Data Management: It's Fundamental

A case study of managing environmental reporting data in Antarctica

Sadie Rusby

Advisor: Dr. Andrew Read

April 2022

Masters project submitted in partial fulfillment of the requirements for the Duke Environmental Leadership, Master of Environmental Management degree in the Nicholas School of the Environment of Duke University.

EXECUTIVE SUMMARY

This master's project leverages process improvement methodologies to build a case for developing standardized and streamlined data systems for analyzing the environmental impacts of the U.S. Antarctic Program (USAP). As one of fifty-four signatories to the Antarctic Treaty, the United States has a responsibility to monitor environmental impacts of all activities associated with conducting research on the coldest and harshest continent in the world. Scientific and support activities affect the environment due to the presence of people in Antarctica and their use of machinery, vehicles and infrastructure.

The regulatory drivers for protecting the Antarctic environment began in 1991 with the enactment of the Environmental Protocol to the Antarctic Treaty; and for the United States in 1996, with its adoption to U.S. law the Antarctic Science and Tourism Conservation Act. Over the last three decades, the amount of data required for environmental reporting has increased, but the associated tools and resources needed to process and manage data have not evolved at the same rate and are now outdated. As a result, most of the data collected over the years are not accessible or saved in useful formats.

The environmental engineering team of the Antarctic Support Contract (ASC), a subcontracting organization for National Science Foundation (NSF), monitors USAP activities, collects data and reports these to NSF. Effective environmental monitoring provides awareness of the adverse impacts of activities and an environmental monitoring program with data systems in place could assess accumulative impacts. However, without systems in place to produce consistent and accurate data, environmental impact analyses cannot be done in a meaningful way.

The purpose of my master's project is to implement process improvements to better organize and streamline data management. In October 2020, I started by investigating USAP environmental reporting data lifecycles, and then identified process improvements with objectives to: 1) produce higher quality and accessible data for analysis; and 2) to reduce resources required to manage data by streamlining procedures. This paper discusses data improvement projects, which are currently ongoing. These projects included three main phases—data standardization, improving data management, and improving data reporting. The time and resources needed to evaluate data management processes and data lifecycles are not insignificant, but the outcomes of these projects are already resulting in reduced processing time and touchpoints, and improved data quality and reporting.

Standardizing environmental monitoring data collection and data management will lead to accessible and accurate data for analysis in the future. Data collected by ASC could be used to assess USAP activities to inform decisions on how to reduce USAP environmental impacts. Currently, however, these data are not accessible, organized or clearly understood. With clean, relevant, and reliable data at the foundation of our analyses, we can better set and meet goals of reducing environmental impacts in the US Antarctic Program.

Table of Contents

Data Management: It's Fundamental

A case study of managing environmental reporting data in Antarctica

INTRODUCTION

United States Antarctic Program

Americans have had a continuous presence in Antarctica since 1955 when the US military began constructing McMurdo and South Pole research stations. Science teams and support staff make up the United States Antarctic Program (USAP), which upholds the Nation's goals of "supporting the Antarctic Treaty, fostering cooperative research with other nations, protecting the Antarctic environment, and developing measures to ensure only equitable and wise use of resources" (NSF, 2021). Today, approximately 3,000 people travel to Antarctica every year to work at USAP facilities, which are comprised of three year-round stations, two scientific research vessels, and numerous seasonal field camps. While the mission is to conduct important scientific research, getting scientists and scientific equipment to Antarctica involves an enormous amount of support including coordination, materials, and operations to run remote research stations. This includes infrastructure to house scientists and support staff, runways for airplanes, and heavy equipment for moving cargo and people, and maintaining roads.

Environmental Impacts

Wherever humans are, so are machines and waste materials. Machines (e.g., airplanes, ships, trucks, buses, loaders, snowmobiles, and generators) rely on petroleum and other chemical materials (e.g., fuel, oil, glycol) to operate, and these materials inevitably get into the environment. Additionally, all waste (e.g., solid, liquid, chemical, hazardous) must be managed so it does not end up in the environment. This is a concern of our industrial and modern world anywhere, but especially in remote research stations in Antarctica. Furthermore, anything sent down to Antarctica, from food to scientific equipment, must be managed carefully as to not end up unintentionally in the environment (i.e., an environmental release).

Environmental impacts are the effects of human activities on the surrounding environment, and Antarctic Support Contract (ASC) collects data to monitor and measure USAP activities. These data are reported to NSF on an annual basis to meet environmental stewardship obligations for the Antarctic Treaty and in compliance with Antarctic Conservation Act (ACA) regulations. A large part of monitoring environmental impact is through data collected of environmental releases, such as emissions, accidental spills, or effluent from wastewater treatment plants. Other data collected include locations of science activities and camps, waste generated and removed from Antarctica, fuel usage, number of people and where they went, and the number of missions of ships, helicopters, and planes, to name a few.

Reducing Environmental Impacts through Better Data Management

To gain insights on what USAP environmental impacts are and how to reduce them, NSF needs clean, relevant, and reliable data. Without data systems in place to produce consistent and accurate data, environmental impact analysis cannot be done in a meaningful way. Every year, resources (e.g., time, labor, financial) are used to collect data to meet compliance and reporting requirements. In addition, data and data analysis are expected to inform management decisions; however, this is an unreasonable outcome without having quality data systems in place that ensure data are accurate and complete.

