

2024 Battelle Chlorinated Conference

June 2-6, 2024 | Denver, Colorado

From Spreadsheets to Scripts:

Optimizing Data Analytics Workflow with an R Shiny Application for a Complex Groundwater Site

Emma Ehret, PE
Karen Kelley, PE

June 5, 2024

William Lai
Chapin Storrar, PE

Tamzen Macbeth,
PhD, PE

Riley Cotter
Jack Elsey, PhD

Hannah Rolston,
PhD



Overview

- Background
- Development Approach
- R Shiny Application
- Lessons Learned and Takeaways

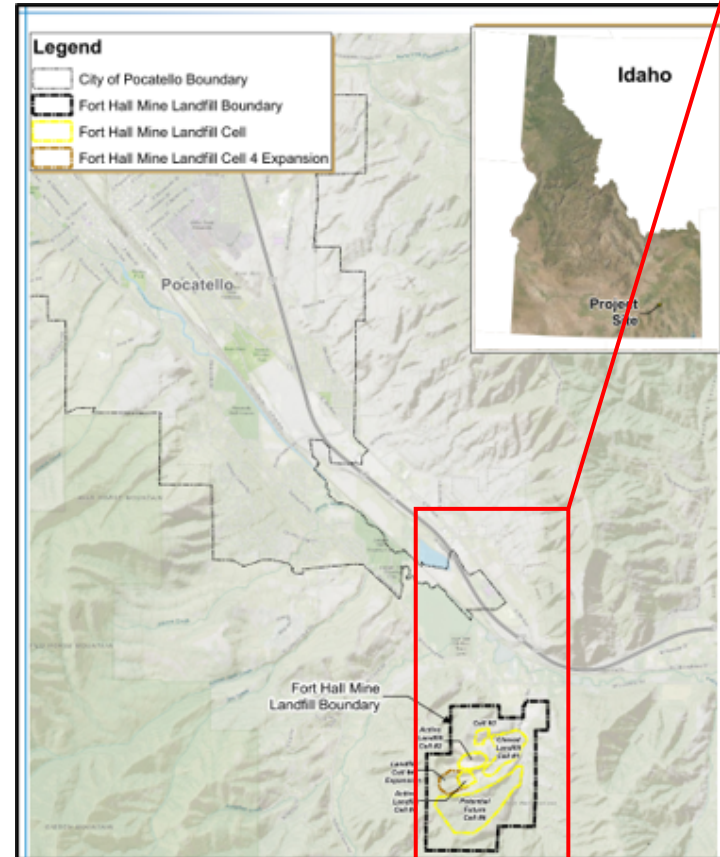


Background



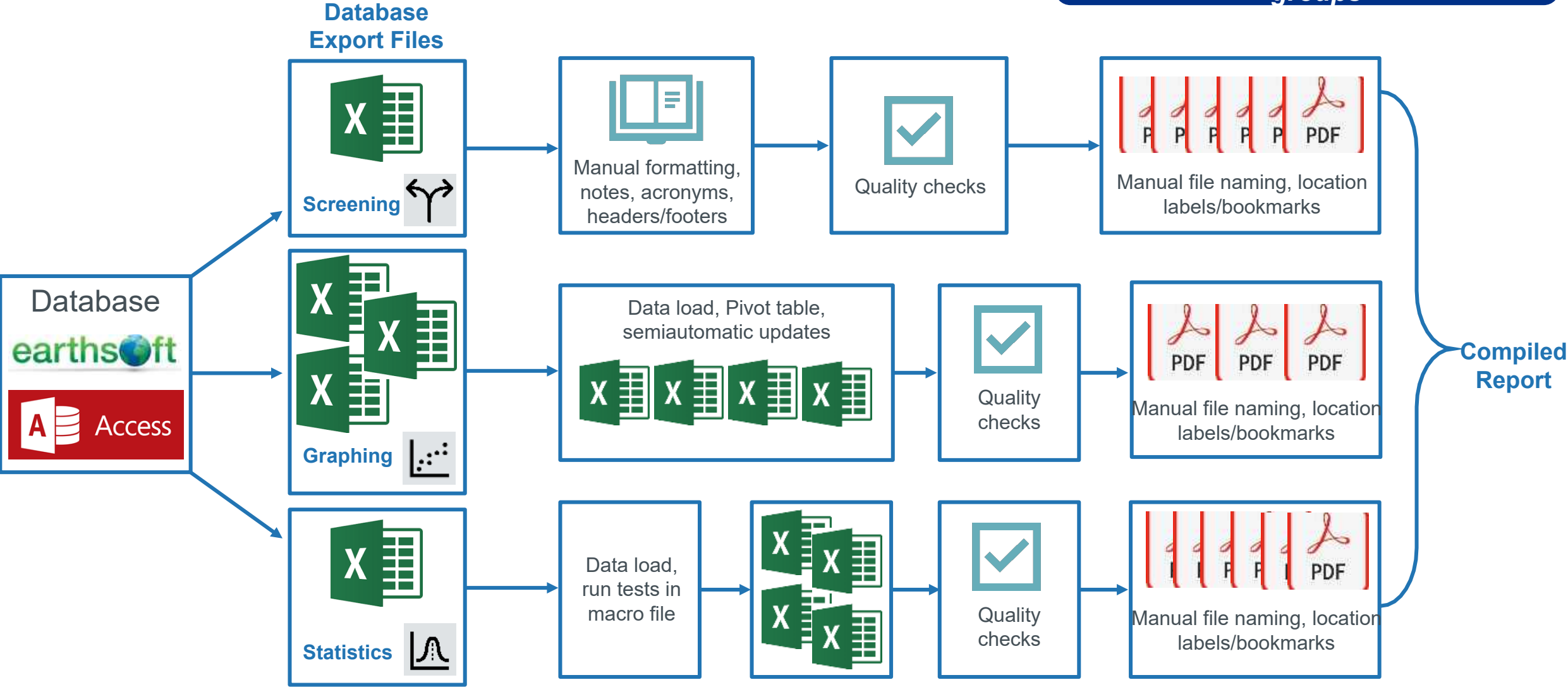
Site Background

- Fort Hall Mine Landfill, ID
- PCE, TCE in groundwater
- 128 wells, 221 regulated analytes
- 1990 – Present; 3 events/year
- RCRA – 3 monitoring tiers
- Pump and treat since 2002 → incomplete capture
- Groundwater monitoring program, treatment system O&M / optimization, site characterization



Spreadsheet Workflow

All Excel and PDF files are organized by location and analyte groups



Project Objectives

Overhaul existing spreadsheet workflow into scripts.
Develop a data analytics tool with...

