verdantas

Exploring AI-Driven Innovations in Environmental Data Management

The 14th Annual International Commission on Environmental Data Management (ICEDM)

Tim Negley and Jamie Combes

May 14, 2025

Disclaimer

All data shown in this presentation are fictional and intended for illustration only. Any resemblance to an actual site is purely coincidental.

Role of AI in Environmental Data Management

What is AI?

- Artificial Intelligence (AI) refers to computer systems capable of performing tasks that typically require human intelligence, such as:
  - Data collection and integration
  - Data cleaning and pre-processing
  - Data analytics and visualization
  - Predictive analytics
  - Automation of routine tasks
- AI encompasses a wide range of technologies, including machine learning, deep learning, and natural language processing (NLP)
- Potential to transform our industry by improving efficiency, quality, and creating new opportunities.

Historical Context & Evolution of AI

Early Beginnings (1950s-1970s)

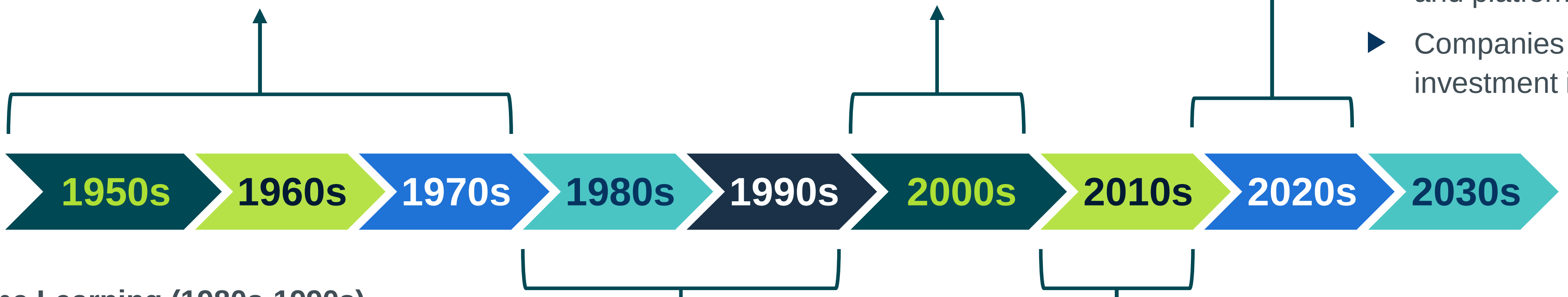
- Concept of artificial intelligence introduced, focusing on symbolic AI & rule-based systems
- Basic data processing techniques developed using statistical methods & simple algorithms

Big Data Era (2000s)

- Big data technologies (Hadoop, NoSQL databases) enable storage/processing of massive datasets
- ML algorithms become more sophisticated

Deep Learning & AI Democratization (2020s)

- Deep learning models revolutionize data analysis
- GPT-3 and GPT-4 natural language processing
- Development of user-friendly interfaces and platforms
- Companies significantly increase investment in AI technologies



Emergence of Machine Learning (1980s-1990s)

- Machine Learning (ML) algorithms introduced
- Data warehousing emerges, allowing storage of large volumes of data
- AI-driven expert systems are being used to make decisions based on predefined rules & knowledge bases

AI & Cloud Computing (2010s)

- Cloud computing provides scalable infrastructure for data storage & processing
- AI integrated into data management platforms, enabling automated data cleaning, preprocessing, & analysis
- Real-time analytics possible, driven by advancements in streaming data technologies

AI is Different than Traditional Programming

Traditional Programming

- Rule-based approach: Programmer uses human intelligence to solve the problem first then gives the computer step-by-step instructions to follow.
- Works well when logic is clearly defined.
- Errors must be manually detected and fixed by the programmer.
- Requires human intervention to apply to new applications.

Artificial Intelligence

- Often conflated with traditional programming.
- Enabling machines to perform tasks that typically require human intelligence.
- Continuously improves by analyzing data and recognizing patterns.
- Trained on massive amounts of data.
- Well suited for complex problems where defining explicit rules is difficult. Continues trying options until it finds a solution, building its own intelligence.

Opportunities for Data Management

Environmental data managers have traditionally faced a range of challenges stemming from the complexity, volume, and variability of environmental data.