BACKGROUND

Environmental Reporting Data

Since the enactment of the Environmental Protocol to the Antarctic Treaty in 1991 and its adoption to U.S. law as the Antarctic Science and Tourism Conservation Act in 1996, USAP collects data and reports activities that affect the Antarctic environment. Over the last three decades, the amount of data required for environmental reporting has increased, however, the associated tools and resources needed to process and manage data have not evolved at the same rate, thus are now outdated. ASC Environmental Engineering compiles data from dozens of sources and is required to report different data to NSF monthly, biannually, and annually.

This project evaluates the current state of USAP environmental reporting, identifies barriers to better data management, implements internal process improvements within Environmental Engineering, and identifies additional recommendations for modernizing and streamlining USAP environmental reporting. USAP environmental reporting is fragmented, labor intensive, and involves redundant requests to data owners. Additionally, without proper quality assurance and quality control procedures (QA/QC) in place, data quality is unreliable. Throughout this paper, data and data management are discussed and these are related to specific data collected, managed, and reported by ASC; these data are required by and outlined in the Antarctic Treaty and ACA.

Scientific Research Data

Americans have been conducting scientific research in Antarctica since 1830 and have had a continuous presence in Antarctica since 1955 when the US military started building McMurdo and South Pole stations. NSF scientific goals in Antarctica are:

- “to understand the Antarctic and its associated ecosystems;
- to understand the region's effects on, and responses to, global processes such as climate;
- and to use Antarctica's unique features for scientific research that cannot be done as well elsewhere” (NSF, 2021).

NSF provides funding for science teams and USAP supports scientific research through coordination of Antarctic-based fieldwork when presence in Antarctica is required for research purposes. NSF funds research in many different disciplines including astrophysics, biology and medicine, geology and geophysics, glaciology, and ocean and climate systems.

Data collected and studied by NSF funded science teams are not discussed in this paper. However, related to the topic of this paper, science teams must include a data management plan (DMP), as part of their grant proposals describing, “how the project will provide open access to quality-controlled and fully documented data and information during and after the project completion” (OPP, 2016).

Antarctic Treaty and Environmental Protection

The Antarctic Treaty (Treaty) provides overall provisions for all activities conducted south of 60° South latitude, including islands, the Southern Ocean, and the continent of Antarctica. Twelve countries signed the Treaty in 1959; it has since been acceded by 54 countries, which are Parties of the Treaty. The Protocol of Environmental Protection to the Antarctic Treaty (Protocol) provides environmental protection and conservation guidance for Antarctic Treaty Nations. Article 2 of the Protocol stipulates that, “the Parties commit themselves to the comprehensive protection of the Antarctic environment and dependent and associated ecosystems and hereby designate Antarctica as a natural reserve, devoted to peace and science” (Protocol, Artcl. 2, 1998).

The Protocol's Environmental Principles (Article 3) specifies that Treaty nations should monitor, plan, and conduct research and operations to limit adverse impacts on the Antarctic environment. Treaty nations should conduct "regular and effective monitoring" to assess the "impacts of ongoing activities" including evaluating accumulated impacts (Protocol, Artcl. 3, 1998). Monitoring will enable early detection of negative impacts and should provide sufficient information needed to assess and make informed judgments about effects on the Antarctic Environment. Furthermore, if research or operational activities will "result in, or threaten to result in, impacts upon the Antarctic environment", they should "be modified, suspended or cancelled" (Protocol, Artcl. 3, 1998).

Antarctic Conservation Act

The Antarctic Conservation Act (ACA)[1] is a U.S. Federal law that "conserves and protects native mammals, birds and plants of Antarctica and the ecosystems they live in" (NSF, 2001). The law applies to the area south of 60° South latitude, U.S. citizens in Antarctica, participants of U.S. government activities or U.S. corporations in Antarctica, and anyone importing Antarctic animals and plants to the U.S. for scientific purposes. The ACA also regulates waste management, designates banned substances and pollutants, and governs Antarctic Specially Protected Areas (ASPA) and associated ASPA management plans. Additionally, a permit system authorized by the ACA, and administered by NSF, authorizes an individual or organization to conduct activities for scientific or other purposes normally prohibited by the law (ACA, 45 C.F.R. Part 671, 1993).

[1] The Antarctic Conservation Act of 1978, Public Law 95-541, as amended by the Antarctic Science, Tourism, and Conservation Act of 1996, Public Law 104-227.

All USAP activities in Antarctica are covered by a Master Permit which "establishes standards for management of all designated pollutants and wastes, including requirements for removal", and reporting of environmental releases (NSF, 2001). USAP participants are required to follow procedures outlined in "the Master Permit for handling, inventorying, storing, monitoring, and disposing of wastes" (NSF, 2001). The ACA provides a framework for tracking and reporting USAP wastes and environmental releases. Waste management standards for the USAP are required as stated in ACA § 671.13, "to provide a basis for tracking USAP wastes, and to facilitate studies aimed at evaluating the environmental impacts of scientific activity and logistic support" (ACA, 45 C.F.R. 671, 1993).

