

# 2021 ITRC Final Proposal Template

## Proposed Project Title

Environmental Data Management Best Practices

## Proposal Contacts

### Brian Pointer

Environmental Program Consultant  
North Carolina Department of Environmental Quality  
Email: [brian.pointer@ncdenr.gov](mailto:brian.pointer@ncdenr.gov)  
Phone: 919-743-8474

### Meghan Eschbaugh

Senior Engineer  
Environmental Resources Management  
Email: [meghan.eschbaugh@erm.com](mailto:meghan.eschbaugh@erm.com)  
Phone: 317-706-2014

## Abstract

This proposed ITRC team will develop Environmental Data Management Best Practices that can be applied across the wide spectrum of environmental data. The best practices developed can serve as the foundation upon which data analyses, visualization and sharing can build. Hand in hand with these best practices the ITRC team will develop tools to aid in data management decision making, on-line training to help implement these best practices, and a community forum to serve as a platform for continued sharing of best practices as available technologies continue to evolve.

## Problem Statement and Highlight the Importance to the States and to the Broader Environmental Community

### Problem statement

The U.S. spends billions of dollars each year planning, collecting, and analyzing environmental monitoring data. The volume of data increases as we address more sites, discover new contaminants of concern, and use automated monitoring tools, including mobile devices. Concurrently, industry stakeholders, consultants, and government regulators are pushing for real-time decision making based on these data sets.

However, there is a lack of standard and consistent data management practices necessary to ensure high quality, rapidly available data and minimize the risks involved in making costly decisions on potentially bad or incomplete data. Data management practices across the industry are often reactive and targeted, rather than planned and strategic. Data management tools and practices vary widely across the environmental industry, from mature, standardized applications and practices, to spreadsheets and ad-hoc data manipulation. There is currently a patchwork of guidelines created by regulators and consultants, but data interoperability remains difficult and cumbersome. In addition, much of the existing guidance is in dire need of updating to match new technology and techniques. Undoubtedly, the environmental community – regulators, the regulated community and the broader public - would benefit from a clear set of Data Management Best Practice guides for Environmental Data in the modern era.

This proposed ITRC team intends to focus on **Environmental Data Management Best Practices** because good data management serves as the foundation for all subsequent data analyses, visualization, and sharing. Several nation-wide initiatives are heavily focused on the use of application program interfaces (APIs) and data analysis techniques to integrate and visualize environmental data. These are all fantastic tools that deserve to be utilized broadly to help better understand complex environmental issues. However, several states, tribes and citizen science groups are still struggling with the fundamentals of just managing the growing body of data that they collect. The idea of setting up an API is overwhelming and may be unrealistic to states without a modern data management system, compounded by shrinking budgets and limited staff. Our intent is to create a toolkit that focuses on the fundamentals of data management so that organizations that are still working with PDFs (or pen-and-paper!) can start along the path to more modern and efficient data management and begin to more readily utilize the tools of modern data analysis. States and organizations with well-developed data management systems can also benefit, since this is an ever growing and evolving field.

As one example, North Carolina is currently transitioning to better data management after years with a lack of unified strategy. North Carolina, like many other states, suffers from far too many 'data silos'. Each study group has its own homegrown data system – mostly consisting of spreadsheets, PDFs, and in some instances, just a bank of filing cabinets). None of these disparate systems share a common data format or valid values, and sharing data among systems is so difficult to be practically impossible. Processes for validating and exporting data to central services such as WQX or ATTAINS are time consuming and take weeks to months to complete. Clearly processes needed improvement, and luckily leadership took notice and got behind an effort to modernize data processes around one cohesive system. This process is underway now, but it has been guided by a set of principles laid out in a set of white papers produced by the International Conference for Environmental Data Management (ICEDM). These white papers are the body of work that this proposed ITRC team intends to build upon to become a basis for best practices that any environmental data gathering organization can utilize.