Challenge	Opportunity for AI
Data routinely needs cleaning, which is laborious and prone to human error.	Automate data cleaning
Data originating from disparate sources in incompatible formats can be laborious to normalize.	Write programs to normalize data
Migrating legacy data files (e.g., PDFs) into centralized databases can be time-consuming and prone to human error.	Find analytical data Data extraction and reshaping workflows
Data quality issues sometimes do not become apparent until the data are visualized.	Data QC and visualization

1. AI for Data Cleaning

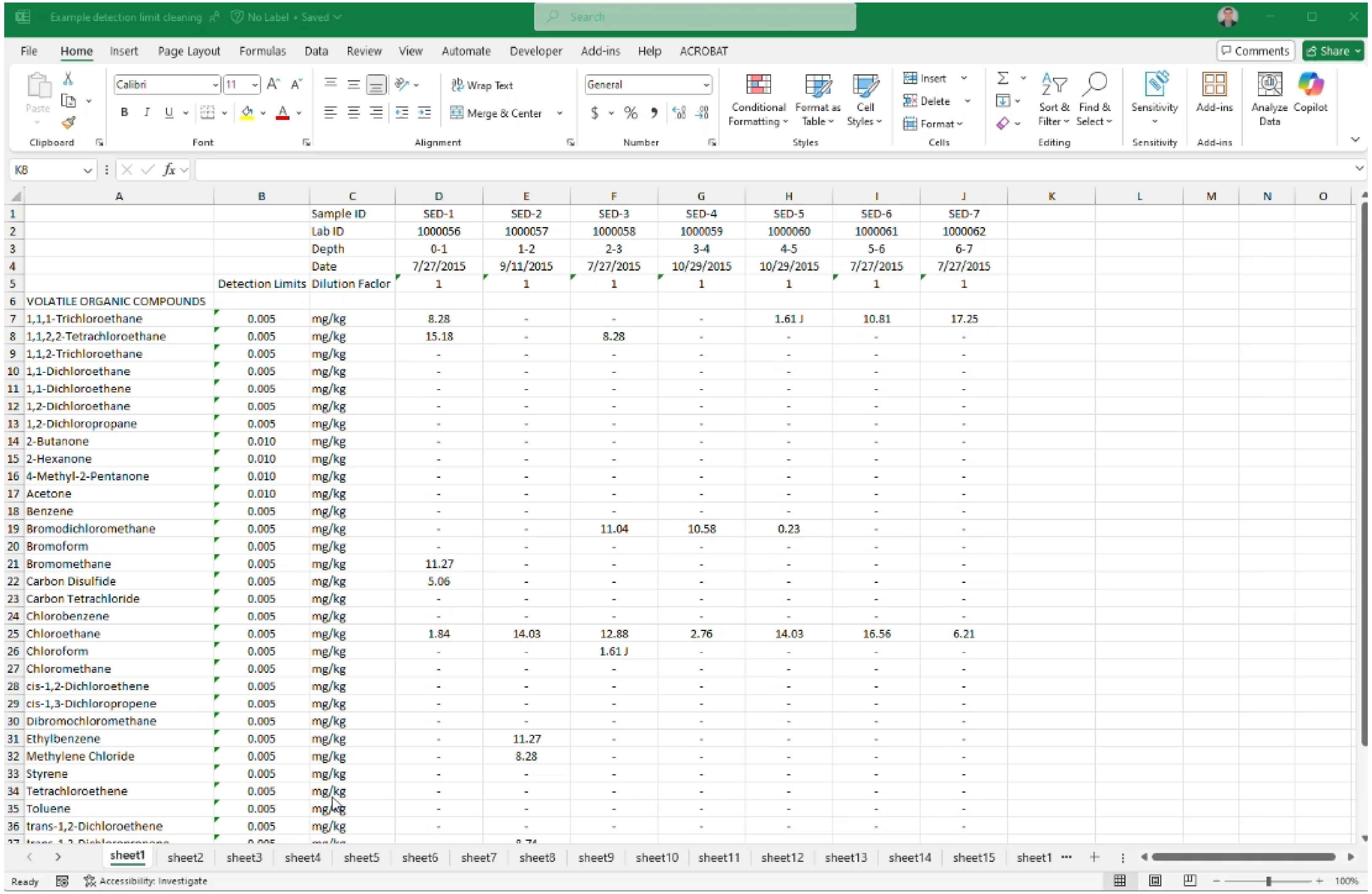
1. Clean chemical names
2. Identify duplicates
3. Identify T/D fraction
4. Split results and qualifiers
5. Harmonize units