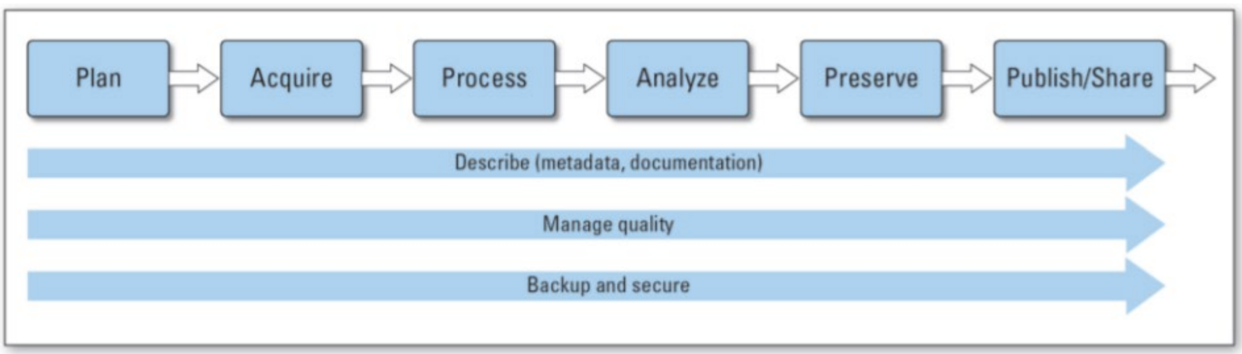
What is Data Management?

"Data management is the practice of collecting, keeping, and using data securely, efficiently, and cost-effectively. The goal of data management is to help people, organizations, and connected things optimize the use of data within the bounds of policy and regulation so that they can make decisions and take actions that maximize the benefit to the organization." (Oracle, 2022)

Data management works to organize and maintain information through developed standards to ensure data is accurate, consistent, accessible and stored securely. Understanding data, through its entire data lifecycle, enables the ability for effective analysis; "without good data management, analysis is practically impossible and unreliable at best" (Vigliarolo, 2019). Data lifecycle models vary by industry and are high-level views of data that consist of phases or stages, from origination through preservation and usage, which guide workflows and data management expectations. They are considered a cycle due to the iterative process of managing data where insights and lessons learned can inform future data projects or initiatives.

This project used the framework outlined by Faundeen et al. in their report *The United States Geological Survey (USGS) Science Data Lifecycle Model*, which includes the data management stages: plan, acquire, process, analyze, preserve and publish/share (2014). Importantly, they state, "the model serves as a structure to evaluate and improve policies and practices for managing data and to identify areas in which new tools and standards are needed" (Faundeen et al., 2014). More details on primary components of each stage of the USGS Science Data Lifecycle Model framework can be found in Appendix 1.

Figure 1. USGS Science Data Lifecycle Model



Furthermore, data must be clean, relevant, and reliable to be used for analysis as defined here:

- <u>Clean data</u> is standardized data; data that are uniform and appropriately labeled and organized. These data are ready for analysis and no additional work is needed to use.
- <u>Relevant data</u> are data associated with the specific issues or problems being measured. Additionally, data being collected unnecessarily or for an unknown future purpose is not considered relevant data and would be a waste of resources.
- <u>Reliable data</u> is when the entire data lifecycle is understood and documented, and data are accessible for use (i.e., not in multiple spreadsheets, document tables or PDF reports); and all this information should be documented in a regularly reviewed Data Management Plan (DMP).

My Interest in Data Management & Process Improvement

As part of most of the jobs I have had in the last 15 years, I have managed data without being a trained data professional. What I have gleaned over the years is that data is everywhere, but in many cases, it is managed poorly. I have held mostly contract or limited-term positions through various employers, including a private business, a state university, and multiple federal agencies and government contracting companies. People in these organizations do their best to manage data with limited tools and resources. My approach has always been to streamline and standardize processes with the idea that I should leave the job better than I found it. To improve my skills, I have taken process improvement courses (Lean, Six-Sigma) and used what I learned to coordinated a long-term Lean process improvement project for a previous employer. The skills and tools learned from these experiences helped me with the work conducted for this master's project.

METHODS

Data Management for Environmental Reporting

For this master's project, I investigate data lifecycles used for environmental reporting to build a case for developing standardized and streamlined data systems that are fundamental for analyzing environmental impact. This includes evaluating internal and external processes of ASC departments, NSF, and a subcontractor to understand all aspects of environmental reporting data. This project resulted in three main phases:

1. *Data Standardization* – Evaluated and standardized two data sets managed by Environmental Engineering: Remotely Deployed Equipment (RDE) inventory and Spill Report data.

2. *Improving Data Management* - Investigated lifecycles of RDE and Spill Report data to identify data management improvements to increase quality and usability of data.
    a. RDE inventory - the format of the environmental end of season (EOS) form was updated to enable automated data uploads and to standardized all data collected in the form; RDE inventory data is one of several datasets collected by the EOS form.
    b. Spill Report data - collaborated with stakeholders to develop an online data collection form to collect standard data at the beginning of the spill reporting process instead of the middle.
3. *Improving Data Reporting* - compiled a centralized list of data required for annual reporting to NSF, which includes data sources, owners, and purposes. It also included creating a SharePoint site to manage data requests and data submittal for reporting.

Investigating Environmental Reporting Data

As part of the data investigations, I started developing a DMP for USAP environmental monitoring data. A DMP includes evaluating and defining every aspect of datasets including scope, metrics, governance, organization, roles, processes, and technology. The purpose of Environmental Engineering DMP is to identify, define and describe data, data sources and information used for USAP environmental reporting. The data lifecycle investigations throughout this project identified data sources, defined requirements for current and future data, and identified sensitive data where governance (security) measures need to be established.

