costs associated with a lack of data quality are harder to understand, their impacts can be seen in areas such as loss of productivity from poor decision-making, time spent

correcting errors, lessened user satisfaction and user confidence, and reputation concerns. The effect on overall project, program, organizational, and opportunity costs highlights the need to address data quality head on and throughout the project and data lifecycles. Issues can parlay into bottom-line costs as data managers or users spend valuable time and resources working to identify and fix data quality problems. In essence, the effects of bad data are costly!

5.2 Budgeting Considerations in Obtaining Quality Data

Data managers have long understood the effects and implications of data quality issues and ensuring data quality; however, these concerns are not always well communicated with decision makers. Identifying the pitfalls of poor data quality and the benefits that good data quality provide can make this an easier conversation.

Although there is not a way to address all budgeting considerations specifically in this document, there are some key types of activities to remember to allow time and budget for when it comes to data quality. Time and effort applied in the planning and defining of key data criteria and needs are well worth the budget. Allowing for the time and budget in the front end of your project, program, or data effort can prevent later and potentially most costly issues. Table 3 outlines a few key considerations.

Table 3. Tasks to consider when budgeting for data quality

Project Stage	Activities	Key Data Quality Task(s)
Developing scope and planning documents	Allow for sufficient time and budget to complete scoping and planning documents. Identify and inventory all data needs and activities.	Produce quality documentation, e.g., QAPP, DMP, SAP; ensure DQOs have been clearly identified, communicate with project team.
Developing scope and planning documents	Allow time for input by a data manager so that key details can be addressed if not previously scoped.	Receive input from data managers.
Developing the data framework needs	If required, allow time and cost associated with the development of a database framework.	Set up database framework.
Executing the project	Allow time for communication and coordination with field staff.	Allow time for field coordination and tracking and review of sampling procedures, including required QC samples and their frequency, and understanding of QAPP specifications. Allow time to ensure correct and complete documentation.
Executing the project	Allow time for communication and coordination with laboratory staff.	Ask the key questions when initiating the laboratory; include tracking and status, communication, and coordination steps.
Evaluating the data	Understand the type of data review required, who will perform it, and how it will be applied.	Determine how data review will be conducted, consider that this should be by an impartial party, allow time for a usability review.
Monitoring the project/data	Allow time and resources for the monitoring of the project/data, including any adaptive management activities.	This could include the monitoring of data collections, as well as protocols and data quality through the data and project lifecycles.
Closing out the project	Apply corrective action and continuous improvement procedures as needed.	Complete all documentation requirements (for example, metadata, data dictionary), review storage and retention policies, determine lessons learned, and allow time to apply to future processes.

6. RE-EVALUATING THE OVERARCHING DATA QUALITY QUESTIONS

As you enter the “Close” stage of the project lifecycle, your project team is concurrently entering the “Publish/Share” and “Retain” stages of the data lifecycle. As Figure 2 illustrates, the project and data lifecycles can be an iterative series of processes and working through the final stages gives your project team the opportunity for adaptive management and data quality improvements. Because it is expected that your data will persist beyond the timeframe of the project lifecycle as data are published, shared, and retained, it is important that your requirements for data quality, or DQOs, have been met. In other words, at these stages of your project it is important to re-evaluate your initial overarching data quality questions. Are the data you acquired the right type and quality for your project? Did you convey the information needed to the audience who needs it?

For example, consider that you are a citizen scientist collecting data about water quality in a local stream. Your audience is a citizens’ environmental group with interest in water quality parameters potentially deleterious to fish populations. The data are collected using basic field meters or test strips during the same month at roughly the same stream location, recorded on paper forms, and shared with the group to inform the community about general trends. These methods have seemed to be sufficient but given new knowledge of state regulatory criteria for fish in streams, the data seem to be trending in the wrong direction. Several new questions then arise: At what stream locations and date/time were the values recorded, and what methods for measurement were used? For comparison to ecological criteria, were the necessary precision/significant figures or reporting levels met by field measurements, or is a different analytical program required? Would there be a need to determine statistically significant trends or correlations with other stream attributes for fish health? Both adaptive management and data quality improvements as part of the iterative project and data lifecycles may help answer these questions.

The preceding example is just one of many for environmental data collection activities, and, as Figure 1 illustrates, the answers to the overarching data quality questions will vary for different projects and intended uses of the data. Regardless of experience or professions of the project team, the path to acceptable data quality should not be taken without careful data quality planning.

In this data quality overview document, we have presented overarching data quality questions, considerations for planning and planning documentation, data quality dimensions for each aspect of the data lifecycle, the cost of data quality, and how each of these fits into the project and data lifecycles. These guiding concepts, along with planning and re-evaluating the overarching data quality questions, will help your team achieve the data quality necessary for a successful project.

7. REFERENCES AND ACRONYMS

The references cited in this fact sheet, and the other ITRC EDM Best Practices fact sheets, are included in one combined list that is available on the ITRC web site. The combined acronyms list is also available on the ITRC web site.