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:

- roles and definitions related to data access, sharing, and security
- privacy considerations related to environmental data
- securing environmental data

Additional information related to data management planning is provided in the ITRC fact sheets on Data Management Planning; Data Governance; Data Lifecycle; Data Storage, Documentation, and Discovery; and Data Disaster Recovery.

1 INTRODUCTION

Data access, sharing, and security are key components of data governance policies. Key concepts related to environmental data access, sharing, and security include but are not limited to:

- **Data access** governs how data are accessed from the source data system, including internal, external, and public access protocols. Government organizations are often subject to legislative or statutory requirements to provide public access to data.
- **Data sharing** defines how data are transferred from and used outside the source data system.
- **Data users** include people who are responsible for data acquisition or data management processes and people who view, analyze, and present the data.
- **Data ownership** can be defined in multiple ways and may include consideration of who is responsible for the data, who creates the data, who maintains the data, and/or who may claim legal ownership of the data.
- **Data privacy** relates to concerns surrounding sensitive or confidential data, including personally identifiable information (PII) and Health Insurance Portability and Accountability Act (HIPAA) considerations.
- **Data security** protocols prevent unauthorized users from gaining access.

2 ACCESS AND SHARING

Data governance policies should describe how data are accessed from the source data system, including access by internal, external, and public users.

Data sharing should be addressed via governance if data are shared or may be shared within or between different organizations. Data sharing agreements may be needed to define restrictions or appropriate use of shared data. When data are shared, associated metadata should also be provided.

For government agencies, data acquired by the agency may become subject to open records requests or legal discovery, so it is important to have well-defined data sharing protocols in place. Be cognizant of your organization’s policies regarding records requests and your state’s open records laws.

3 DATA USERS

Data users include all people interacting with the data at any point in the data lifecycle, including those who need to access or share the data. User access rights/privileges to data depend on the user’s needs and tend to be defined case-by-case, but a best practice is to limit rights/privileges allowing data modification as much as possible to protect data integrity. It is also a best practice to ensure that more than one user can access the data to ensure continuity of any functions that are dependent on the data.

3.1 Stakeholders

Stakeholders include any persons or organizations, including responsible parties or agencies that are invested in or impacted by a situation or issue. It may be difficult to identify all of the potential stakeholders for a data set, so a recommended best practice is to assume a wide range of potential stakeholders when planning, developing, and maintaining a data set. Stakeholders can also be identified by conducting a stakeholder analysis, which is an analysis used to identify potential stakeholders and their level of interest, level of influence, and ability to participate in a given project, situation, or issue.

Other best practices include developing and maintaining a stakeholder registry for tracking stakeholders, grouping stakeholders by appropriate access levels, and documenting processes for engaging stakeholders in a communications plan.

3.2 Internal Users

Internal users within the organization that owns and maintains the data include anyone involved in the generation, documentation, processing, updating, or transfer of data. These internal users include those who build and maintain the database, as well as those who extract information from the database to use for their analyses. Internal users and their roles/responsibilities should be explicitly defined in data governance policies. In many organizations, internal user roles may be distributed among program staff responsible for conducting business processes and information technology staff whose primary role is providing technology support for the organization. The division of responsibility will vary by organization. As such, a program/information technology structure cannot be defined herein that would be best for all organizations. However, it is recommended that all organizations define roles for managing data and information.

Staffing Considerations

Despite the ubiquity of technology and automation in society, people remain an important element in most data management processes. It is therefore best practice to hire qualified candidates and ensure that they have adequate training and support.

Data management roles and positions should be included in organizational change management and succession processes to ensure that these valuable functions are available to manage an organization's data.

The following roles should be filled, and may be performed by one or multiple people depending on the scope of data management activities and the size of your organization:

- **Data steward –** Data stewardship is the accountability and responsibility for data and processes that ensure effective control and use of data assets and is enabled by defining data stewards within the organization. A data steward is the person or persons who are responsible for managing a data set and ensuring it is usable to the organization and end users. Data stewards are responsible for the maintenance of main data and metadata, as well as the data management system, to ensure quality, compliance, availability, access, and preservation. Data stewards are subject matter experts who drive the execution of processes and procedures to enforce data governance policies. They must be familiar with the data security requirements of the project, as well as any applicable requirements for public access and data sharing. Core responsibilities of a data steward include:
  - keeping the database up-to-date while ensuring that all appropriate quality assurance/quality control (QA/QC) procedures are followed
  - providing a point of contact for internal and external users to access the data
  - managing access and permissions for internal and external users

  A data steward must have the ability to:

  - create and manage metadata for the data set
  - document rules and standards
  - execute operational data governance activities
  - set and manage guidelines around data and its use
  - ensure data complies with applicable regulations
  - correct data and data-related issues

  A data steward should understand the importance and utility of the data from business (how the data are used), data management (how the data are acquired/maintained), and technical (how the data are structured/stored) perspectives. To ensure data usability, every data set should have a data steward. A single data steward may be responsible for multiple data sets.