The many distinct types of data, complexity of data collection, and varied purposes for USAP environmental reporting, made developing a DMP a project in itself. The number of data sources range from 70-100+ depending on the USAP activities each year, and many have varying collection and management procedures. The DMPTool.org[2] website and templates provided the basic outline for sections and guiding questions used for the Environmental Engineering DMP development. These sections and associated questions are:

1. Data Summary
    a. What are USAP environmental monitoring and reporting data?
2. Data Sources, Collection & Management
    a. Where does USAP environmental monitoring data come from?
    b. How are data collected and managed for environmental monitoring?
3. Existing Data
    a. What data already exists and where; and how are these integrated with future use or data analysis?
4. Data Formats and Standards
    a. What type/formats of data are used and what formats do we need?
5. Dissemination, Access, and Sharing Data
    a. How are data shared and with whom?
    b. What levels of access or restrictions are needed?
6. Archiving Data
    a. Where are data stored, short-term or long-term?
    b. Are there size constraints for storage?

When complete, the Environmental Engineering DMP will follow best practices, which will document all aspects of every data lifecycle. The DMP will be a living document to be reviewed annually and updated regularly as needed.

[2] The DMPTool is a free, open-source, online application that helps researchers create data management plans. It also has help text for answering questions, and data management best practices resources. (DMPTool, 2022)

ASC Environmental Engineering Data Collection and Management

As part of coursework conducted for this graduate program, I developed a strategic plan for Environmental Engineering. In the development of the strategic plan, I evaluated Environmental Engineering work structure and tasking and identified a substantial portion of Environmental Engineering tasking that are data driven (i.e., tasks that require data collection and data management). Overall, work structure and tasking fall under three main categories: Monitoring, Stewardship and Reporting.

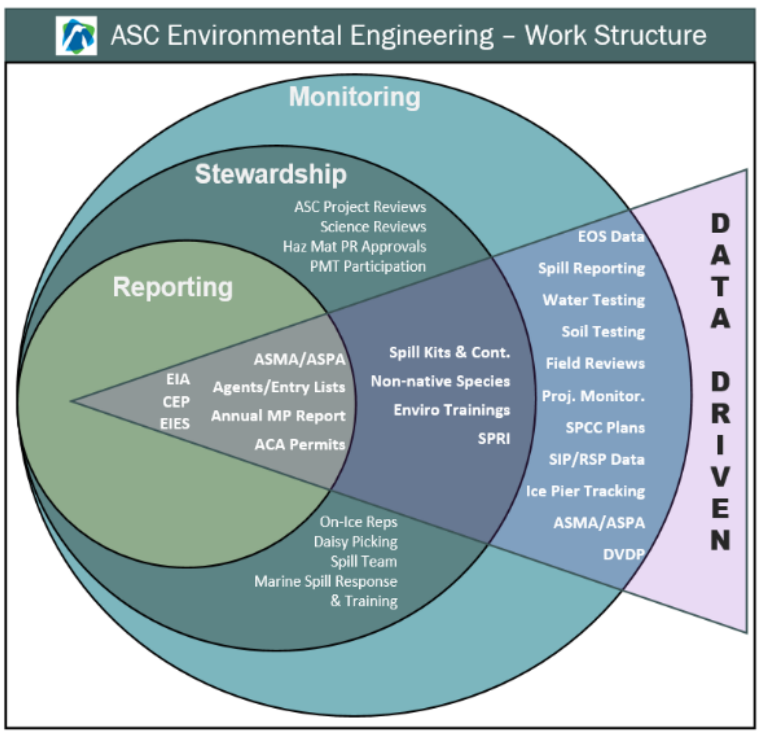


Figure 2. ASC Environmental Engineering work structure is made up of three categories: Monitoring, Stewardship and Reporting. Primary tasks are listed in each category. The purple triangle identifies all tasks that involve managing data.

Environmental Engineering handles data in every phase of the data lifecycle; however, Environmental Engineering does not manage most data used for environmental reporting throughout the entirety of each lifecycle (i.e., from beginning to end). Data handled by

Environmental Engineering fall into three main categories, 1) Data collected and managed by Environmental Engineering; 2) Data collected by others, but managed by Environmental Engineering; 3) Data owned by others, but gathered by Environmental Engineering for environmental reporting.

Each of these categories have differing controls or quality checks in place that lead to variable levels of data accuracy or completeness.

1. Data collected and managed by Environmental Engineering have established standard procedures in place, including essential control measures and quality reviews. This leads to high certainty of accurate and complete data.
2. Data collected by others, but managed by Environmental Engineering have varying levels of established standard procedures, limited control measures in place or quality reviews conducted. This leads to some uncertainty of the accuracy or completeness of the data. In this category, accuracy of this data depends on meaningful engagement of others, clearly defined instructions and expectations, and the practical use of data collection tools. Additionally, these data require more resources to coordinate its collection and review when compared to data collected and managed by Environmental Engineering alone.
3. Finally, for data owned by others, but gathered by Environmental Engineering for environmental reporting, it is difficult to assess if controls are in place or quality reviews are conducted, and assumptions must be made that the data is accurate and complete. A key factor to consider when working with data owners is to clearly communicate data

requirements and data purposes for environmental reporting. This clarity will lead to more thorough data collection and better quality data.

Table 1. ASC Environmental Engineering (ASC EE) has different levels of involvement with environmental reporting data. Data examples sorted by who collects and manages data, and the reporting requirements.