Comprehensive,
efficient, accurate
analyses

Flexible data
input options

Reusable
intellectual
property (IP)

Accessibility



Development Approach

Project Solutions

Overhaul existing spreadsheet workflow into scripts
Solutions:

Comprehensive,
efficient, accurate
analyses:

**Data analytics
tool coded in
R language**

Flexible data input
options:

**API and flat
file data
upload
options/data
management**

Reusable intellectual
property (IP):

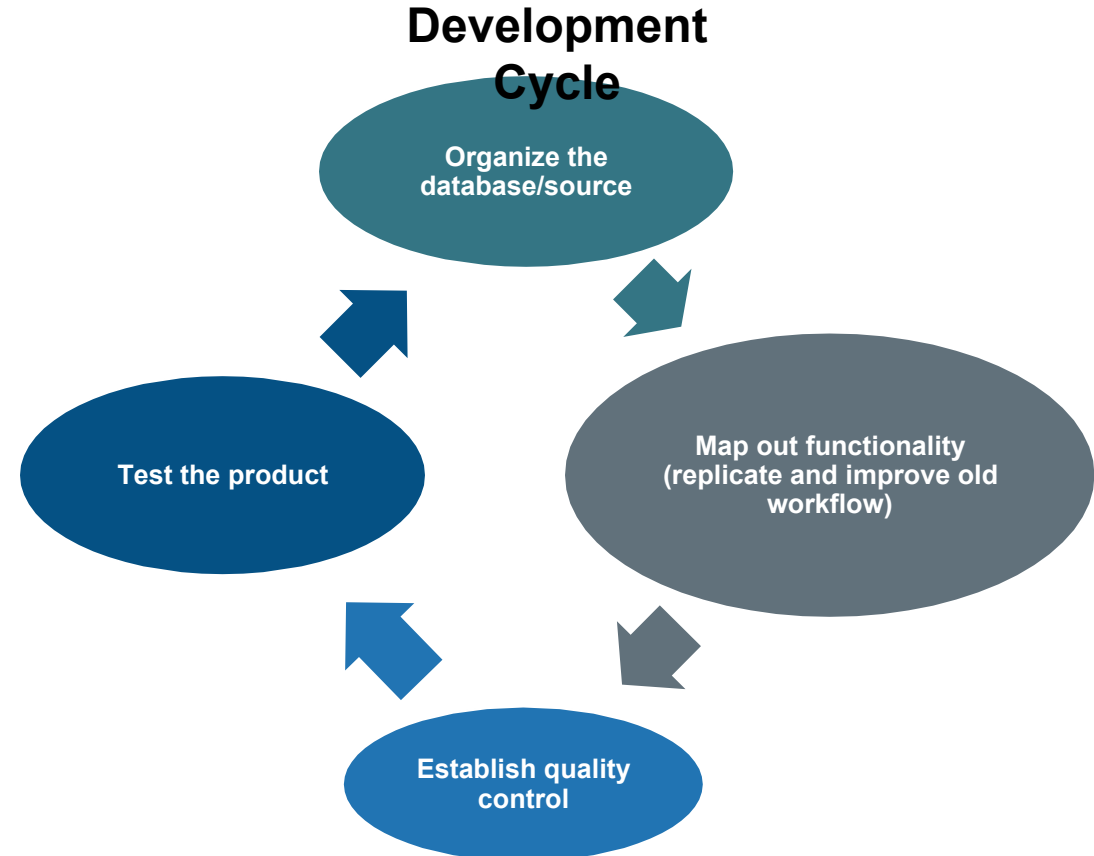
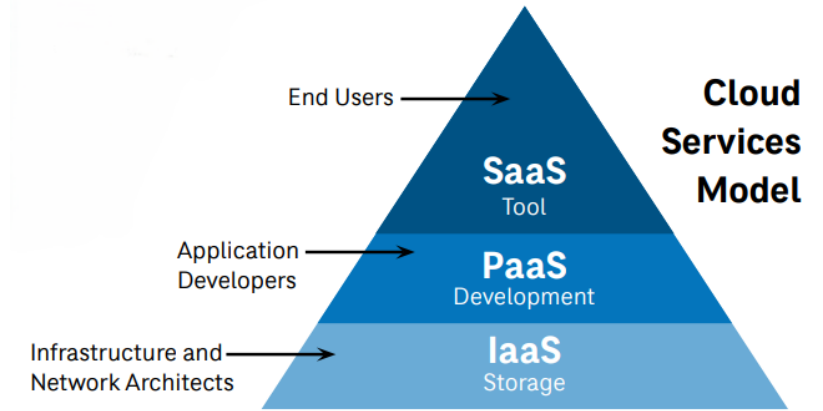
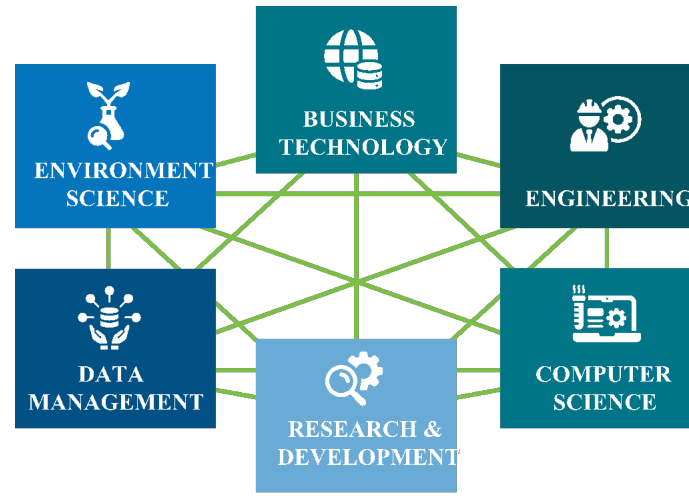
**R scripts and
GitHub
repository**

