1 INTRODUCTION

The Minnesota Pollution Control Agency (MPCA) began a three-year project to compile groundwater data for the state-funded Superfund sites and provide a web-based interactive map of groundwater contamination in 2017. This project, completed in 2020, resulted in the production of the Minnesota Groundwater Contamination Atlas, which maps areas of groundwater contamination concern and provides information to both technical staff and the general public about the contamination. The Minnesota Groundwater Contamination Atlas can be found at the following link: https://www.pca.state.mn.us/data/minnesota-groundwater-contamination-atlas (MPCA Undated).

The initial scope was limited by design to include only sites where the MPCA directly oversees the investigation and owns all of the data. After completing the initial phase, the platform was expanded to include additional information from other MPCA remediation sites. As of April 2022, the Minnesota Groundwater Contamination Atlas has been successfully expanded to include Minnesota's Closed Landfill Program remediation sites along with select petroleum remediation sites.

Data from remediation sites are some of the most requested data across all state government agencies in Minnesota. Prior to the project start, most of that data were embedded in documents. To fulfill data requests, MPCA staff needed to sort through physical reports or digitized versions of those reports to provide the data. The first step in filling the data gap was to migrate this historical data from documents into a database.

2 PROCESS

2.1 Historical Data Migration

2.1.1 Planning

The historical data migration process began by identifying the desired work products and then the data elements needed to produce them. After that information was determined, the project members reviewed existing data standards and valid values list to ensure that the proposed work conformed to existing data standards. The MPCA uses a commercial off-the-shelf database system and works with the vendor to develop both the electronic data deliverable (EDD) format and the reference values. In addition, the State of Minnesota has a separate database, the Minnesota County Well Index, that is the database of record for well information, and the project needed to conform to those standards (Minnesota Geological Survey (MGS) Undated). While a good portion of information could fit into existing enterprise database systems, the team was tasked with determining what changes needed to be made for the remaining information. These changes included:

- augmenting reference value sets
- refining EDD formats
- ensuring that unique identifiers were consistent

While the scope was limited as part of this initial historical data migration, the decisions made as part of this migration needed to be carefully mapped out to ensure that future data migrations for other remediation sites managed by the MPCA and submittal of newly collected data could follow the standards and practices established as part of this work.

MPCA staff began evaluating both enterprise systems currently in use by the State of Minnesota and systems used by other governmental agencies. After staff determined the eventual home of the data, they met with the administrators of those data systems to ensure that they were following the established procedures and valid values for these systems. By conforming to these established standards, the team was able to avoid duplicative efforts and allow for a smoother migration process. In addition, this allowed easier sharing of information across the state government.

2.1.2 Non-Analytical Data Transcription

The most intensive step in the data migration process was the data transcription. While the driver for the data migration process was analytical data to develop the areas of groundwater contamination concern, the sample locations for each of these samples needed to be established beforehand. This information was determined from reading through historic reports for each of the sites to gather sample location names and coordinate data.

Establishing the groundwater sample locations involved working with the state's enterprise data system for well information to search for existing monitoring locations and gathering metadata about these wells. Although a good number of the wells

were found in existing enterprise systems, a number of other location types, such as temporary groundwater samples, were explicitly not included in this data system. Staff worked with the database administrators to evaluate these locations and allow them to fit into schema already established.

2.1.3 Analytical EDD Acquisition and Refinement

Once the locations had been established, staff began the process of collecting and preparing the analytical data for loading into the enterprise system. Staff had to evaluate several paths forward for this data migration, including transcription of the data from limited tables included in reports, attempts to use optical character recognition on scanned copies of physical reports, or attempts to reach out to the data collectors and laboratories that analyzed the data to potentially receive the data in a digital format. Staff decided to begin with the path of reaching out directly to the laboratories themselves to try and receive copies of the analytical data in a digital format.

Minnesota laboratory accreditation requirements state that laboratories keep the results from their sample analyses for a period of 5 years. Staff took advantage of this requirement and limited their scope of analytical data collection to the previous 5 years' worth of data to ensure that the data set was relatively full and robust. Although staff were able to receive EDDs from the laboratories, because these samples were not submitted with the intention of generating an EDD, they were missing some important metadata. Because these samples were up to 5 years old, staff received what EDDs could be generated as-is and began to process these.