Data		Collected and Managed by ASC EE	Collected by Others; Managed by ASC EE	Owned by Others; Gathered by ASC EE for Reporting	Reporting Required by ACA/NSF	Reporting Required by Treaty
ASC EE Data	Drinking Water Testing	X	X		X	
	Wastewater Testing	X	X		X	
	Non-Permitted Releases	X			X	
	Non-Native Species	X			X	X
	Environmental Training	X			X	
	Field Camp Reviews	X			X	
	EIA Reviews	X			X	X
USAP Field Only Data	Field Activities & Locations		X		X	
	Environmental Disturbances		X		X	
	Environmental Releases (physical & acoustic)		X		X	
	Waste Generation		X		X	
	Wastewater Discharge		X		X	
	Fuel Usage		X		X	
	UAS Flights		X		X	
	Deployed Equipment Inventory		X		X	
	Field Cache Inventory			X	X	
USAP Operational Data	Environmental Releases		X		X	
	Spill Reporting		X		X	
	SPCC Plans			X	X	
	Population			X	X	X
	Wastes Generated (liquid & solid)			X	X	
	Hazardous Waste			X	X	X
	Aircraft Missions & Hours			X	X	X
	Fuel Usage			X	X	
	Wastewater discharged			X	X	
	Radioisotope Use			X	X	

As the coordinator of USAP's annual environmental reporting to NSF, Environmental Engineering is in a unique position within ASC to handle data from many sources across USAP. Data comes from ASC personnel and science teams who make up many stakeholders for environmental reporting. ASC departmental stakeholders include station work centers such as, Facilities, Utilities (Water and Wastewater plants), Operations (Waste, Fuels and Fleet Operations), and field teams who have activities outside a station footprint. Examples of Antarctic environmental impact data include GPS locations of all USAP activities, amount and type of fuel used, amount and type of waste generated and disposal methods, emissions generated, material released to the environment, accidental spills, station and camp populations, and number of helicopter and airplane flights and landing locations.

RESULTS

Finding Efficiencies through Better Data Management

The two Environmental Engineering data process improvement projects listed below started in October 2020 and are continuing through to today. These improvement projects involved standardizing data, streamlining data collection, and centralizing data management and organization. The time and resources needed to evaluate data management processes and data lifecycles are significant, but the outcomes of these projects are already resulting in reduced processing time and touchpoints, and improved data quality and reporting.

Spill Report Data Management

Problem Description

Spill reporting is required by the ACA to track environmental releases and measure USAP's impact on the environment. Spill report data are collected when accidental spills of designated pollutants (fuel, oil, glycol, laboratory chemicals, and sewage) occur in the Antarctic environment (water, soil, rock, ice, and snow). The spill report data collection process is inefficient and involves redundant touchpoints. Additionally, there are limited quality controls in place leading to poor data quality lacking standardization. Monthly reporting to NSF involves manually updating Excel charts. Overall, this process takes an average of 25-30 labor hours per year to produce data for limited summary reports of total spill counts, total spill volume, and median spill volume.

Without standardized and reliable data, monthly and annual summary report capabilities are limited; and without extensive data processing, more dynamic data analysis is not possible. Enhanced data analysis could identify areas of focus for spill prevention, reduction actions, and training, or could highlight other risks and opportunities within ASC.

Problem Solving Approach

In the first project phase, Environmental Engineering standardized data by separating free-text entries into multiple categories (e.g., locations, spilled materials, and work centers), researched spills and added missing details, and brainstormed what data was needed for any additional analysis. Next, a Power BI dashboard that includes interactive visualizations was created to replace static Excel charts emailed monthly.

The second phase of the project included collaboration of a cross-functional team of key stakeholders to evaluate how to collect standard and comprehensive data at the beginning of the spill reporting process instead of in the middle of the process. The team included Environmental Engineering, Waste Management, a Business System Analyst and Software Systems Engineering.

The team developed an online spill report data collection form to collect standardized data entries, where data could be used to populate an email template for spill notifications. When finalized, data collected in the online form will be in a centralized database and accessible to

Environmental Engineering to use for analysis and reporting. Currently, the spill response team at McMurdo station is testing the online form; this process improvement is expected to be fully implemented by October 2022.

Project Results

The two phases of the project reduce labor hours needed to manage and report spill data, and increase the data quality and analysis capabilities. With the interim phase of the project complete, labor hours were reduced by 30% from the original process. Even though the data collection process is still manually entered (albeit, now in a standardized way), the Power BI dashboard published on a centralized reporting site allows for automated dashboard updates whenever new data is added. Through better data management standards and with QA/QC procedures in place, data can be used for dynamic spill reporting analysis and higher quality data is now reported to NSF. When this project is complete, it is estimated that Environmental Engineering labor hours for this process will reduce by 63% from the original process.

Field Data Collection Form Update

Problem Description

All field teams that leave the footprint of a permanent station (McMurdo, South Pole, and Palmer) fill out an environmental end of season (EOS) form. In a typical season, ASC staff and grantees return 70-90 EOS forms to Environmental Engineering. At the end of the season, EOS data are compiled manually and reported to NSF; however, accumulated data are not stored for future use. The EOS form is an Excel spreadsheet and data collected is extensive including, but not limited to, locations of field camps, aircraft landings, scientific sampling sites, deployed equipment, type and totals of generated waste, and environmental releases. If EOS data were compiled every year and stored in an accessible format, data could be used to assess accumulated environmental impacts over time. Additionally, not all collected data are reported to NSF; this means data are collected with the intention of future use, but are not being used. A significant amount of time and labor are spent entering data (ASC staff and grantees) and managing EOS forms (Environmental Engineering and a subcontractor). If these data were compiled and stored in a central location, resources to manage data would be reduced for Environmental Engineering and their subcontractor, and value could be added since data would be available for analysis.