Accessibility

**Shiny web
application for
internal staff**

General Process

1. Identify funding
2. Understand available tools
3. Identify the team
4. Set up server and subscription
5. **Development cycle**
6. Expand users, incorporate feedback
7. Release product





R Shiny Application

Terminology

R

- Programming language
- Environment for statistical computing and graphics
- Free software
- Wide range of packages and extensions
- Large user community

Shiny

- Framework for building web applications using R



```
library(shiny)

# Define UI for app that draws a histogram ----
ui <- fluidPage(

  # App title ----
  titlePanel("Hello Shiny!"),

  # Sidebar layout with input and output definitions ----
  sidebarLayout(

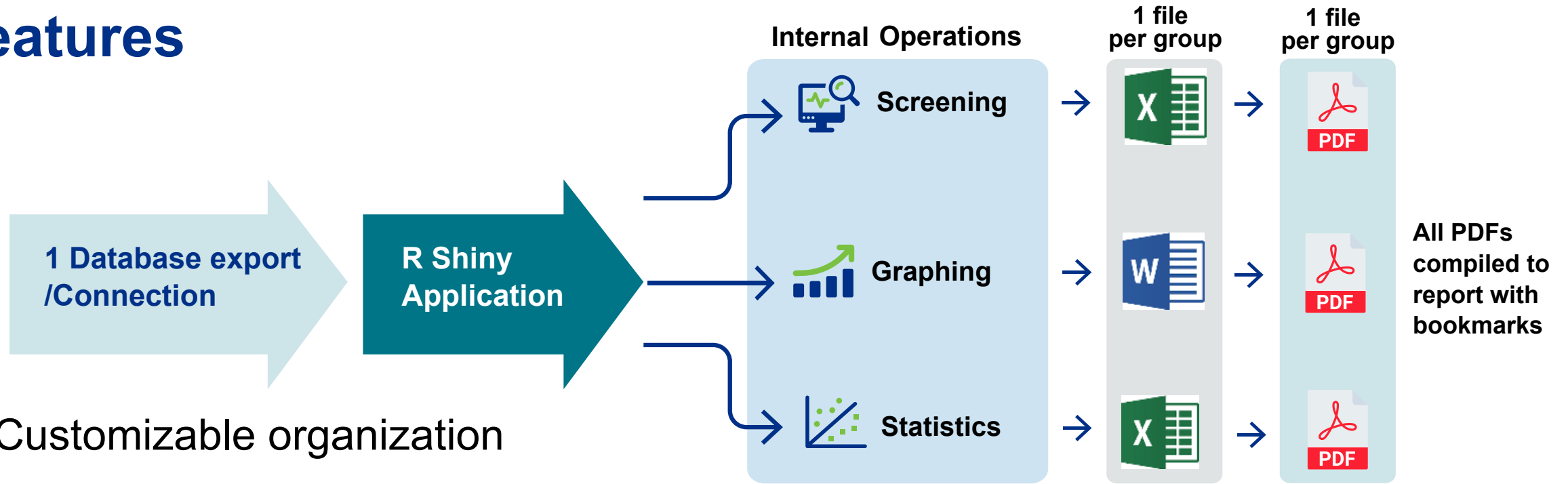
    # Sidebar panel for inputs ----
    sidebarPanel(
```

Application

- The entire Shiny tool
- What the developer builds
- Interface for non-programmers
- Various functionality and complexity depending on how built.
- Hosted on a Shiny server

Features

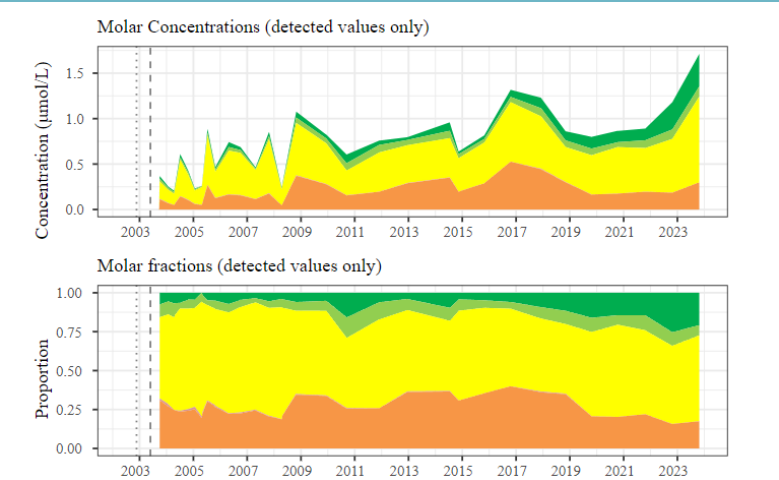
- Customizable organization
- Web-based, interactive
- Flexible data input
- Formatted outputs
- Protected code
- Automated, accurate, reproducible, efficient



Little to no manual formatting
Fewer files -> fewer QC steps
No manual bookmarking