The EDD cleaning process included:

- reconciling the valid values in the historic EDDs with current valid values
- determining unique identifiers for each sample (for example, concatenation of date to the well identifier)
- replacing the common name "MW-1" with the unique well identifier that is used in both the Minnesota County Well Index database and across other agencies in the state
- matching quality control samples with the appropriate batch data to ensure public trust in the data used to make regulatory decisions and allow for third-party data quality evaluations

2.2 Supporting Framework and Documentation for Historical Migration

In addition to the data migration, additional infrastructure building was required to drive the online interactive portion of the Minnesota Groundwater Contamination Atlas. Discussions between technical staff and information technology (IT) staff took place to ensure that data that would be shared with the public met the minimum data requirements and were accurate. This involved disseminating the data quality standards that were involved in the data migration, explanations of the valid values contained within the files, and sharing resources related to data structure and electronic data deliverables.

After the data migration was complete, staff took care to compile standard operating procedures and metadata that were generated as part of the migration. This information is used both internally at the MPCA for adding new sites to the Minnesota Groundwater Contamination Atlas data migrations and shared externally to ensure new data submittals meet the established standards. The team recognized the amount of work that was required to perform the data migration and wanted to ensure that the lessons learned and information were saved to allow future data migrations to follow this path. Like the activities that occurred in the beginning of this project referencing existing enterprise solutions, the staff wanted to ensure that future data migrations could use the information and procedures gained to expedite the process and ensure data quality.

Metadata created for the project is available online at the following links:

- https://www.pca.state.mn.us/sites/default/files/c-rem1-15.csv
- https://resources.gisdata.mn.gov/pub/gdrs/data/pub/us_mn_state_pca/env_mn_gw_contamination_atlas/metadata/metadata.html

3 CHALLENGES

Funding for this three-year project was provided by the Minnesota Environment and Natural Resources Trust Fund (ENRTF) as recommended by the Legislative-Citizen Commission on Minnesota Resources. Because of this funding source, there was

a hard deadline at the end of the three years. This required the project to have a defined scope and a sense of urgency to complete it. In addition, dedicated staff members were hired to complete this project to ensure that it was done on time. Even with dedicated staff and a rigid timeline, the project required tremendous in-kind contributions from other funding sources at the MPCA to allow staff who were not funded by the ENRTF to work on the project (in addition to their other job duties).

The process of loading the analytical data was difficult and frustrating for team members. In particular, reconciling common names such as MW-1 or Observation Well 2 with a standardized unique identifier was difficult given that a number of these sites had existed for decades. The relational nature of analytical data, where results are related to tests that are related to samples, caused errors to cascade and initially caused confusion and despair among the team.

Providing a work product that met the requirements of multiple enterprise systems and conformed with the up-to-date valid values was also problematic. The historic data could not be changed and some of the metadata associated with it were lost to time. This caused some of the entries to be lacking quality control information. Staff were required to identify samples where information was missing and, because this was a historical data migration, flag the data appropriately to ensure that it is used appropriately.

The three-year project required over six full-time position equivalents, a personnel budget of $400,000, an in-kind contribution from existing MPCA staff of approximately $399,000, and over $140,000 in IT costs.

4 RESULTS

The project made analytical data and maps accessible to internal staff, state contractors, and the public. Migration of the data from documents into a database allows the data to be easily accessible, well documented, meaningful, and usable. Considering such factors as valid values and EDDs as part of the migration not only ensured a high-quality product but also allowed the data migration process to be applied to other data sets.

This historic data migration involved over 14,000 unique sample locations established in Minnesota’s enterprise system for storing environmental data (see Figure 1). From these over 14,000 locations, over 36,000 individual samples were migrated from historic lab deliverables to current standards. These 36,000 samples included close to 15,000 unique samples collected from the field that were indicative of environmental conditions, with 545,000 individual results associated with those unique samples. Although some of the data were not able to be migrated due to missing valid values or unclear sample location information, the framework set up by the project allows for work to continue on these efforts.