Issue	Location	Sample	Matrix	Date	Chemical Name	Result and Qualifier	Result Unit	Dete
Extraneous information in chemical names	MW-1	MW-1_20180930	Groundwater	9/30/2018	DO (mg/L)	12400	mg/L	Y
	MW-1	MW-1_20130925	Groundwater	9/25/2013	1,2-Dichloroethene (1,2-DCE)	4620	mg/L	Y
	MW-1	MW-1_20160928	Groundwater	9/28/2016	DO (mg/L)	3720	mg/L	Y
	MW-1	MW-1_20170929	Groundwater	9/29/2017	DO (mg/L)	1540	mg/L	Y
Duplicate Records	MW-1	MW-1_20170929	Groundwater	9/29/2017	1,2-Dichloroethene (dissolved) (1,2-DCE)	1410	ug/L	Y
	MW-1	MW-1_20170929	Groundwater	9/29/2017	1,2-Dichloroethene (dissolved) (1,2-DCE)	1410	ug/L	Y
	MW-1	MW-1_20170929	Groundwater	9/29/2017	cis-1,2-Dichloroethene	1410	ug/L	Y
Duplicate Records	MW-1	MW-1_20120830	Groundwater	8/30/2012	Tetrahydrofuran (THF)	440	ug/L	Y
	MW-1	MW-1_20120830	Groundwater	8/30/2012	Tetrahydrofuran (THF)	440	ug/L	Y
	MW-1	MW-1_20170929	Groundwater	9/29/2017	Vinyl Chloride	1410	ug/L	Y
	MW-1	MW-1_20130925	Groundwater	9/25/2013	Vinyl Chloride	370	ug/L	Y
Total or dissolved embedded in chemical name	MW-1	MW-1_20180930	Groundwater	9/30/2018	1,2-Dichloroethene (dissolved) (1,2-DCE)	210	ug/L	Y
	MW-1	MW-1_20110421	Groundwater	4/21/2011	1,2-Dichloroethene (total) (1,2-DCE)	240	ug/L	Y
	MW-1	MW-1_20180930	Groundwater	9/30/2018	cis-1,2-Dichloroethene (dissolved)	280	ug/L	Y
	MW-1	MW-1_20130403	Groundwater	4/3/2013	Vinyl Chloride	210	ug/L	Y
Results and qualifiers merged	MW-1	MW-1_20130403	Groundwater	4/3/2013	1,2-Dichloroethene (dissolved) (1,2-DCE dissolved)	930 T	ug/L	Y
	MW-1	MW-1_20110421	Groundwater	4/21/2011	cis-1,2-Dichloroethene	5 UJ	ug/L	Y
	MW-1	MW-1_20130403	Groundwater	4/3/2013	cis-1,2-Dichloroethene	1,100 E	ug/L	Y
	MW1	MW1_20130925	Groundwater	9/25/2013	1,2-Dichloroethene (dissolved) (1,2-DCE)	1810 B	ug/L	Y
	MW-1	MW-1_20130925	Groundwater	9/25/2013	Tetrahydrofuran (THF)	240	ug/L	Y
Inconsistent concentration units	MW-1	MW-1_20150927	Groundwater	9/27/2015	Vinyl Chloride	210000	ng/L	Y
	MW-1	MW-1_20140922	Ground watei	9/22/2014	Tetrahydrofuran (THF)	0.210	Mg/L	Y
	MW-1	MW-1_20120830	Groundwater	8/30/2012	1,2-Dichloroethene (dissolved) (1,2-DCE)	210	ug/L	Y
	MW-1	MW-1_20120830	Groundwater	8/30/2012	cis-1,2-Dichloroethene	0.210	mg/L	Y
	MW-1	MW-1_20130403	Groundwater	4/3/2013	Tetrahydrofuran (THF)	200	ug/L	Y
	MW-1	MW-1_20170929	Groundwater	9/29/2017	Toluene	200	ug/L	Y
	MW-1	MW-1_20170929	Groundwater	9/29/2017	Xylenes (dissolved)	190	ug/L	Y
	MW-1	MW-1_20100422	Groundwater	4/22/2010	Tetrahydrofuran (THF)	170	ug/L	Y
	MW-1	MW-1_20170929	Groundwater	9/29/2017	Chloroethane	170	ug/L	Y
	MW-1	MW-1_20100126	Groundwater	1/26/2010	Tetrahydrofuran (THF)	150	ug/L	Y
	MW-1	MW-1_20110421	Groundwater	4/21/2011	Tetrahydrofuran (THF)	140	ug/L	Y
	MW-1	MW-1_20130925	Groundwater	9/25/2013	1,1-Dichloroethane (1,1-DCA)	140	ug/L	Y
	MW-1	MW-1_20170929	Groundwater	9/29/2017	1,1-Dichloroethane (1,1-DCA)	140	ug/L	Y
	MW-1	MW-1_20120128	Groundwater	1/28/2012	Tetrahydrofuran (THF)	110	ug/L	Y
	MW-1	MW-1_20130403	Groundwater	4/3/2013	1,1-Dichloroethane (1,1-DCA)	93	ug/L	Y
	MW-1	MW-1_20140922	Groundwater	9/22/2014	1,2-Dichloroethene(dissolved) (1,2-DCE dissolved)	89 T	ug/L	Y
	MW-1	MW-1_20140922	Groundwater	9/22/2014	cis-1,2-Dichloroethene	87	ug/L	Y