Select Outputs

Visualization



Navigation

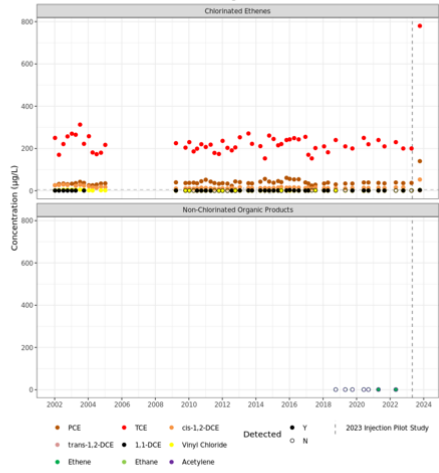
Search document

Headings Pages Results

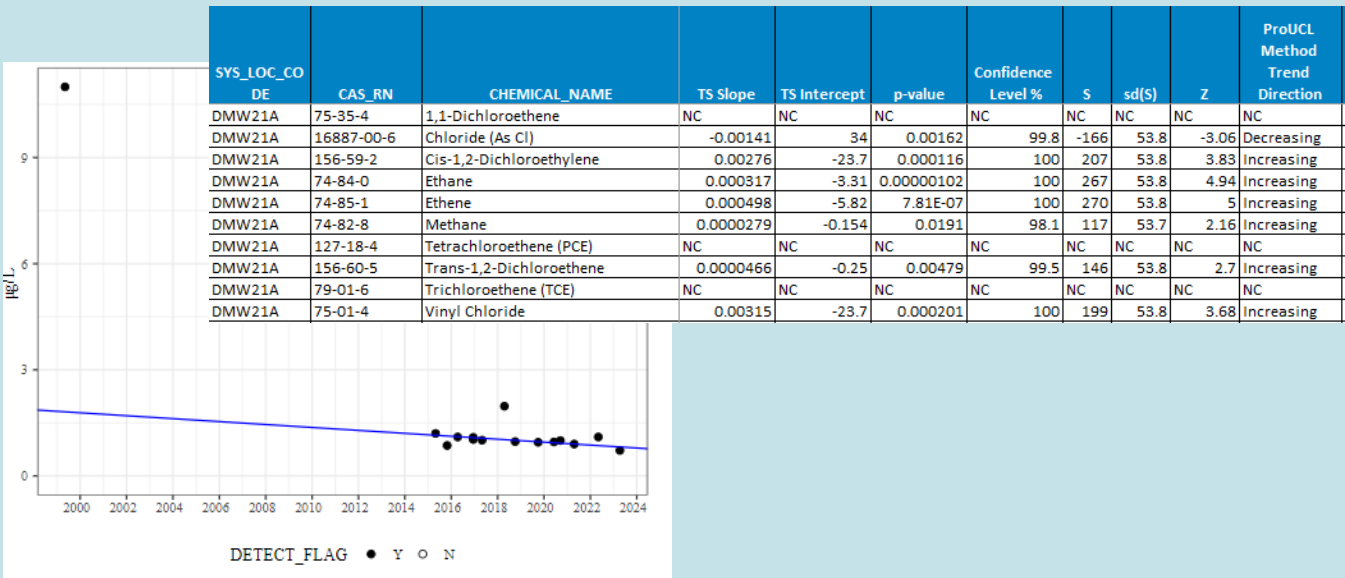
- Data filters
- Graph notes
- Well summaries
 - MP-1
 - Water Quality
 - TOC and Redox Parameters
 - Chlorinated Ethene and Daughter Product Mass
 - Chlorinated Ethene and Daughter Product Mass (log10)
 - Chlorinated Ethene and Daughter Product Concentration
 - MP-2
 - Water Quality
 - TOC and Redox Parameters
 - Chlorinated Ethene and Daughter Product Mass
 - Chlorinated Ethene and Daughter Product Mass (log10)
 - Chlorinated Ethene and Daughter Product Concentration
 - MP-3
 - Water Quality
 - TOC and Redox Parameters
 - Chlorinated Ethene and Daughter Product Mass
 - Chlorinated Ethene and Daughter Product Mass (log10)
 - Chlorinated Ethene and Daughter Product Concentration

Chlorinated Ethene and Daughter Product Mass

MW-10SD Chlorinated Ethene and Daughter Product



Statistics



Screening

CHEMICAL_NAME	Standard	Sample Name	CMI-DMW-21A-20231025		CMI-DMW-21A-20231025-L	
		Well ID	DMW21A		DMW21A	
		Sample Date	2023-10-25		2023-10-25	
		Unit	Result	Qualifier	Result	Qualifier
Cis-1,2-Dichloroethylene	70	µg/L	29		26	
Trans-1,2-Dichloroethene		µg/L	0.51	U	1.3	U
Vinyl Chloride	2	µg/L	59		51	

Other Opportunities: ProUCL Efficiency

1 Database export
/Connection

R Shiny
Application

Automate input formatting – all data

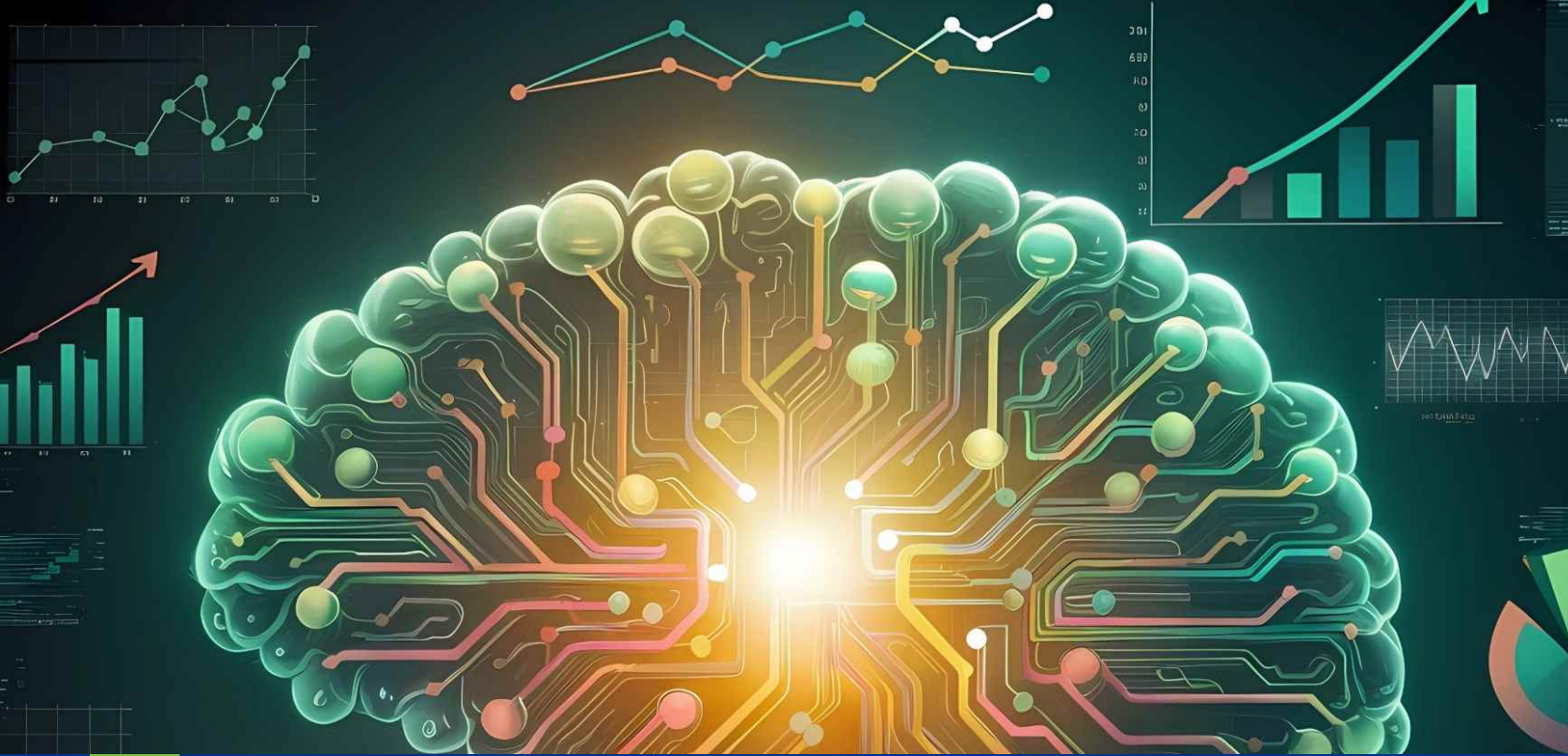
SYS_LOC_CODE	PCE	TCE	D_PCE	D_TCE
RW-2076F	15	210	1	1
RW-2076F	12	150	1	1
RW-2076F	15		1	
RW-2076F	6.4	90	1	1
RW-2076F	6.4	90	1	1
RW-2076F	15	130	1	1
RW-2076F	190		1	
RW-2076F	190	260	1	1
RW-2076F	290	290	1	1
RW-2076F	0.88	6.9	1	1
RW-2076F	0.66	6.9	1	1
RW-2076F	0.24	2.31	1	1
RW-2076F		1.3		1
RW-2076F	24	230	1	1
RW-2076F	14	160	1	1
RW-2076F	12	160	1	1
RW-2076F	13	160	1	1
RW-2076F		0.67		1