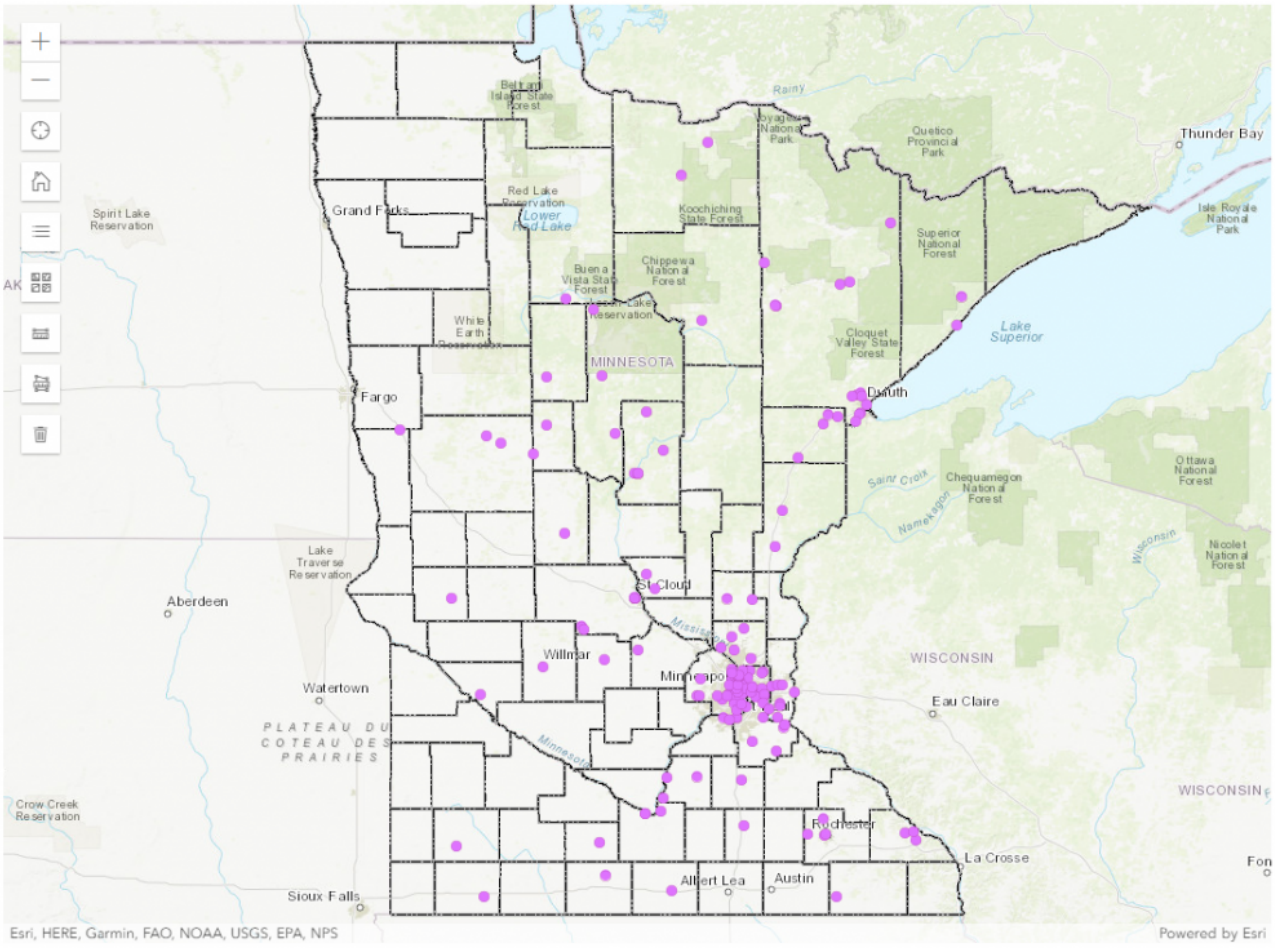


Figure 1: Interactive web-based map showing the sites in the Minnesota Groundwater Contamination Atlas. The online portion of the Minnesota Groundwater Contamination Atlas was viewed by over 1,100 non-MPCA users in the first month of release; in the first six months there were over 13,000 views of the online portion. The expandable framework is serving its purpose—along with the well documented nature of the historic data migration, additional remediation programs are being incorporated into the atlas.

5 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.