ITRC has developed a series of fact sheets that summarize the latest science, engineering, and technologies regarding environmental data management (EDM) best practices. This fact sheet provides an overview of data exchange and data migration, introduces some considerations for both data exchange and data migration that are detailed in subsequent fact sheets (Valid Values, Electronic Data Deliverables and Data Exchange, and Data Migration Best Practices), and introduces two related case studies from the U.S. Geological Survey (see USGS Challenges with Secondary Use of Multisource Water Quality Monitoring Data Case Study) and Minnesota Pollution Control Agency (see Historical Data Migration Case Study: Filling Minnesota's Superfund Groundwater Data Accessibility Gap).

1 INTRODUCTION

Data exchange is the transfer of data between a source and target, often two environmental data management systems (EDMS). The source and target each have a data structure or schema, with rules that govern the data formats. The transfer involves copying data to an electronic data deliverable (EDD), which is a temporary, intermediate format used to transform the data.

Exchanging data isn't an easy process. Data sets vary in complexity or completeness, making them a challenge to intermesh. Additionally, data fields and values can differ in name and/or definition between EDMSs.

Certain practices help to ensure consistency and completeness of the data within an EDMS. Those practices become even more important during data exchange. The following fact sheets and case studies provide best management practices and examples for data exchange.

- Valid Values
- Electronic Data Deliverables and Data Exchange
- Data Migration Best Practices
- Historical Data Migration Case Study: Filling Minnesota's Superfund Groundwater Data Accessibility Gap
- USGS Challenges with Secondary Use of Multisource Water Quality Monitoring Data Case Study

In addition to the fact sheets, the EDM Best Practices Team compiled a noncomprehensive list of resources for valid values and EDD formats.

2 VALID VALUES AND ELECTRONIC DATA DELIVERABLES

EDMSs have rules that govern their structure or schema. The rules include what data fields are included and constraints on data field type (for example, number or text, field length, etc.).

Additionally, certain data fields allow only specific, predefined values, known as valid values. Valid values include codes or names and their definitions. Examples of valid values include what analytical method was used or what projection or datum a spatial coordinate is in. Managing valid values includes development of new values, changing values, and communicating values and changes.

EDDs are powerful tools in data exchange. They collect, transform, and deliver data from source to target. EDDs can transmit data into or between EDMSs. They can also be used to update or add information to what is already in an EDMS. EDDs contain data in a specific format. The format reflects the rules governing both the source and destination data structures, including valid values. EDDs themselves can also take on multiple forms, like a spreadsheet or comma-separated values (CSV) text file.

3 DATA EXCHANGE AND DATA MIGRATION

Data exchange uses an existing import/export process. The EDD is already established for the purpose of transferring data from source to target. Data exchanges are normally planned in advance and can occur with regular frequency. Data exchanges typically transfer current data between active EDMSs.

Data migration is a subset of data exchange. The primary difference is that there isn't an existing export/import process between the source and target. Data migrations are also typically unplanned until the need arises. They are usually one-time

events. The source data is usually not from an active system. The differences between data exchange and data migration are summarized in Table 1.

Data migrations are sometimes referred to as “historical” data migrations because they are often used to get historical data sets into an EDMS. The term “historical” is omitted here to promote generality and because the line between “historical” and “current” data sets isn’t always clear.

For example, migrating data from 20-year-old paper documents is clearly a historical data migration. Migrating data from an EDMS that was actively maintained and distributed until one month ago isn’t exactly “historical,” but it’s still considered a data migration.

Although data migrations are usually one-time events, they can sometimes be repeated. For example, the source data set might continue to grow or more likely, data sets in the same format as the original migrated data set are discovered. Subsequent migrations are more like data exchanges because there is an existing import/export process. Data migrations are challenging because of the formats, missing metadata, and media encountered in older data sets.

Table 1: Differences between data exchange and data migration

Data Exchange	Data Migration
Existing data export/import process	NO existing data export/import process
EDD established	NO EDD established
Planned in advance	Conducted as needed
Occur with regular frequency	One-time event
Active data source and target	Data source usually isn’t active

4 CASE STUDIES—USGS NUTRIENT AND MPCA

The U.S. Geological Survey (USGS) and Minnesota Pollution Control Agency (MPCA) case studies highlight the importance of consistent metadata and valid values. See the USGS Challenges with Secondary Use of Multisource Water Quality Monitoring Data Case Study and the Filling Minnesota’s Superfund Groundwater Data Accessibility Gap Case Study.

5 A NOTE ON DATA QUALITY

When exchanging data, all available metadata and information relating to data quality should be included. Data quality is important for evaluating potential uses, assessing uncertainty in decision-making, and communicating effectively to the public. Some data quality–related fields are fairly standard, such as result data qualifiers or monitoring location horizontal precision. However, for other data, such as depth measurements, observations, and screening tool data, the data management community isn’t consistent about how quality and usability are documented. For these instances, applicable metadata such as instrument calibration or other criteria should be included in the data exchange.

6 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.