Another example that clearly demonstrates the need for good environmental data management - and the cost of poor data management - is the 2016 U.S. Geological Survey (USGS) study "[Challenges with Secondary Use of Multi-Source Water-Quality Data in the United States](#)" (Sprague, Lori A., et al. Water Research, vol. 110, 2017, pp. 252-261). The study addressed challenges USGS encountered when assimilating nutrient data from numerous systems across the U.S. They thought that using existing data could help counterbalance diminishing resources for stream monitoring and lead to regional and national insights. The study surveyed more than 25 million nutrient records collected by 488 organizations in the U.S. since 1899. However, they found that nearly 14.5 million of the records (57%) had missing or ambiguous information for one or more key metadata elements. The cost of these ambiguous data was estimated at \$12 billion. As a result, USGS partnered with U.S. EPA and several states to form a nutrient data management best practices workgroup. The data management best practices they developed will ultimately facilitate better data analyses, visualization, and most importantly, data sharing and reusability. This workgroup shows the need for data management best practices and has served as inspiration for the establishment of this proposed ITRC team.

This proposed ITRC team does not intend to replace or supplant other efforts encouraging data management, but, conversely, hopes to be more of a platform for integration and cohesion across environmental disciplines and programs. The data management best practices developed can be applied across the wide spectrum of environmental data. We hope to engage with E-Enterprise, U.S. EPA, States, private sector entities, citizen science groups, and others to encourage the quality data management

practices that should be the backbone of any environmental monitoring program. The deliverables that this Team proposes should be able to support and supplement existing efforts by ECOS, E-Enterprise, Internet of Water and others. We have engaged in discussions with several of these groups while developing this proposal and hope to maintain a collaborative relationship throughout the project period (and beyond).

A strategic component of this engagement will be with E-Enterprise for the Environment and the Exchange Network. The work of the proposed ITRC team dovetails with the work of the Exchange Network. The Exchange Network seeks to make data sharing easier, leading to better decision-making. This ITRC team will help establish the foundational data management needed to facilitate that data sharing.

## **Project Summary**

Topics to be covered in ITRC data management best-practice fact sheets will include:

- Planning, developing, and implementing data governance across environmental data collection, and analysis efforts including a [data management tool selection matrix](#).
- Drafting, maintaining, and adhering to a Data Management Plan (DMP).
- Determining minimum laboratory electronic data deliverable (EDD) structure requirements and standards to guide the creation of analytical laboratory EDDs, with the goal of producing an EDD structure that makes data use more efficient, thus saving time and money.
- Managing valid values in an environmental data management system (EDMS).
- Strategies for management and migration of historical project data.
- Field data collection and management for field staff and data managers including best practices for field notes (paper and digital), as well as graphical/pictorial data collection.
- Training and Communications
- ‘Big Data’ Management
- Data Security and Integration
- Case studies

In addition, the team will develop online training videos to accompany the publication of the guidance document. The videos will be developed in modules of approximately 30 minutes in length and be paired with the content of the best management practice fact sheets.

## **Project Deliverables**

The new team would begin in January 2021. The team will initially research current data governance practices used by regulators and industry. The team will then work on development of the fact sheets and case studies. Upon completion of the documents, the online training videos will be produced. The team would plan to complete the work by December 2023.

ITRC data management products to be produced:

- Best management practice fact sheets – see above

- Case studies
- Online training videos
- Web forum for ongoing discussion of changes in available tools and technology
- Tool selection matrix

### **Additional Information**

The team would be composed of any interested academia and regulatory and regulated entities including: state and federal agencies, tribes, consultants and industry.

#### State Team Leads:

Brian Pointer – NC Department of Environmental Quality

#### State Agencies:

Washington Department of Ecology, Chris Neumiller

North Carolina Department of Environmental Quality, Brian Pointer

Kansas Department of Health and Environment, Dr. Elizabeth Smith

New Mexico Bureau of Geology & Mineral Resources, Stacy Timmons

Minnesota Pollution Control Agency, David Vaaler

#### Local Agencies:

City of Seattle, Jennifer Arthur

#### Federal Agencies:

#### Academia:

Internet of Water @ Duke University, Dr. Ashley Ward, Dr. Lauren Patterson

#### Consultants:

Environmental Resources Management, Meghan Eschbaugh

#### Professional Organizations:

International Conference of Environmental Data Management (ICEDM)

#### Industry and Subject Matter Experts - Interested Parties:

David Myers, Environmental IT Manager – Dow Chemical