Integrating R, Power BI, and ArcGIS Online to Guide Sediment Investigations and Allocation Toward Better Decision-Making

Battelle Eleventh International Conference on the
Remediation and Management of Contaminated Sediments

Tim Negley (tnegley@intell-group.com), Jamie C. Combes (jcombes@intell-group.com), and Katie Ives (kives@intell-group.com) (TIG Environmental, Syracuse, NY, USA)

Background/Objectives. The need to evaluate large environmental data sets is common at contaminated sediment sites, but project teams often encounter obstacles that prevent evaluating such data quickly and efficiently. One obstacle is the limited ability to rapidly evaluate different data types (e.g., structured and unstructured data), often collected by separate entities. For example, structured data can include analytical chemistry data collected at upland sites and in adjacent soils and waterbodies. The ability to view, analyze, and interpret these data in real time is frequently needed to guide further research and investigation. Other types of structured data can include geographic information systems (GIS) data, such as hydrodynamic modeling outputs, property boundaries, discharge pathways, historical site layouts, sewer networks, outfalls, depositional areas, and sediment management areas. However, in some cases, such as in allocation matters, it is essential to evaluate unstructured data. Unstructured data can include information in PDF files and other text documents. Information can relate to potentially responsible parties, process operations, environmental discharge history, chemical use information, and descriptions of inadvertent onsite pollutant releases. Unstructured data can also include media files such as digital images and video. Enabling the project team to interact with these data sets at the beginning of the process in real time, and not at the end, is important for decision-making and guiding the investigation and allocation process.

Approach/Activities. In this presentation we will demonstrate an application of Microsoft Power BI as an interactive and user-friendly data visualization platform that allows users to interact with big data sets from multiple sources and of different data types. Users can interact with the information through a user-friendly website without the need to install any desktop software. Content developers can embed ArcGIS for mapping spatial data, R and python programming languages for custom visualizations and advanced data analytics, and Microsoft Data Analysis Expressions (DAX) for sophisticated data querying and evaluating "what-if" scenarios. When designed properly, the system is highly scalable, allowing users to evaluate millions of data entries with ease. Information from different data types can be summarized in interactive maps, charts, and tables. Interactive visuals allow users to turn data on or off, reduce data with filters, or select a group of points on a map. Analytical chemistry data can be overlaid on GIS base layers along with research findings of potential sources. Content developers can use the R programming language to achieve more advanced analytics.

Results/Lessons Learned. The data visualization system has been used successfully internally and externally. Project teams can evaluate millions of rows of data with no need to install software. The system has been useful during team meetings to evaluate questions in real-time as they arise. The ability to evaluate what-if scenarios has also been a useful tool. For example, data can be screened against variable screening criteria and summarized immediately in tables, maps, and figures eliminating the need to generate static interim figures only for one-time use.