

## Data and Information Management Exercise

This document outlines the steps for the supplemental exercises for the Data and Information Management role. The exercise includes basic retrieval, manipulation, reporting of data, and a reflection at the end. You are free to use whatever method or tools you are comfortable with to complete the task.

Please take no more than **2 hours** to get as far as you can in this exercise. If you are unable to complete the task, please include a brief write up or bullet points of the steps you would have taken with more time.

Skeo performs and tracks work on an hourly basis. It is especially important to not exceed the time limit. Failure to do so may result in Skeo having higher than likely expectations for what you can accomplish in that timeframe. This is also a good opportunity for you as an applicant to consider whether working under an hourly paradigm is a good professional fit for you.

### 1. Obtain Source Data

At Skeo, we often analyze and report data for Superfund/National Priorities List Sites. Navigate to the [Superfund Site Information Report Page](#) to obtain a list of sites and site information.

To do so, select the 'View a list of all NPL and SAA sites' as shown in the screenshot below:

[View a list of all NPL and SAA sites](#)

Then the 'Download Excel file containing values for all search criteria' (as shown in the screenshot below)

To conduct another search, return to the [Search Superfund Site Information](#) page.  
[Download Excel file containing values for all search criteria](#)

*Note: your PC may warn you about opening the download. It is safe to open. This warning is due the .xls extension of the file.*

### 2. Clean and Join with additional data

There are a variety of related data about each site that EPA uses to help support all stages of the cleanup process including remedies, site geographic boundaries, and reuse. Choose one of the following data sources to join or crosswalk with the Sites table you downloaded in Step 1. Data source options include:

- Superfund National Boundary Features – [EPA Link](#)
- Remedy Data – *In Data Source Folder*

Using any of the data sources above, clean the Sites table and related table, and join the two in a way that helps complete the 3<sup>rd</sup> step.

### **3. Pivot and Report Data**

When reporting out data, it is common for us to pivot and group. Common fields to group or pivot by include Region, NPL Status, and Federal Facility Indicators. However, you can use any categorical fields for this step. Think about what questions could be answered by the related data you selected in Step 2. Then, reshape the data to answer these questions. Some potential questions include (feel free to use these or come up with your own):

*Which Region/NPL Status have*

- *Highest number of Site Boundaries, Media, Remedy*
- *Relative percentage Site Boundaries, Media, Remedy*
- *Missing Site Boundaries*

*How does each Region/NPL Status break down by*

- *Boundary Feature Type*
- *Remedy selected*

You can report these results out in any way you think is a good fit including tabular, narrative (with tabular supplement), or through visualization.

### **4. Process Enhancement Brainstorm**

Reflect on the data analysis and reporting you've conducted, considering you have more resources and time to enhance the project. With your skills and experience in mind, please describe how you would improve or expand this reporting process. You can focus on elements such as automation of data retrieval and processing, advanced data visualization techniques, documentation, or any other approaches.

### **5. Share Results and Methods**

After you have finished or have worked through the 2-hour window, please upload any steps taken and results to the folder linked in the email. You could include a brief write-up or bullet points of the methods you used. Additionally, it could be helpful to include code snippets, a Jupyter Notebook, or an Excel workbook.