- **Data acquisition manager**—The data acquisition manager should have the experience and training necessary to ensure all acquisition requirements are met. The data acquisition manager determines the process through which the data will be acquired and stored. This person may need to be familiar with data exchange, including electronic data deliverable (EDD) formats (see Electronic Data Deliverables and Data Exchange Fact Sheet), database platforms, document archive systems, etc.
- **Data management systems administrator—**The data management systems administrator is the person

responsible for identifying and maintaining data storage infrastructure. This role may be handled by an organizational department or committee, possibly in an information technology (IT) department or group. This role is responsible for determining the system through which the data will be updated and maintained and may need to develop and document workflow plans for data import/export and identify the people who will be doing the day-to-day work of database maintenance.

- **Quality control manager/officer**—People responsible for ensuring that quality objectives are met in accordance with the quality assurance project plan (QAPP) or organizational or project-based quality standards. Responsibilities may include setting content standards and functional rules, performing or supervising quality control activities, and ensuring results are of adequate/usable quality.
- **Data analyst/end user**—Person who extracts information from the database to use for a specific purpose.

3.3 External Users

For data that contain sensitive or restricted information, data sharing agreements may be required. An organization’s access policy or acceptable use policy would be appropriate tools for specifying data sharing requirements. Data sharing agreements should address the following issues:

- data that should not be publicly released
- metadata or supporting data that must be provided with the primary data set(s)
- citations that should accompany the primary data set(s)
- sources to allow end users to get more information about the shared data set(s)

Other users may exist outside of the users who are using the data for its intended purpose. Although these users are using the data for an unintended purpose, the organization should have processes in place to allow data use by these users.

4 DATA OWNERSHIP

From a legal and policy perspective, data may be owned by the organization or individual responsible for creating, storing, and maintaining it, but some organizations choose to also designate individuals with the data owner role. A data owner is defined by the Data Management Association (DAMA) as an individual responsible for definitions, policy, and practice decisions about data within their area of responsibility (DAMA 2016). Data owners may also be data stewards, but not all data stewards may be data owners, dependent on the policies and management structure of the organization hosting the data.

It is important to note that the definition of data owner discussed above is from the perspective of implementation of data management practices. However, the data ownership term itself may also refer to the legal and/or financial ownership of data. The legal definition of data ownership is a rapidly evolving topic, and therefore legal considerations must be considered prior to, during, and after any data collection activities. Per the U.S. Department of Health and Human Services Office of Research Integrity (Steneck 2007), the following should be considered when evaluating data ownership:

- **funding**—Those who provide financial support for generating data will likely claim some sort of ownership rights to the data, which may allow them to dictate how the data may or may not be used.
- **institutions**—If data are collected on behalf of an organization or institution, the institution may claim ownership rights over data collected with funds given to the institution.
- **data sources**—Entities that are the source of data may seek some control over data derived from them and therefore claim an ownership stake in the end results.

Per the Office of Research Integrity (Steneck 2007), the following questions should be answered prior to undertaking any data-generating activities:

- Who owns the data I am collecting?
- What rights do I have to publish the data?
- Does collecting these data impose any obligations on me?

Due to the nature of environmental work, multiple parties are often involved in the collection and reporting of environmental data. For example, consider the situation where a groundwater sample is collected from a private residential drinking water well. In this scenario, the sample itself was collected from private property owned by the resident. It may have been collected by a third-party environmental consultant, working on behalf of a client or responsible party that is financing the data collection activities, and likely was collected at the request of a regulatory agency. In this scenario, the private resident, environmental consultant, responsible party, and regulatory agency may all potentially claim partial ownership of the data collected. It is important to be aware that any party involved in the data collection, management, and/or evaluation process may have the right to claim at least partial ownership of the data. To the extent possible, it is advised that when answering the questions above the answers be recorded in writing and agreed to by those with potential ownership stake in the data.

5 PRIVACY

Privacy refers to maintaining the anonymity of people and organizations associated with data. Organizations should develop policies and take measures to comply with privacy requirements and ensure that data are protected during storage, access, and transfer. Access should be limited to authorized users, and data should be used without revealing sensitive information about the associated people or organizations. PII and HIPAA considerations should be taken into account. The Department of Homeland Security (DHS) defines PII as any information that permits the identity of an individual to be directly or indirectly inferred, including any information that is linked or linkable to that individual (DHS 2021). This information includes a person's first name or first initial and last name in combination with the person's address, telephone number, and/or the person's Social Security number or driver's license number. HIPAA is a federal law that requires the creation of national standards to protect sensitive patient health information from being disclosed without the patient's consent or knowledge (CDC 2018). Examples of privacy concerns that may be encountered during environmental data management include information related to residential or commercial private properties where sampling occurs, environmental health information, or location information. Collection and use of private or sensitive data may require that consent be obtained and the record of consent be stored with the related data set.

