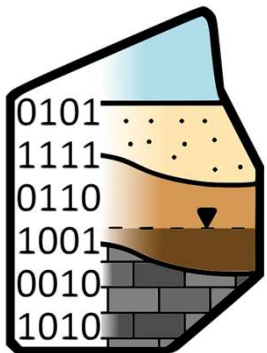


STGEC

Southeastern Transportation
Geotechnical Engineering Conference

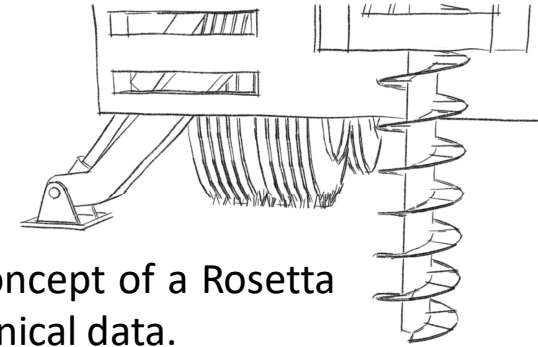
Building a Repository of Public Geotechnical Data



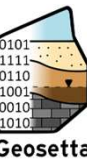
Geosetta



What is Geosetta?



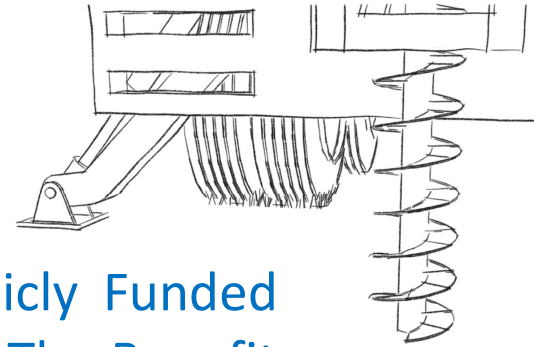
- A non-profit Maryland based company. The name gets its inspiration from the concept of a Rosetta stone that unlocks access to the tremendous amount of available historic geotechnical data.
- Provides a platform for hosting subsurface /geotechnical data from various publicly funded sources throughout the United States.
- Provides a preliminary understanding of anticipated subsurface conditions at any geographic area, thus assisting in the design of a cost effective and efficient subsurface exploration program.
- Geosetta has developed geospatial and visualization tools, with machine learning techniques applied.
- Geosetta is NOT a substitute for site-specific subsurface investigation



Geosetta

What Is Our Mission?

Build a Repository of Subsurface Investigation Data from Publicly Funded Sources and Derive Valuable Geotechnical Engineering Tools for The Benefit the Geotechnical and Civil Engineering Profession.



Geosetta is partnering with and contributing to the development of DIGGS.



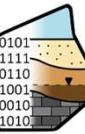
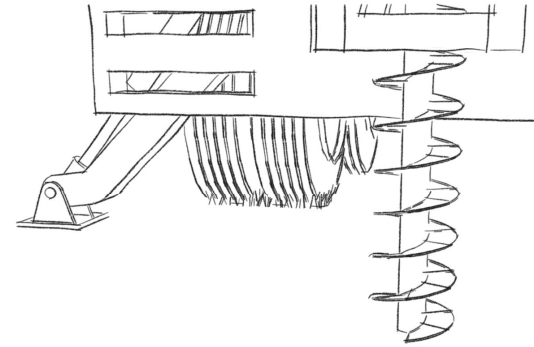
Why is Historic Data Important?



IMPORTANT NOTE:

Why a non-profit?