ProUCL Output – 1 dataset

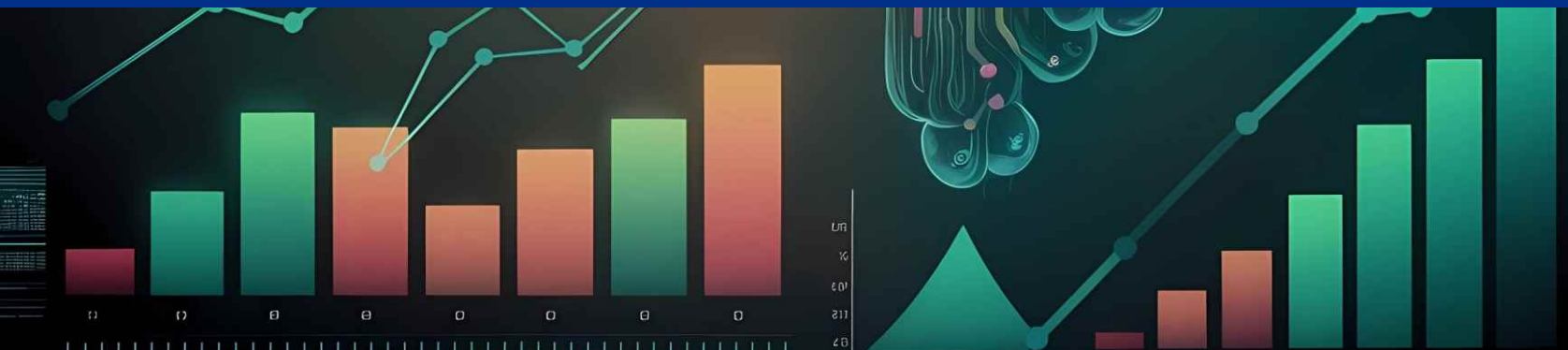
Mann-Kendall Trend Test Analysis			
User Selected Options			
Date/ Time of Computation	ProUCL 5.2 11/22/2023 4:46:46 PM		
Full Precision	OFF		
Confidence Coefficient	0.95		
Level of Significance	0.05		
Trichloroethene-rw-2076f			
General Statistics			
Number of Reported Events Not Used	0		
Number of Generated Events	47		
Number Values Reported (n)	50		
Number Values Missing	3		
Number Values Used	47		
Minimum	0.588		
Maximum	290		
Mean	67.05		
Geometric Mean	30.97		
Median	45		
Standard Deviation	71.08		
Coefficient of Variation	1.06		
Mann-Kendall Test			
M-K Test Value (S)	-405		
Critical Value (0.05)	-1.645		
Standard Deviation of S	109		
Standardized Value of S	-3.707		
Approximate p-value	1.0502E-4		
Statistically significant evidence of a decreasing trend at the specified level of significance.			

Automate output summaries – grouped datasets

Item	General Statistics			Mann-Kendall Test		
	Minimum	Maximum	Mean	M-K Test Value (S)	Approximate p-value	Result
Trichloroethene-rw-2076f	0.588	290	67.05	-405	1.0502E-4	Decreasing
Trichloroethene-rw-2140h	2.4	78	38.77	-74	4.9505E-4	Decreasing
Trichloroethene-rw-2151h	0.79	250	74.59	-57	3.0981E-4	Decreasing
Trichloroethene-rw-2172f	0.16	0.398	0.246	1	0.5	No trend
Trichloroethene-rw-2172h	16	207.1	72.08	-36	8.3586E-4	Decreasing
Trichloroethene-rw-2203h	13	250	86.5	-40	0.00116	Decreasing
Trichloroethene-rw-2213f	0.16	0.398	0.283	-4	0.217	No trend
Trichloroethene-rw-2237h	5.3	44.3	19.25	-11	0.106	No trend
Trichloroethene-rw-2879m	0.16	0.398	0.246	1	0.5	No trend
Trichloroethene-rw-5898b1	0.26	6.69	3.397	-45	0.00125	Decreasing
Trichloroethene-rw-6998p	0.14	0.2	0.175	1	0.5	No trend
Trichloroethene-rw-7070b	2.9	3.4	3.167	1	0.5	No trend
Trichloroethene-rw-7200p	0.16	0.3	0.253	2	0.27	No trend
Trichloroethene-rw-7244p	0.16	0.398	0.22	-3	0.186	No trend
Trichloroethene-rw-7350p	0.16	0.398	0.22	-3	0.186	No trend
Trichloroethene-rw-7505p	0.16	2.5	0.511	-14	0.155	No trend
Trichloroethene-rw-7549p	0.16	5.5	1.095	-51	0.0285	Decreasing
Trichloroethene-rw-7586p	0.16	0.398	0.22	-3	0.186	No trend
Trichloroethene-rw-7588p	0.16	0.398	0.22	-3	0.186	No trend
Trichloroethene-rw-7677p	2.95	33.8	14.43	-39	0.00979	Decreasing