2. AI for Writing Data Management Scripts

Leveraging AI in place of your traditional programming tasks.

Example: Use AI to populate detection limits in messy report tables.

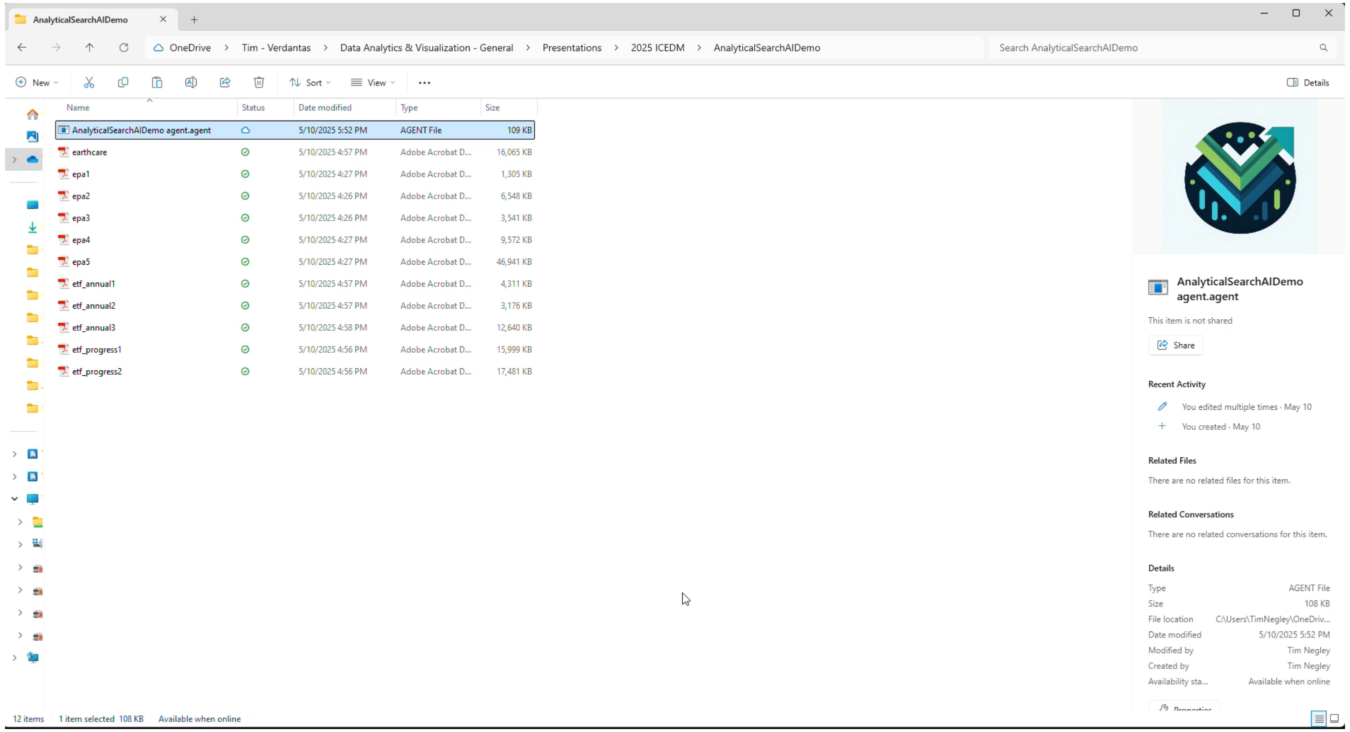


	A	B	C	D	E	F	G	H	I	J
1			Sample ID	SED-1	SED-2	SED-3	SED-4	SED-5	SED-6	SED-7
2			Lab ID	1000056	1000057	1000058	1000059	1000060	1000061	1000062
3			Depth	0-1	1-2	2-3	3-4	4-5	5-6	6-7
4			Date	7/27/2015	9/11/2015	7/27/2015	10/29/2015	10/29/2015	7/27/2015	7/27/2015
5		Detection Limits	Dilution Factor	1	1	1	1	1	1	1
6	VOLATILE ORGANIC COMPOUNDS									
7	1,1,1-Trichloroethane	0.005	mg/kg	8.28	-	-	-	1.61 J	10.81	17.25
8	1,1,2,2-Tetrachloroethane	0.005	mg/kg	15.18	-	8.28	-	-	-	-
9	1,1,2-Trichloroethane	0.005	mg/kg	-	-	-	-	-	-	-
10	1,1-Dichloroethane	0.005	mg/kg	-	-	-	-	-	-	-
11	1,1-Dichloroethene	0.005	mg/kg	-	-	-	-	-	-	-
12	1,2-Dichloroethane	0.005	mg/kg	-	-	-	-	-	-	-
13	1,2-Dichloropropane	0.005	mg/kg	-	-	-	-	-	-	-
14	2-Butanone	0.010	mg/kg	-	-	-	-	-	-	-
15	2-Hexanone	0.010	mg/kg	-	-	-	-	-	-	-
16	4-Methyl-2-Pentanone	0.010	mg/kg	-	-	-	-	-	-	-
17	Acetone	0.010	mg/kg	-	-	-	-	-	-	-
18	Benzene	0.005	mg/kg	-	-	-	-	-	-	-
19	Bromodichloromethane	0.005	mg/kg	-	-	11.04	10.58	0.23	-	-
20	Bromoform	0.005	mg/kg	-	-	-	-	-	-	-
21	Bromomethane	0.005	mg/kg	11.27	-	-	-	-	-	-
22	Carbon Disulfide	0.005	mg/kg	5.06	-	-	-	-	-	-
23	Carbon Tetrachloride	0.005	mg/kg	-	-	-	-	-	-	-