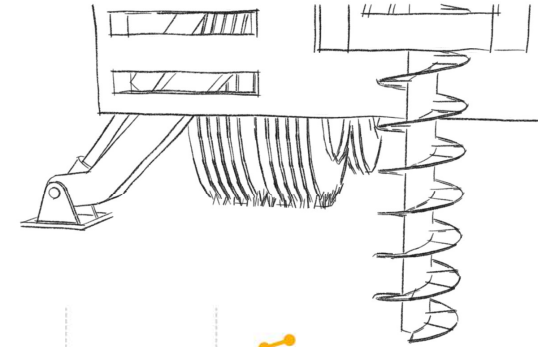
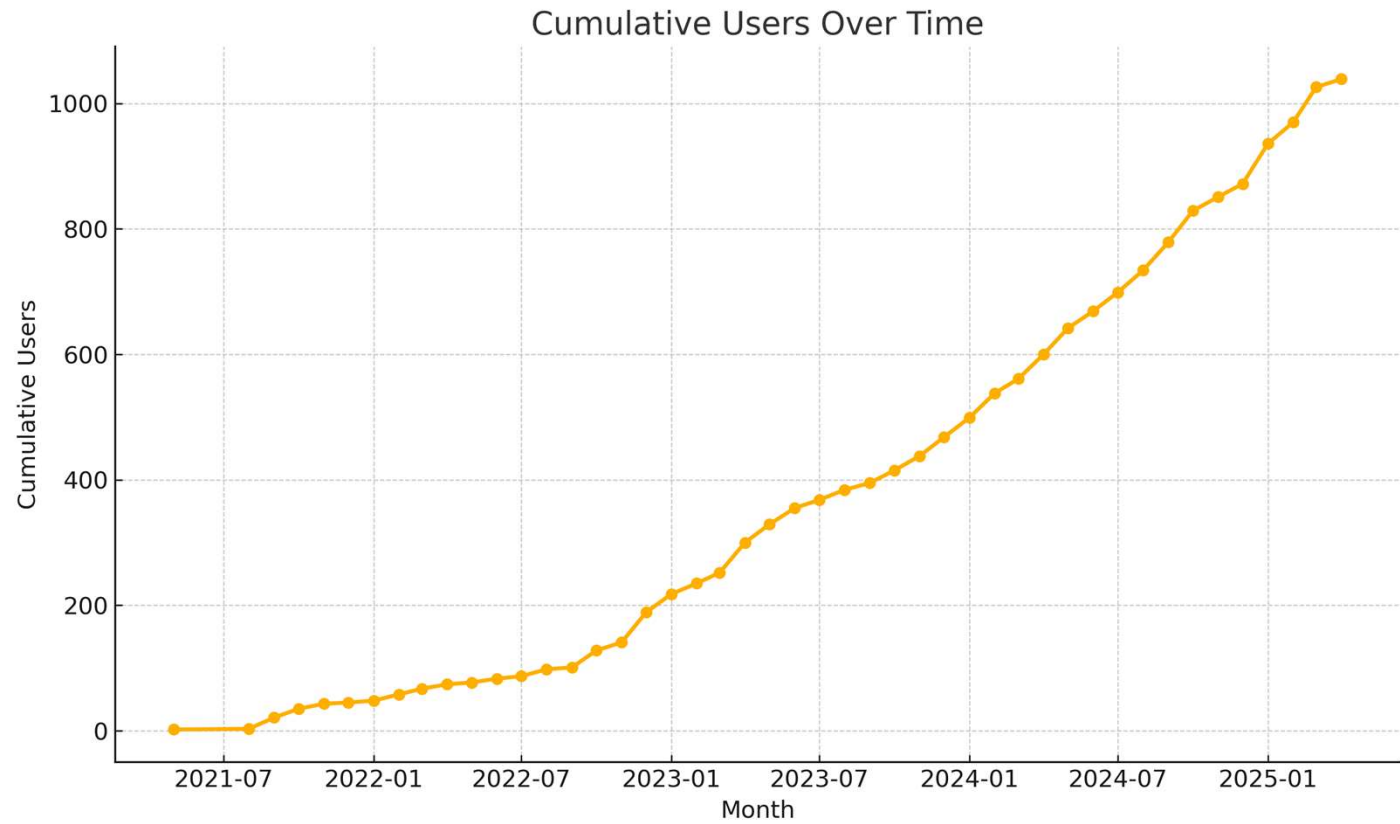
A non-profit structure ensures that Geosetta remains a resource for our profession. This structure means that Geosetta cannot be acquired by a larger firm or start competing with other firms. A non-profit gives trust that we are an independent steward of their data.



Geosetta

Geosetta in Numbers:

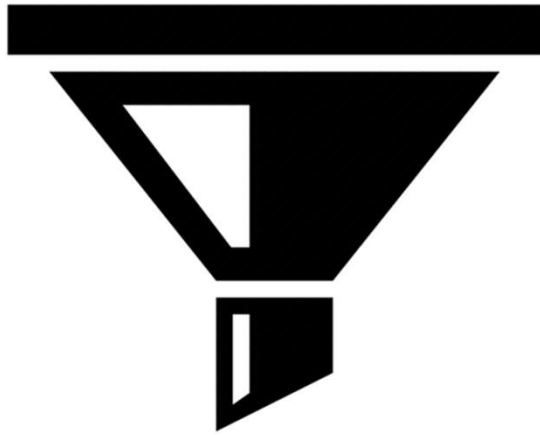
Constant Growth:
We recently broke
1,000 active users!