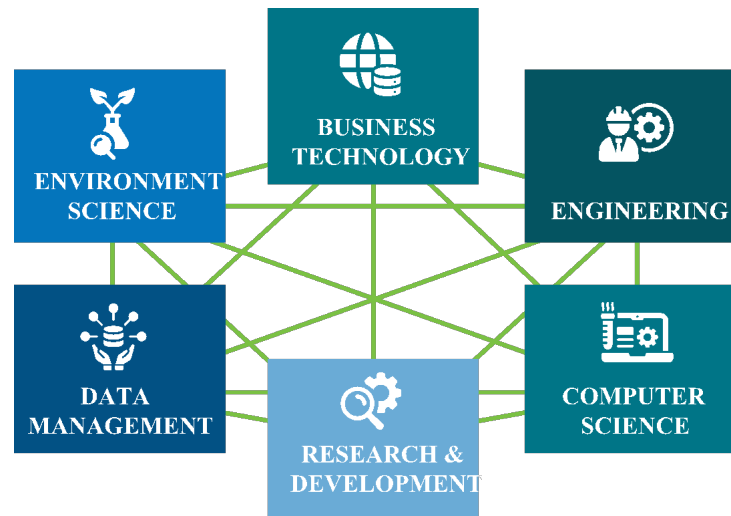


Lessons Learned and Takeaways

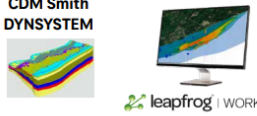


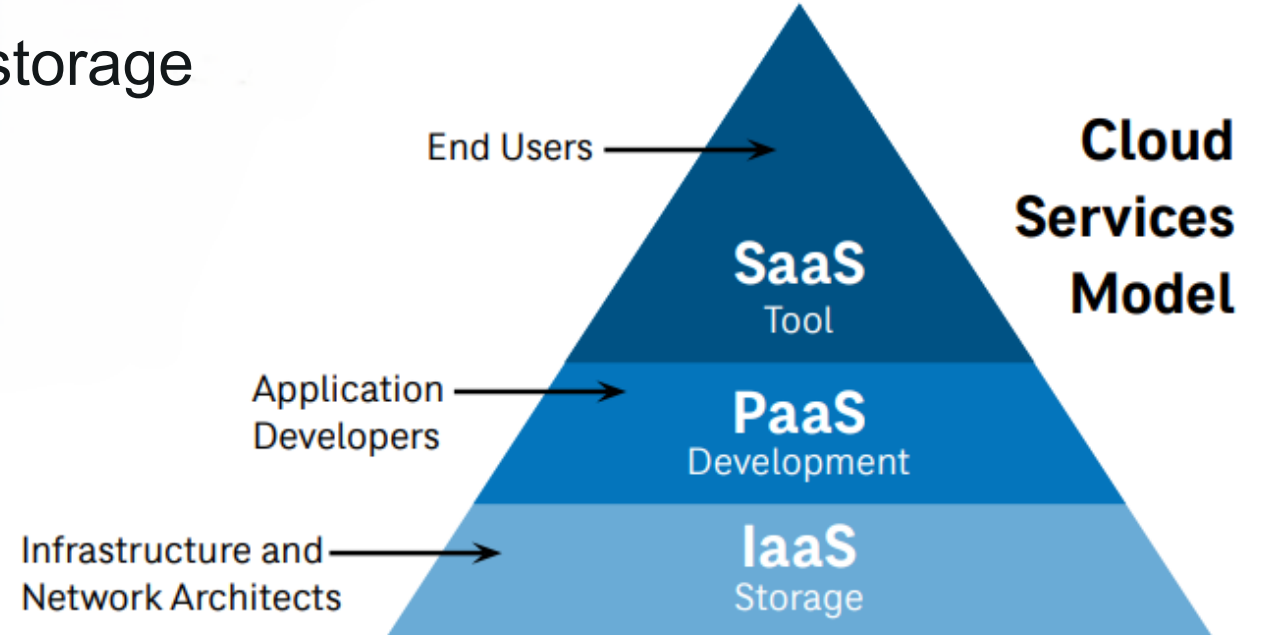
General Success Factors

- Understand available tools
- Align professional resources
- Collaborate
- Develop quality control procedures
- Data management and preparation
- Intellectual property protection and storage