Problem Solving Approach

A cross-functional team including Environmental Engineering, Software Systems Engineering, and a Business System Analyst investigated options within current technology and resources to identify data upload and management capabilities within ASC. Barriers to collecting EOS data centrally are 1) the format of the EOS form is not conducive to automated data uploads and, 2) there are limited capabilities within ASC to centrally manage datasets. The POLAR ICE[3] program was identified as a place to upload and manage data since all EOS forms are associated with either a science event or an internal ASC departmental field group in POLAR ICE. However, the program is 20 years old and resources to manage it or make updates are limited.

The team decided that no matter where the data would be stored, updating the EOS form to enable automatic retrieval of data would be beneficial and a partial solution to the problem. Additionally, a thorough review of the data collected by the EOS form for relevance and purpose

[3] “Participant On-Line Antarctic Resource Information Coordination Environment (POLAR ICE) is the web-based software application used to assist in deployment preparation, and coordinate and facilitate support for Antarctic research missions” (USAP, 2022). It was first implemented in 2003.

was an added benefit. Environmental Engineering updated the format of the EOS form by separating data entry sections into separate tabs within the Excel workbook. This allows for EOS form standardization and for data to be accessible and consistent for retrieval through computer programing.

In addition, the EOS form was updated with the end-user in mind. End-users are teams living and working in cold temperatures and extreme weather, without internet connectivity, and using laptops with small screens and varying operating systems. Making the form easy to navigate and understand was a high priority. Finally, data fields in the EOS form were reviewed for their intended purposes to ensure that data being requested were relevant and not superfluous.

Project Results

While automated data uploads are still not possible, the updated EOS form was tested during the 2021-2022 season and feedback was positive from EOS form users, from field teams who entered data, to those who compiled data for end-of-year reporting. In general, field teams entered higher quality data than collected on previous EOS forms. This is due to using data validation formats (e.g., drop down lists, specific cell formatting), clearer instructions and examples, and a planning checklist which, when used, identifies data sections that need to be filled out by the field team.

Unanticipated positive outcome – to ensure the successful implementation of the updated form, during the season the Environmental Specialist met with every field team, before they went into the field, to review the form and answer questions. These meetings also enabled general discussions about environmental reporting and the importance of collecting accurate data. Field teams were encouraged to provide feedback on the use of the new form, which some did. Overall, these interactions were valuable and constructive; furthermore, they demonstrate that building good relationships is an important part of collecting reliable data. Environmental Engineering plans to continue these environmental field team meetings in future seasons.

In conclusion, until data systems are in place to centrally manage EOS data and enable automated data uploads, data will continue to be compiled manually and only used for end of year reporting. While data collected in the updated EOS form is higher quality data, data is still inaccessible and cannot be used for analysis.

DISCUSSION

Data Management Challenges

A significant percentage of Environmental Engineering's tasks are data driven. Current data practices are inefficient and require excessive amounts of time to manage or manipulate data for use, which inhibits the ability to respond nimbly to NSF requests for data. Standard practices for data collection and management are not in place, which leads to data management decisions being implemented without proper planning or understanding the entire data lifecycle. This leads to significant resources being used to manage data inefficiently and to report poor quality data to NSF.

There are hundreds of distinct data being collected, managed, and reported every year. These data have various data collectors, data owners and data managers, and Environmental Engineering coordinates some or all aspects of these data. Data are being collected, but not used effectively, and current data are inaccessible or not in usable formats for data analysis.

Additionally, since there is minimal planning or regular reviews of data, resources are used to collect and report unnecessary data.

USAP environmental reporting data should be used to inform future planning within ASC Environmental Engineering and USAP. An example of a data-informed planning action would be evaluating accidental spill reporting data to inform who is and is not reporting spills. This might lead Environmental Engineering to work with specific departments to investigate if spills are not really occurring, or they are occurring, but not reported. Alternatively, data-informed USAP planning might include developing a strategy to measure and evaluate waste generation and set goals around reducing waste in Antarctica.

Scientific Use of USAP Operational Data

While scientific data is not the subject of this master's project and paper, it should be noted that scientists could use USAP's operational and activity data as part of their research. In fact, a 2016 workshop and report, focusing on the McMurdo Dry Valleys Antarctic Specially Managed Area (ASMA), identified specific recommendations for National Antarctic Programs (e.g., USAP, Antarctica New Zealand) to provide program activity data to the scientific community for incorporation into current and future research studies (Priscu & Howkins, 2016). Additionally, Priscu and Howkins give examples of data requested for public use, "sample and instrument sites, camp locations, landing sites, personnel movements, and environmental incidents" (2016).

The report further recommends that investments be made to inform management decisions and future science planning and to "facilitate consistent documentation, streamlined analysis and easy dissemination of past and present field activities" (Priscu & Howkins, 2016).

USAP activity data could inform current and future science planning in all areas where USAP operates.