To prevent unauthorized use of private or sensitive data, data storage systems should also track and classify data based on privacy and sensitivity considerations by classifying the entire data set and documenting the privacy and sensitivity considerations in the associated metadata. Access rights to the data storage location should be set to limit the individuals who have access to the data.

6 SECURITY

Data security is the practice of protecting digital information from unauthorized access, corruption, or theft throughout its entire lifecycle, encompassing physical security, software security, administrative and access controls, and organizational policies and procedures (IBM 2021). Data security is directly linked with data access, as the purpose of data security is to control who has access to the data and data systems, and to control what they can do with the data. Generally speaking, it is a best practice to keep access to the original data source on a "need to know" basis, where users have access only to the data and data systems they require for their needs.

Several federal agencies responsible for the handling of environmental data, including the United States Environmental Protection Agency (USEPA), the United States Geological Survey (USGS), and the National Oceanographic and Atmospheric Administration (NOAA), have published guidance regarding data security for their organizations that can be useful and relevant resources when developing data governance policies specific to environmental data. Key concepts from these guidance documents are summarized below.

6.1 USEPA

The USEPA Information Security Policy (USEPA 2021) includes the following items as part of their information security program:

- periodic assessments of the risk and magnitude of harm that could result from the unauthorized access, use, disclosure, disruption, modification, or destruction of information and information systems
- policies and procedures that:
    - are risk-based
    - reduce security risks cost-effectively

    - address information security throughout the information system lifecycle
    - ensure compliance with information security directives

- systems security engineering principles, concepts, and techniques are employed
- supply chain risk management principles are employed to protect against:
    - the insertion of counterfeits
    - unauthorized production, tampering, theft, insertion of malicious software
    - poor manufacturing and development practices

- definition and effective implementation of minimum mandatory technical, operational, and management security controls or other compensating countermeasures
- subordinate plans for providing adequate security for all networks, facilities, and individual or groups of information systems
- mandatory security awareness training and role-based information security training
- periodic testing and evaluation of management, operational, and technical controls
- continuous monitoring of information security controls
- a process for planning, developing, implementing, evaluating, and documenting remedial actions to address deficiencies in information security controls
- capabilities for detecting, reporting, and responding to security incidents
- plans and procedures to ensure continuity of operations so that security controls remain effective over time

6.2 USGS and the Department of the Interior

USGS has published several best practices for backing up and securing data (USGS undated). These best practices recommended by USGS include:

- share metadata but keep confidential or sensitive information unavailable
- create codes to make data anonymous
- keep the data dictionary secure
- when transferring sensitive data to another party, encrypt the data
- make sure your computer has antivirus and firewall software that updates regularly
- make sure data are physically protected in a locked drawer or on a secure network

USGS has also published guidance specific to data security considerations during the data acquisition phase of the data lifecycle. As discussed in more detail in the USGS guidance, the goal of these policies is to protect data from uninvited disclosure or intentional corruption, and to secure data systems from external attacks to the maximum extent possible. The USGS guidance is based on guidance and policies published by the U.S. Department of the Interior (DOI) Information Assurance Division. Responsibilities of the DOI Information Assurance Division that are relevant to environmental data management include:

- develop enterprise IT security policies, standards, guidelines, and procedures
- ensure the confidentiality, integrity, and availability of information and systems
- oversee system assessments and authorizations
- support cyber security assessment and management (CSAM)
- develop an enterprise risk management framework
- establish an enterprise continuous monitoring program
- develop Privacy Act policies, standards, guidelines, and procedures
- identify relevant IT infrastructure controls for implementation to meet Privacy Act requirements

See the DOI Information Assurance Division website for additional details.

6.3 NOAA

NOAA has also published several resources related to environmental data management and security policies. These resources can be found on the NOAA Environmental Data Management Committee (EDMC) home page. In particular, the NOAA Information Technology Security Policy (NAO 212-13) includes several best practices applicable to general environmental data management, including:

- ensure safeguards exist to protect the confidentiality, integrity, and availability of data and data systems
- protect environmental data and data systems from abuse and misuse
- protect information from unauthorized disclosure, destruction, or modification while collected, processed, transmitted, stored, or disseminated
- apply security policies throughout all phases of an information system’s lifecycle

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.