24	Chlorobenzene	0.005	mg/kg	-	-	-	-	-	-	-
25	Chloroethane	0.005	mg/kg	1.84	14.03	12.88	2.76	14.03	16.56	6.21
26	Chloroform	0.005	mg/kg	-	-	1.61 J	-	-	-	-
27	Chloromethane	0.005	mg/kg	-	-	-	-	-	-	-
28	cis-1,2-Dichloroethene	0.005	mg/kg	-	-	-	-	-	-	-
29	cis-1,3-Dichloropropene	0.005	mg/kg	-	-	-	-	-	-	-
30	Dibromochloromethane	0.005	mg/kg	-	-	-	-	-	-	-
31	Ethylbenzene	0.005	mg/kg	-	11.27	-	-	-	-	-
32	Methylene Chloride	0.005	mg/kg	-	8.28	-	-	-	-	-
33	Styrene	0.005	mg/kg	-	-	-	-	-	-	-
34	Tetrachloroethene	0.005	mg/kg	-	-	-	-	-	-	-
35	Toluene	0.005	mg/kg	-	-	-	-	-	-	-
36	trans-1,2-Dichloroethene	0.005	mg/kg	-	-	-	-	-	-	-

3. AI Agents for Finding PDFs with Analytical Data

Copilot Agents

- AI-powered digital assistant embedded within Microsoft
- Allows user to create customized prompts
- Benefits:
  - Scalable
  - Automates repetitive tasks
  - Reduces human error
  - Fully customizable



Copilot

4. Case Studies: Practical Implementation & Stakeholder Empowerment

Case Study #1: High profile litigation case – PCBs & BEHP

- Client is looking to understand sources of contamination on their property and neighboring properties
- Data only available in PDF – manual entry was taking too much time
- Used AI to extract data from lab reports in 17 different formats
- Approx. 3,500 pages of data, extracted approx. 21,000 results