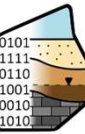
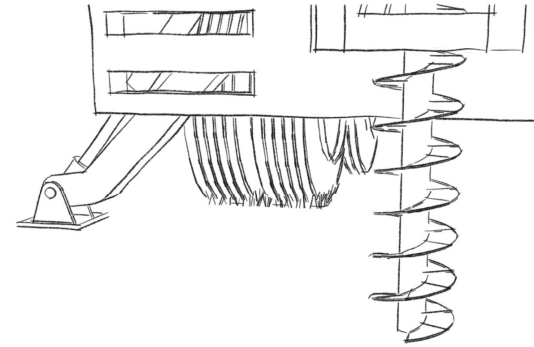
Geosetta

Geosetta in Numbers:

Over 400,000 exploration points from
28 Contributing Agencies.

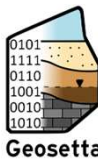
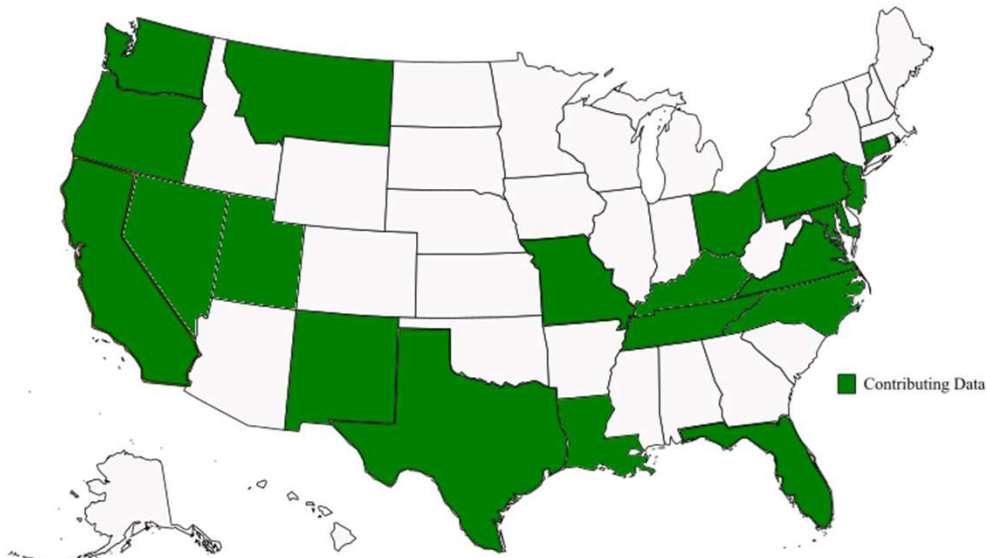


Converted and delivered
into ONE interchangeable
format

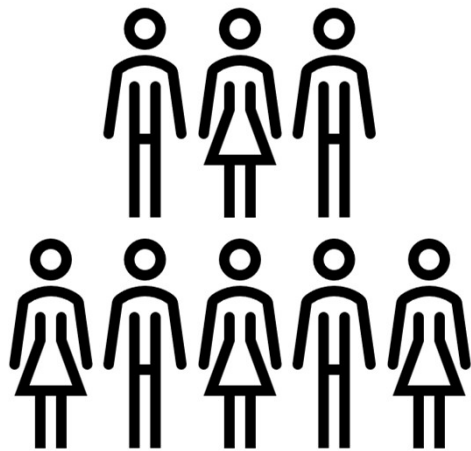
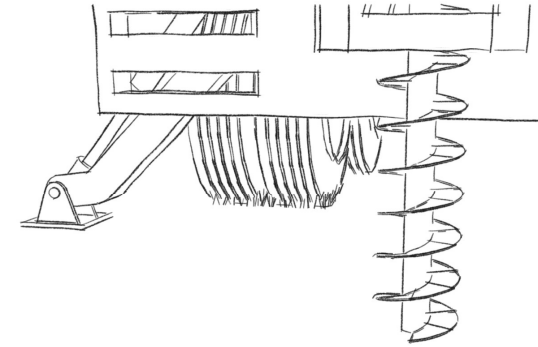


Geosetta

Who is contributing Data?