Key Findings & Recommendations

The project's major finding is that Environmental Engineering is responsible for coordinating or managing numerous datasets without clearly defined plans for their use besides reporting for short-term contract or regulation compliance (e.g., ACA). Over the years, various requests for additional data have resulted in multiple year-end reporting deliverables, some of which are specified in the ASC contract and others that are not. Additionally, some data reported to NSF every year are redundant or no longer relevant. My recommendation is to develop a USAP environmental monitoring framework to guide planning, execution and reporting of all environmental reporting data and monitoring activities. This project also identified barriers to better data management and has recommendations for additional process improvements.

Key Findings

ASC Environmental Engineering data management and reporting processes are isolated within the department or managed by a subcontractor. This approach has caused disjointed and redundant data requests for annual environmental reporting. The following four factors are likely contributors to the current state of USAP environmental reporting data:

1. A subcontractor has managed data collection for annual Master Permit reporting for 28 years and ASC Environmental Engineering does not have visibility or clear understanding of data requirements or purposes.

2. The Antarctic Treaty's Electronic Information Exchange System (EIES)[4] involves biannual data collection and reporting of USAP operational activities. This is an internal process within Environmental Engineering but is managed separately and not integrated with other environmental reporting tasks, which leads to redundant requests for data.
3. Over the years, singular requests for data or information were added to yearly reporting and were either added to the scope of the annual Master Permit report or required by NSF to be reported separately. Furthermore, the overall reporting structure has not been reviewed regularly, possibly resulting in continued data requests that are unnecessary or data collected without well-defined plans for use.
4. Most data collected by ASC Environmental Engineering are collected inefficiently through emailed requests. Standardized reporting would be appropriate for data requests that rarely change from year to year and are requested at the same time every year.

Recommendations

Recommendation 1, for ASC Environmental Engineering – Continue to collaborate with ASC departments to map out data lifecycles used for environmental reporting with objectives to eliminate redundancies, clarify misunderstandings, and identify improvements to streamline data reporting. Additionally, work with ASC Systems Engineering to develop standards for integrated and centralized data management systems for environmental reporting data. Finally, develop an Environmental Engineering Strategic Plan with goals and objectives to 1) develop a USAP Environmental Monitoring Program in order to guide environmental monitoring activities, and 2) finalize a Data Management Plan to document data lifecycles used for environmental reporting.

Recommendation 2, for Antarctic Support Contract – Develop integrated data systems to streamline data processes and reduce redundant and segregated data management. Standard,

[4] Since 2012, EIES is the official means of exchanging information between Treaty nations. Requirements to exchange information started in 2001. Summarized reports can be found at Summarized Reports (ats.aq) (ATS, User Guide, 2021).

enterprise-wide, end of year reporting would improve data management, reduce wasted resources managing redundant data, and increase reliability and relevance of data reported to NSF; this includes weekly situation reports (SITREP), and annual end of season reporting.

Recommendation 3, for National Science Foundation – Review annual environmental reporting requirements for data relevance, purpose, and future use. Additionally, review and develop data use plans for environmental data collected by ASC but not currently reported or used. Develop a long-term vision, strategy, and set goals to measure and reduce USAP environmental impacts. Lastly, develop a system for governing and sharing USAP environmental reporting data with the science community or use a similar framework as the USAP Data Center.

Conclusion

Current NSF requirements for collecting and reporting data will not lead to meaningful action to reduce USAP's environmental impact unless strategic goals for monitoring, measuring, and reducing impacts are developed. Strategic planning is needed to develop an environmental monitoring program with data systems in place to produce consistent and reliable data. An integrated and comprehensive data management plan will provide the framework to organize data and data management practices throughout environmental monitoring program initiatives. In conclusion, when clean, relevant, and reliable data are at the foundation of our analyses, we can gain insight into how to set and meet goals to reduce environmental impacts in the US Antarctic Program.

REFERENCES

Antarctic Conservation Act (ACA). (1993). *Waste Management Regulations* (45 C.F.R. 671): https://www.ecfr.gov/current/title-45/subtitle-B/chapter-VI/part-671

Antarctic Treaty Secretariat (ATS). (2021). *Electronic Information Exchange System.* Retrieved from Antarctic Treaty Secretariat: https://documents.ats.aq/ie/eies_userguide.pdf

DMPTool. (2022). Retrieved from DMPTool: https://dmptool.org/

Faundeen, J., Burley, T. E., Carlino, J. A., Govoni, D. L., Henkel, H. S., Holl, S. L., Hutchison, V. B., Martín, E., Montgomery, E. T., Ladino, C., Tessler, S., & Zolly, L. S. (2014). *The United States Geological Survey Science Data Lifecycle Model* (Report No. 2013–1265; Open-File Report, p. 12). USGS Publications Warehouse. https://doi.org/10.3133/ofr20131265

Michener W. K. (2015). Ten Simple Rules for Creating a Good Data Management Plan. PLoS computational biology, 11(10), e1004525. https://doi.org/10.1371/journal.pcbi.1004525

National Science Foundation (NSF). (2001). *Antarctic Conservation Act of 1978 as amended by Antarctic Science, Tourism, and Conservation Act of 1996.* Retrieved from https://www.nsf.gov/geo/opp/antarct/aca/nsf01151/aca1_intro.pdf

National Science Foundation (NSF). (2021). *United States Antarctic Program*. Retrieved from https://www.nsf.gov/geo/opp/antarct/usap.jsp