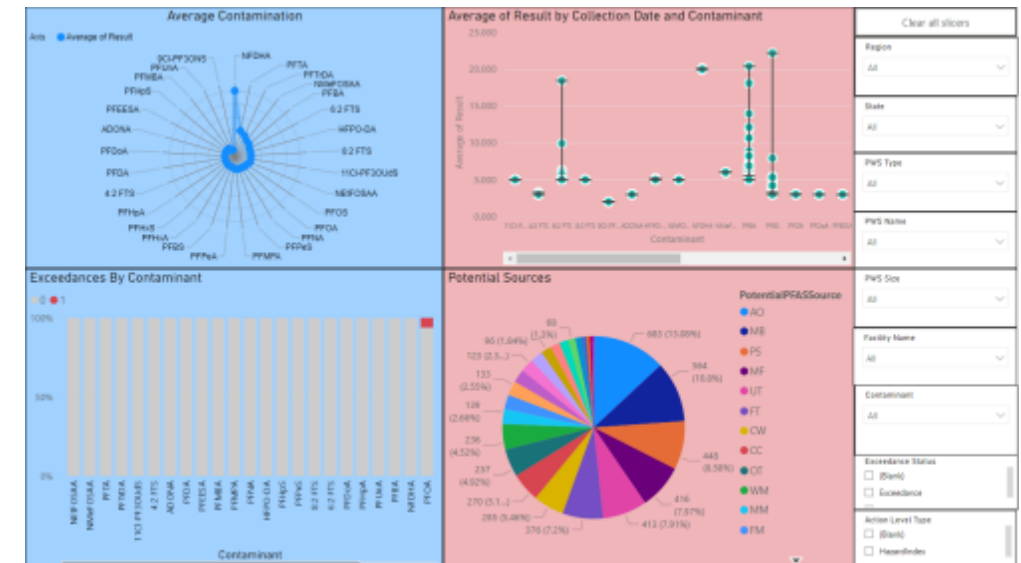
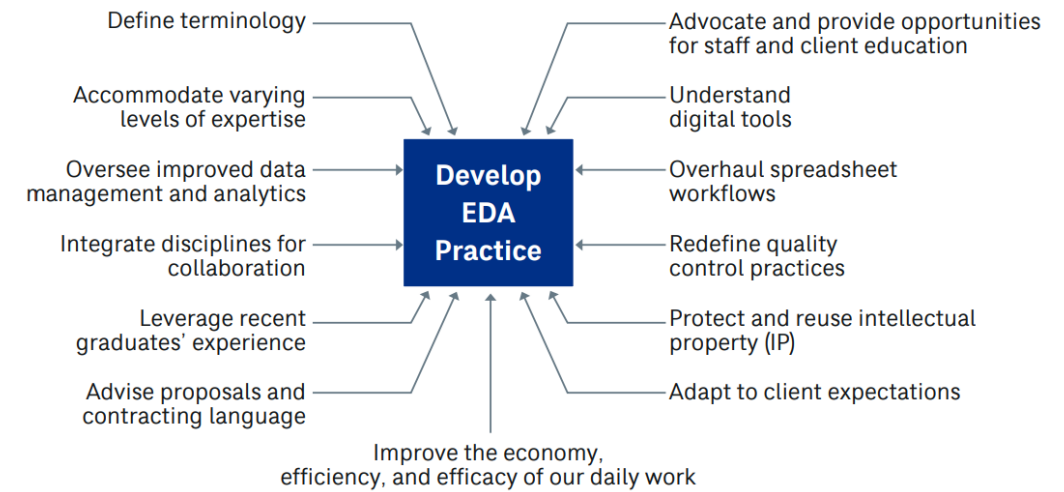
Tools

 SPREADSHEETS	 DATA COLLECTION	 SPATIAL
 RELATIONAL DATABASES	 NUMERICAL MODELS	 STATISTICAL
 PROGRAMMING LANGUAGES	 PROPRIETARY	 OPEN-SOURCE

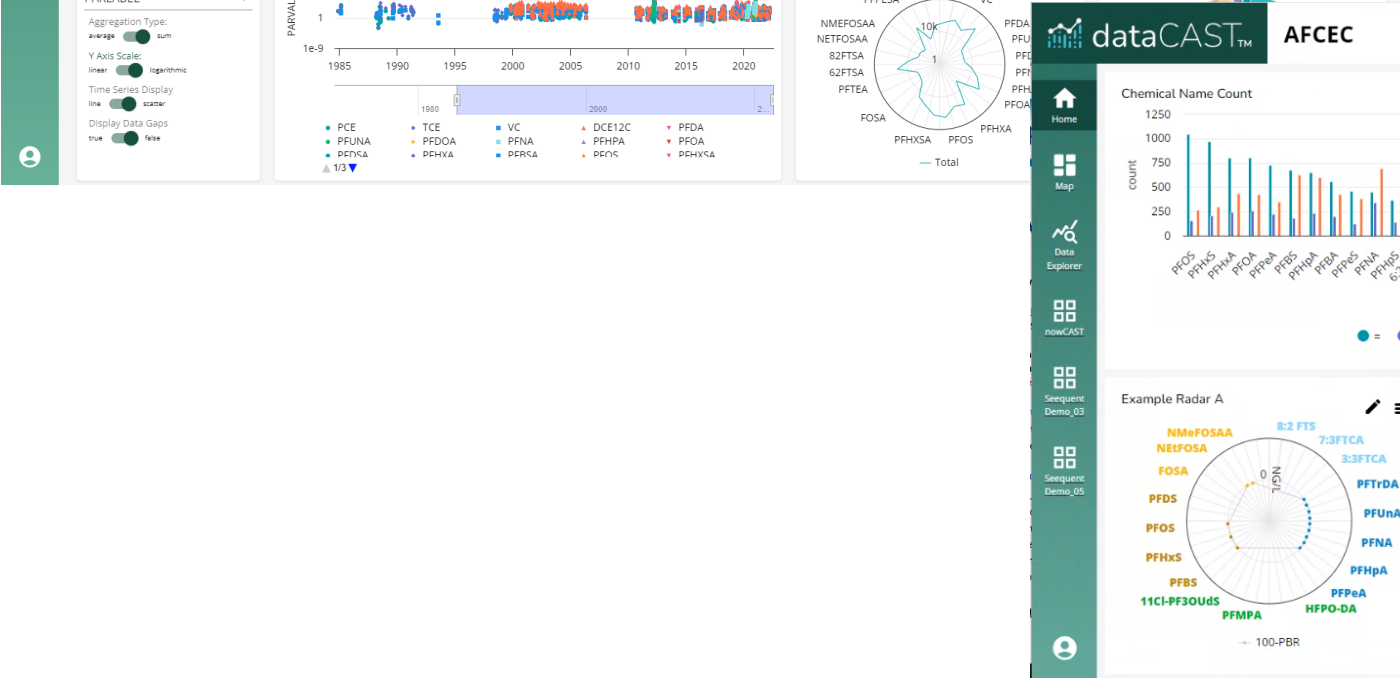
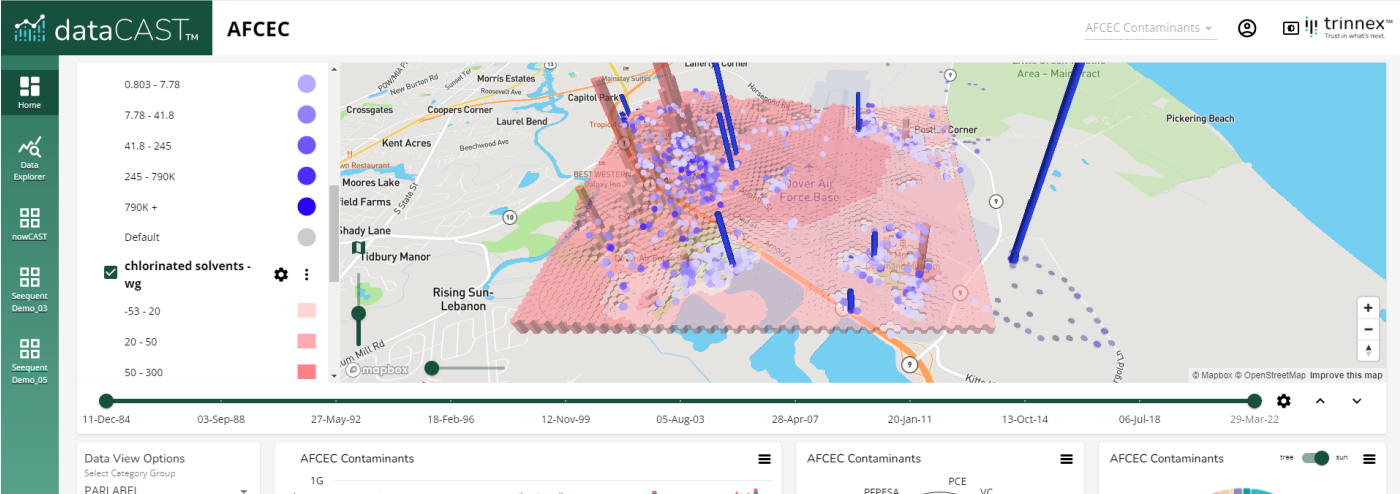