Who is contributing Data?



All of us!

Our web crawler identifies publicly available geotechnical reports

- Discovers only what's already public online
- Automatically geolocates data
- Converts isolated documents into standardized, searchable data
- Transforms public records into collective industry knowledge



Geosetta

Who Benefits from this Data?

Academic Community

- Researchers conducting regional geological studies
- Faculty developing case studies for engineering courses
- Students working on thesis projects and design competitions

Practitioners

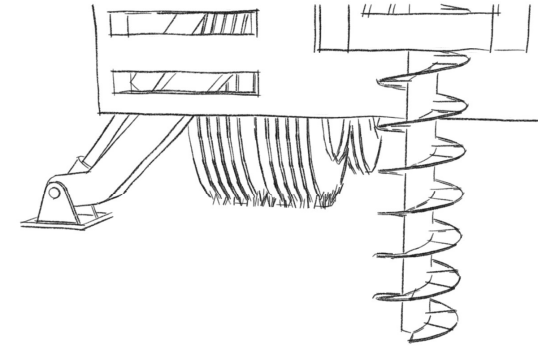
- Engineers establishing preliminary site conditions
- Consultants optimizing exploration plans
- Construction professionals assessing regional challenges

Project Owners

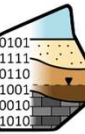
- Developers evaluating potential site constraints
- Infrastructure managers planning maintenance and upgrades
- Property owners understanding regional geological risks

Public Agencies

- Transportation departments planning infrastructure projects
- Municipal engineers evaluating development risks
- Emergency management officials assessing natural hazard potential

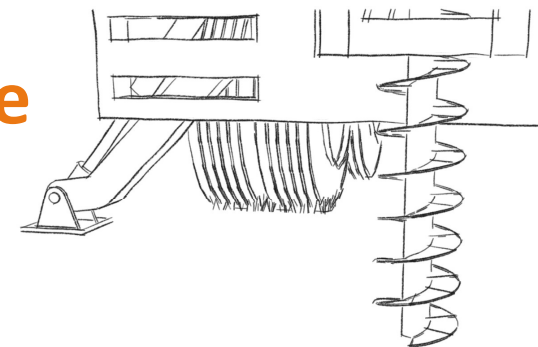


Our Profession



Geosetta

Beyond Data Collection: Extracting Intelligence



From Raw Data to Actionable Insights

- Transforming fragmented historical records into predictive tools
- Leveraging patterns across hundreds of thousands of data points
- Developing machine learning models that "learn" from previous explorations

What Our Algorithms Can Predict:

- Soil classification probability by depth and location
- Expected SPT blow count ranges
- Depth to bedrock estimations

The Goal: Provide engineers with preliminary subsurface expectations before the first drill bit touches soil - reducing uncertainty and optimizing field exploration programs.



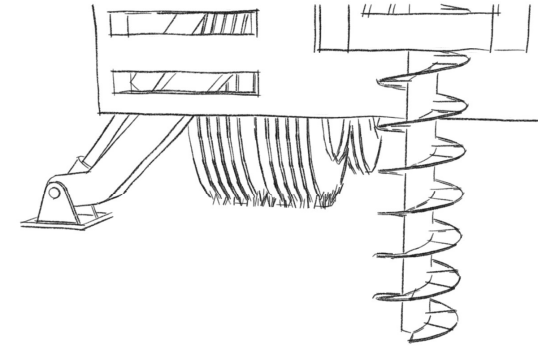
Geosetta

Geosetta's Features



IMPORTANT NOTE:

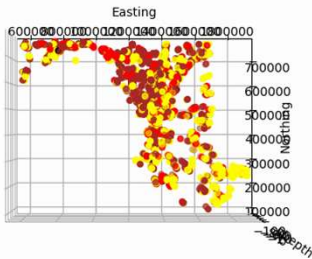
- We are not doing magic!
- Imagine you have a driller in your office that has been around from the beginning and was there for every boring.
- You ask this driller what they think you will encounter at a new project. What do you think they would say?
- Does this advice have value?
- Would you use their advice to replace field testing?