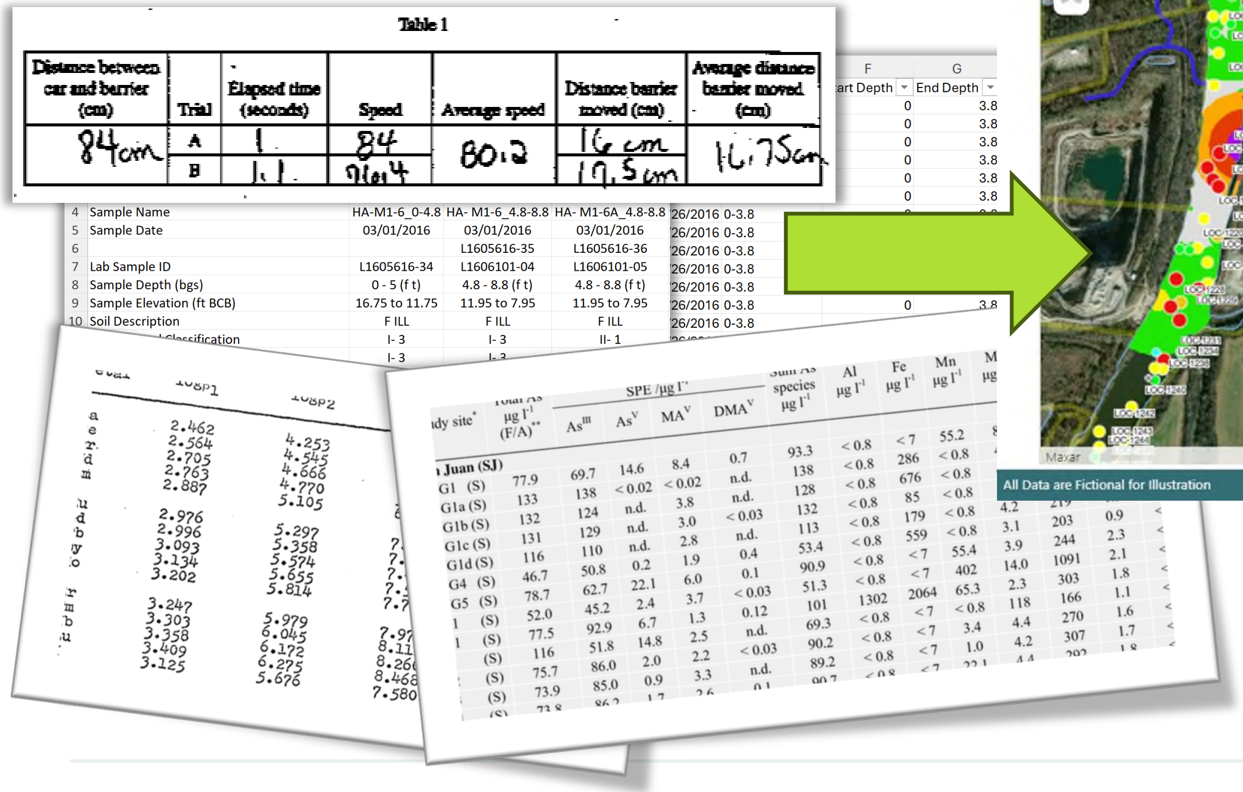
Case Study #2: PFAS litigation case

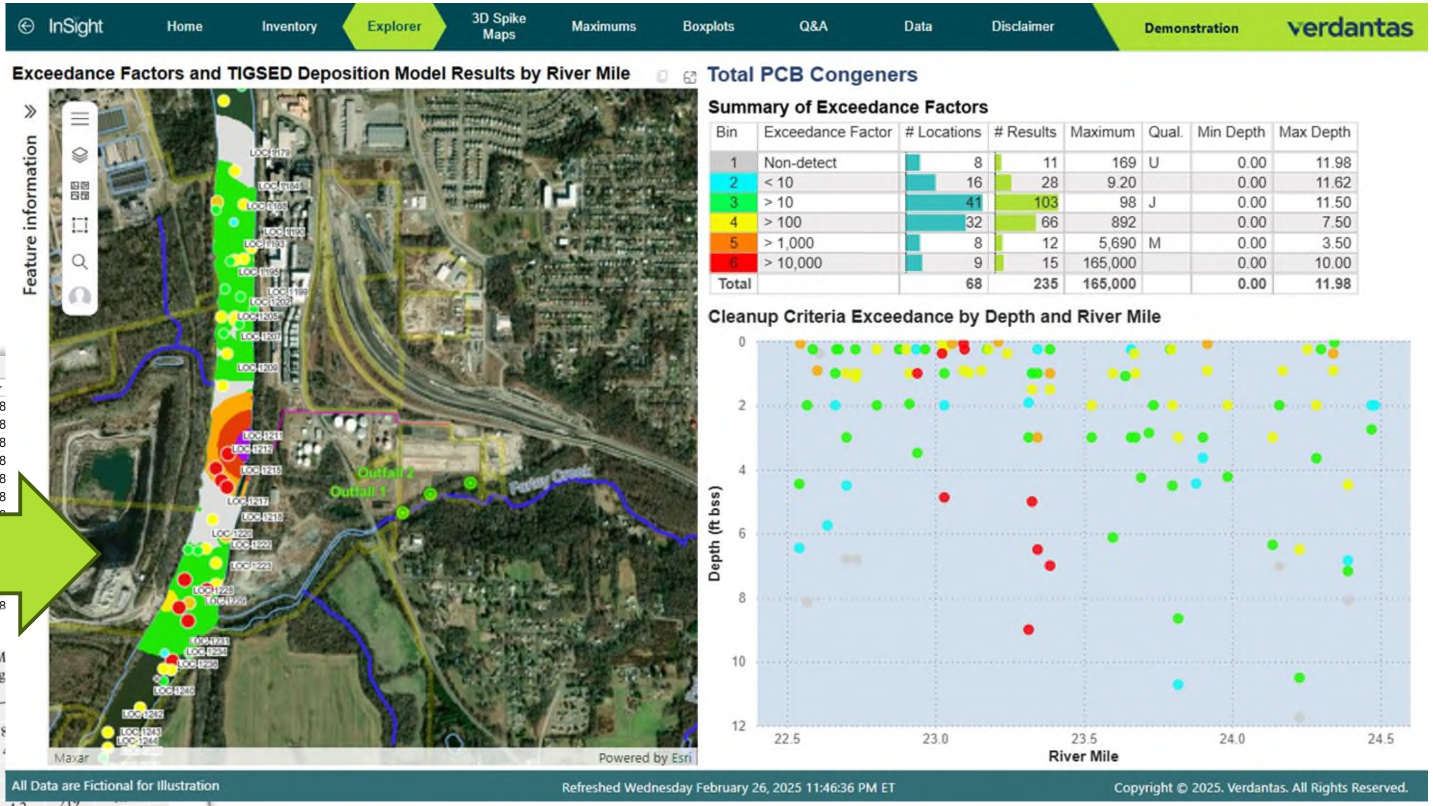
- Client is looking to compile information from production batch sheets, recipes, and sales invoices
- Used AI to extract data from multiple document types
- Many were hand-written documents
- Approx. 120,000 documents

InSight Data Visualization Dashboards

Leverage TIGER & AIDE to find and extract the data

- Algorithm picks out tabular data amongst collections of PDF and/or TIF files





Use InSight to visualize the data

- Data are uploaded to customizable dashboards
- Provides real-time access to data and facilitates evaluation and analysis

InSight Data Visualization Dashboards

AI in Environmental Remediation

Advances in AI are completely disrupting the AEC industry, and analytics is no exception. We are leveraging these advances to increase the analytical capabilities of employees across the Verdantas platform.

Future Directions and Considerations

Emerging trends and future directions in AI-driven EDM

- AI Governance: Establishing robust governance frameworks to oversee AI development and deployment.
- Collaborative AI: Enhancing human-AI collaboration to leverage the strengths of both.
- Sustainable AI: Developing AI technologies that are environmentally sustainable and energy-efficient.
- Explainable AI (XAI): There is a growing emphasis on making AI models interpretable and explainable to ensure transparency and trust.
- Edge AI: AI is moving to the edge, enabling real-time analytics on devices such as IoT sensors and mobile phones.
- Quantum Computing: The potential of quantum computing to solve complex data analysis problems is being explored.

verdantas

Thank you

Tim Negley

Senior Consultant, Data Analytics and Visualization

tnegley@verdantas.com

Jamie Combes

Senior Consultant, Data Management

jcombes@verdantas.com