Office of Polar Programs (OPP). (2016). *Document Library, NSF 16-055, Dear Colleague Letter: Data Management and Data Reporting Requirements for Research Awards Supported by the Office of Polar Programs.* Retrieved from National Science Foundation: https://www.nsf.gov/pubs/2016/nsf16055/nsf16055.jsp

Oracle. (2022). *What Is Data Management?* Retrieved from OCI: https://www.oracle.com/database/what-is-data-management/

Priscu, J. C., & Howkins, A. (2016). *Environmental Assessment of the McMurdo Dry Valleys: Witness to the Past and Guide to the Future.* Special Publication, College of Agriculture, Montana State University, Department of Land Resources and Environmental Sciences. Retrieved from https://mcm.lternet.edu/sites/default/files/Priscu_MDV_2016_workshop_FINAL_0.pdf

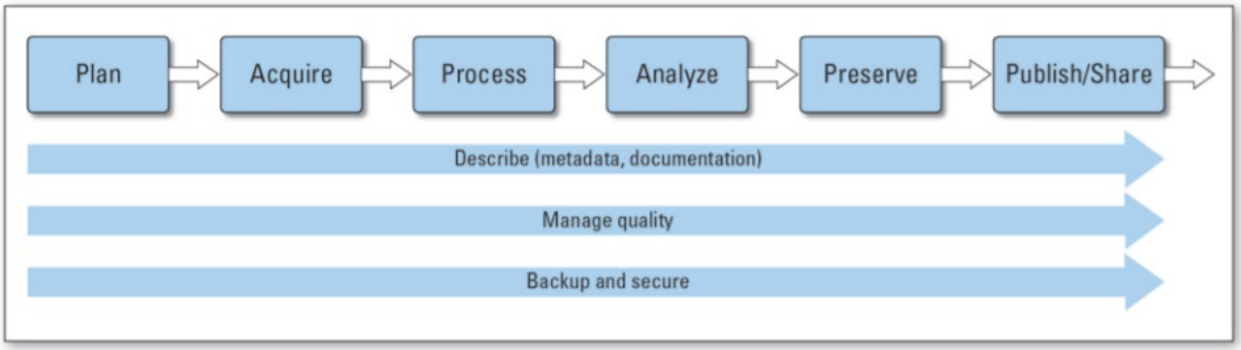
Protocol on Environmental Protection to the Antarctic Treaty (Protocol). (1998). Retrieved from Antarctic Treaty Secretariat: https://documents.ats.aq/keydocs/vol_1/vol1_4_AT_Protocol_on_EP_e.pdf

U.S. Antarctic Program (USAP). (2022). POLAR ICE. Retrieved from USAP: https://www.usap.gov/sciencesupport/polarice/

Vigliarolo, B. (2019). *Data Management: A cheat sheet*. Retrieved from TechRepublic: https://www.techrepublic.com/article/data-management-a-cheat-sheet/

Appendix 1

The USGS Science Data Lifecycle Model



Additional details outlining the USGS Science Data Lifecycle Model (Faundeen et al., 2014):

1. *Plan* – This stage of the process assists data managers in considering all activities related to handling data assets from project inception to publication and archiving. During this stage of the model all elements of the model should be evaluated, addressed, and documented. Considerations include approaches, needed resources (e.g., funding and personnel), and intended outputs for each stage of the data lifecycle. A data management plan (DMP) is the common output of the Plan element of the model.

2. *Acquire* – the second model element represents the activities through which new or existing data are collected, generated, or considered and evaluated for re-use. Some examples are meteorological or physical science sensor data, historical maps, geographic information system datasets (GIS), biological records, satellite observations, scientific observations, and data collected through forms and surveys. It is important to design data acquisition techniques to address research questions while also considering relevant organizational policies and best practices that maintain the attribution and integrity of the data as an information product. The outputs of this element of the model are the data inputs or datasets acquired.

3. *Process* – the third model element represents various activities associated with preparation of new or previously collected data inputs. Processing input data may entail defining data elements; integrating disparate datasets; data extracting, transforming, and loading operations; and standardizing data to prepare for analysis.

A guiding principle of this model element is to align data with established organizational standards to build a foundation of data for integrated analysis. The outputs of this element are datasets that are ready for integration and analysis.

4. *Analyze* – the fourth model element represents activities associated with exploring and interpreting processed data, where hypotheses are tested, discoveries are made, and conclusions are drawn. Analytical activities include summarization, graphing, statistical analysis, spatial analysis, and modeling are used to produce scientific results and information. In this element new data are generated, versions are tracked, and processes are documented. Data management during analysis improves the efficiency of data analysis activities, preserves documentation that is critical for scientific integrity, and creates a foundation for future research. The outputs of this element are interpretations or new datasets, which often are published in written reports or machine-readable formats such as map layers or numerical modeling results.

5. *Preserve* – the fifth Model element represents the activities associated with storing data for long-term use and accessibility. USGS puts this stage ahead of Publish/Share stage to ensure a plan is in place for long-term preservation of data recognizing these activities could be neglected due to pressures of project timelines and budgets. The output of this stage is an established framework of data, metadata, ancillary products, application-neutral storage formats, and any additional documentation, to ensure data availability and re-use.

6. *Publish/Share* – the final element in the USGS science data life cycle model considers how data and results are published through traditional peer-review publications, Web sites, data catalogs, social media, and other venues.