Geosetta

Geosetta's Machine Learning

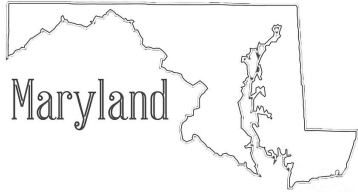
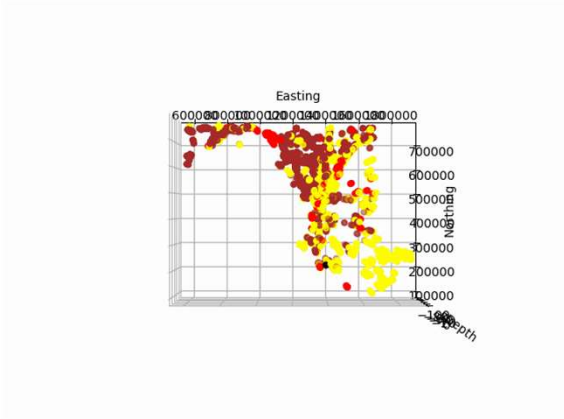
Actual data



Grainsize

- Clay
- SAND
- SILT
- GRAVEL

Model data



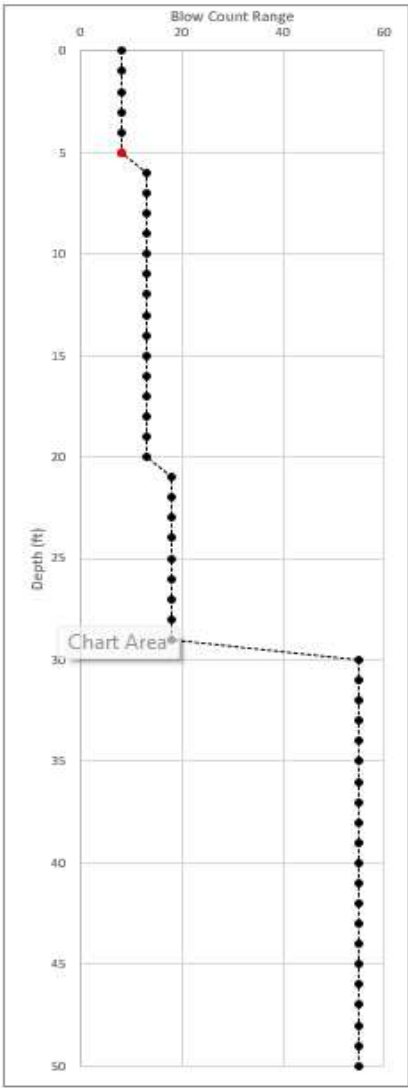
Change Depths Here

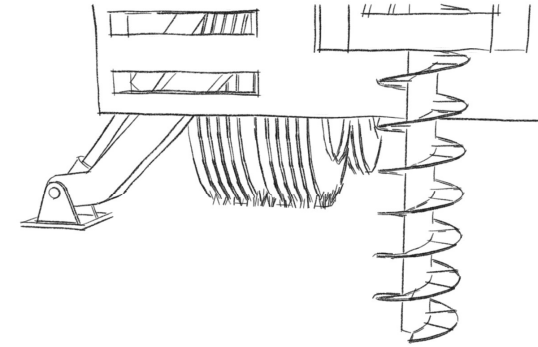
Select Depth

Rock Depth* Range (ft) 11-20

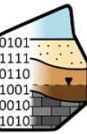
Depth (ft)	Elevation (ft)	Grainsize	N Range
0	1017	GRAVEL	6-10
1	1016	GRAVEL	6-10
2	1015	GRAVEL	6-10
3	1014	GRAVEL	6-10
4	1013	GRAVEL	6-10
5	1012	GRAVEL	6-10
6	1011	GRAVEL	11-15
7	1010	GRAVEL	11-15
8	1009	GRAVEL	11-15
9	1008	GRAVEL	11-15
10	1007	GRAVEL	11-15
11	1006	GRAVEL	11-15
12	1005	GRAVEL	11-15
13	1004	GRAVEL	11-15
14	1003	GRAVEL	11-15
15	1002	GRAVEL	11-15
16	1001	GRAVEL	11-15
17	1000	GRAVEL	11-15
18	999	GRAVEL	11-15
19	998	GRAVEL	11-15
20	997	GRAVEL	11-15
21	996	GRAVEL	16-20
22	995	GRAVEL	16-20
23	994	GRAVEL	16-20
24	993	GRAVEL	16-20
25	992	GRAVEL	16-20
26	991	GRAVEL	16-20
27	990	GRAVEL	16-20
28	989	GRAVEL	16-20
29	988	GRAVEL	16-20
30	987	Refusal	50+
31	986	Refusal	50+
32	985	Refusal	50+
33	984	Refusal	50+
34	983	Refusal	50+
35	982	Refusal	50+
36	981	Refusal	50+
37	980	Refusal	50+
38	979	Refusal	50+
39	978	Refusal	50+
40	977	Refusal	50+
41	976	Refusal	50+
42	975	Refusal	50+
43	974	Refusal	50+
44	973	Refusal	50+
45	972	Refusal	50+
46	971	Refusal	50+
47	970	Refusal	50+
48	969	Refusal	50+
49	968	Refusal	50+
50	967	Refusal	50+