Environmental Data Analytics Practice (EDA)

- We don't build these tools in a vacuum.
- We need an EDA practice.
- Applicable AI/ML.
- Opportunities through collaboration
- Focus growth areas:
 - Cloud services
 - Quality
 - Intellectual property



dataCAST Tools



When optimize data analytics workflow?



High volume of data



Frequent, periodic data collection, analysis, reporting



Various regulatory/ comparison criteria



Multiple dataset groups/categories



Multiple lines of evidence feed into interpretation



Complex decision-making process



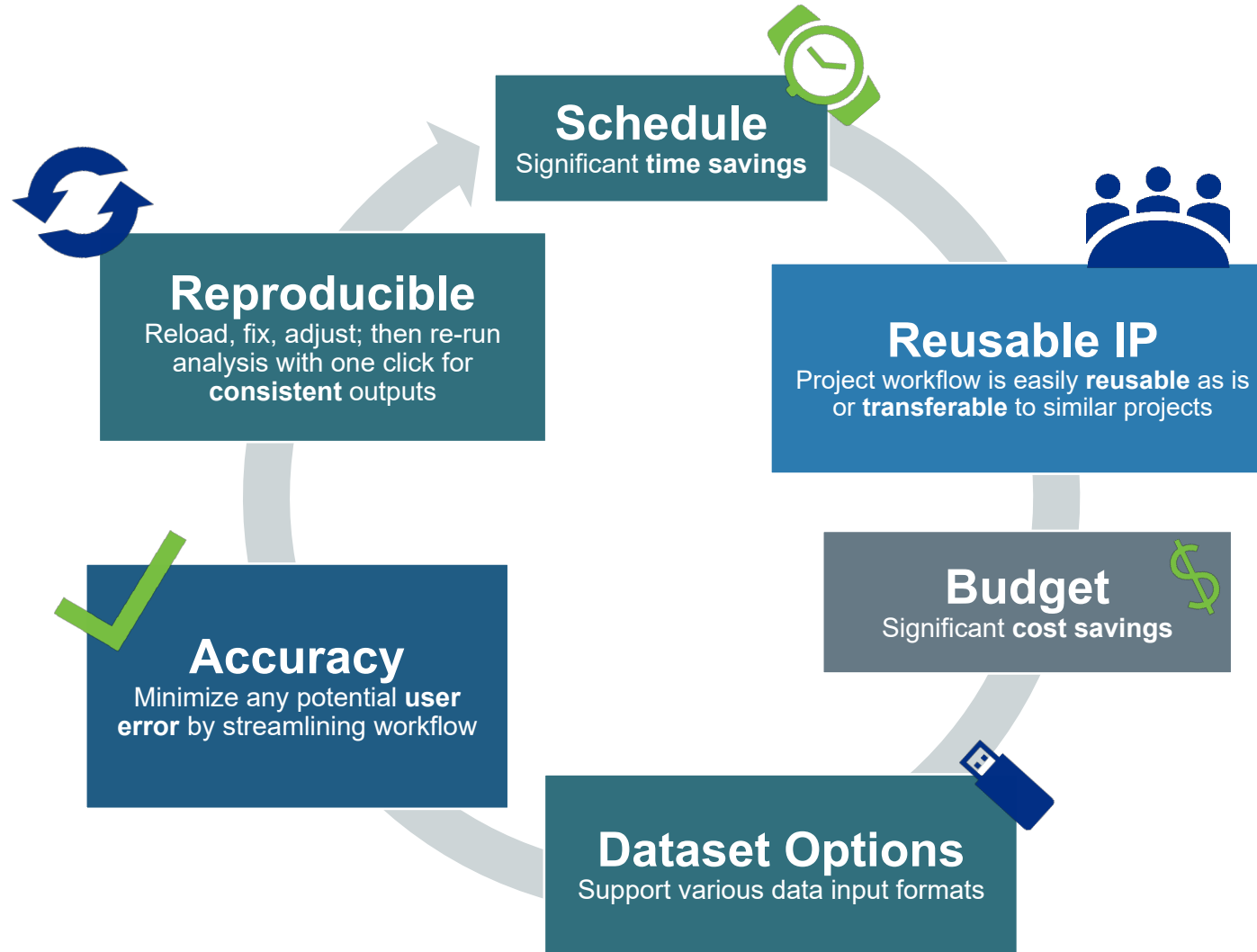
Engaged or questioning clients or stakeholders (need for clear documentation, robust rationale)



Client/stakeholder demands for interactive platforms

It's time to optimize our data analytics workflows for staff and clients.

Benefits of Optimizing Data Analytics Workflows



Thank you!



To learn more, contact:

Emma Ehret, PE

CDM Smith

EhretEL@cdmsmith.com

208-417-2279