Predicted SPT Blow Count Log



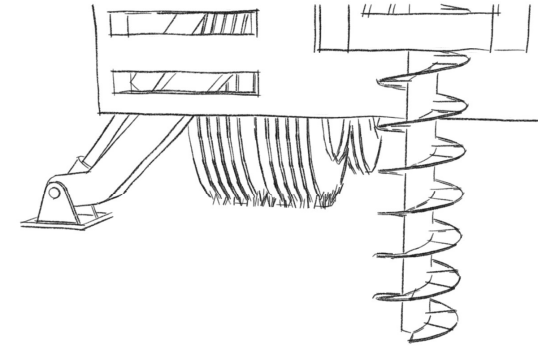


Application Demo



Geosetta

How do we extract these massive collections of data?



Two Complementary Approaches:

1. Structured Data Processing

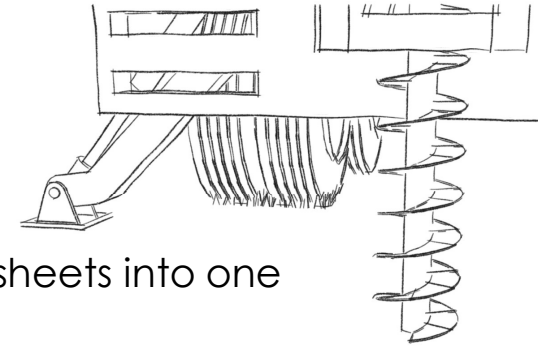
- Extraction from tabular formats (glNT files, Excel spreadsheets, databases)

2. Document Intelligence Processing

- Extraction from unstructured formats (PDFs, scanned reports, images)



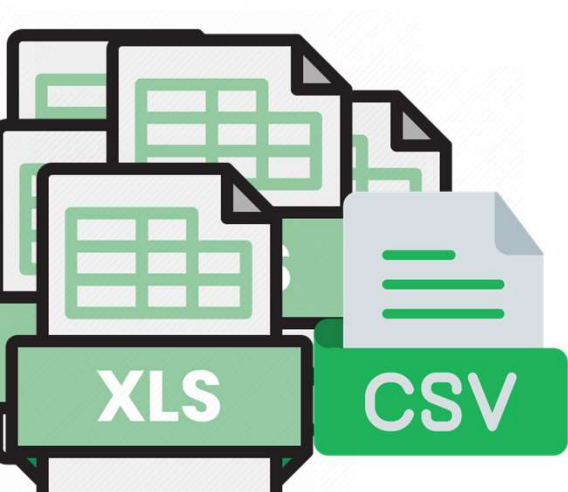
Structured Data Processing Challenge



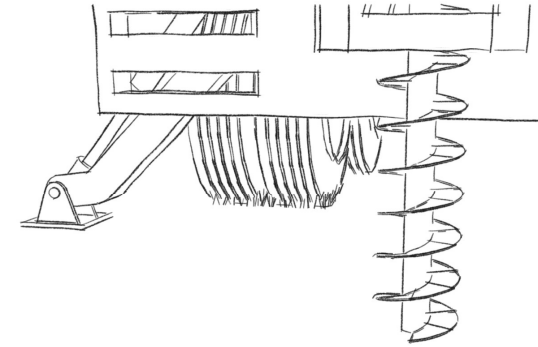
Imagine you are tasked with combining tens of thousands of individual Excel sheets into one Excel sheet with a consistent structure.

What are your challenges?

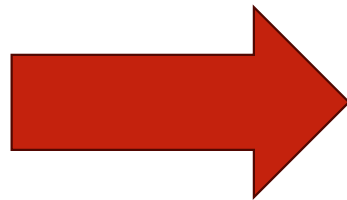
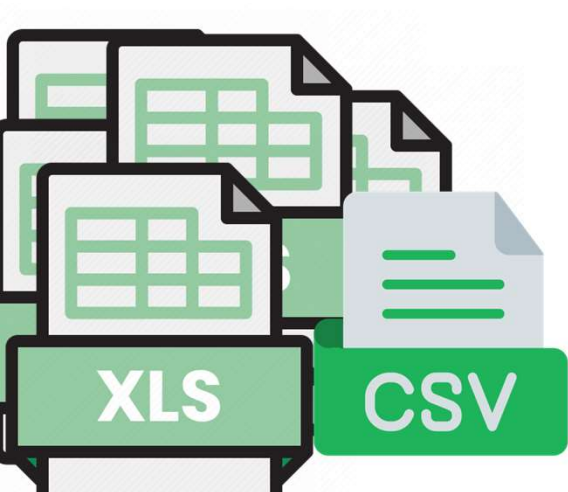
- Each file varies by region, office, and era ✓
- Harmonizing diverse geotechnical parameters (N_Value, BlowCount, N_Field etc) ✓
- Identifying and resolving duplicate exploration points across datasets ✓
- Creating comprehensive mapping algorithms to transform disparate data elements into a uniform DIGGS-compatible database. ✓



Structured Data Processing Challenge



This process is now solved and automated. We are processing massive DOT sized datasets into Geosetta.

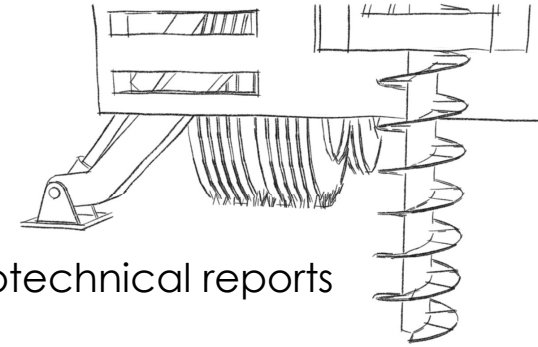


DiGGS
DATA INTERCHANGE FOR GEOTECHNICAL
AND GEOENVIRONMENTAL SPECIALISTS



Geosetta

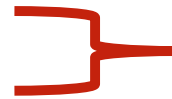
Document Intelligence Processing



Imagine you are tasked with extracting data from tens of thousands of geotechnical reports that been saved as PDF's.

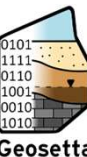
What are your challenges?

- Does the pdf have Geotechnical Data? ☒
- If so, find where the Testing took place? ☒



Powers our web crawler that automatically discovers and incorporates new public geotechnical reports.

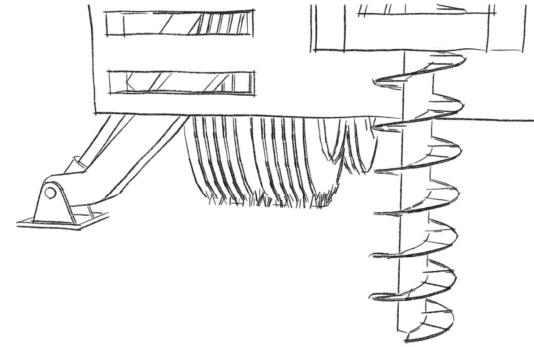
- Extract and Convert document based Geotechnical data into DIGGS
 - We're partnering with Virginia Tech to develop dual extraction methods that enable cross-validation.



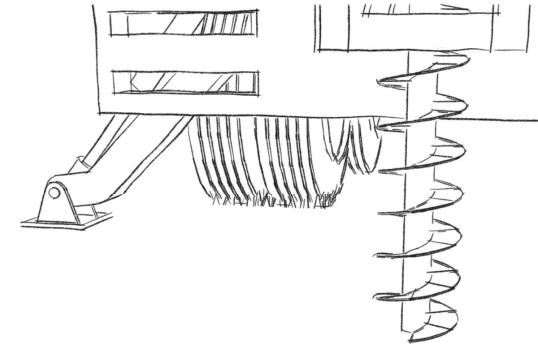
Geosetta

Future Plans

- Back populating laboratory data into our dataset.
- Integrating MWD, Insitu, Geophysics data.
- Expanding our reach beyond State DOTs.
- Providing a public tool to convert PDF Geotechnical reports to DIGGS files.



Call to Action

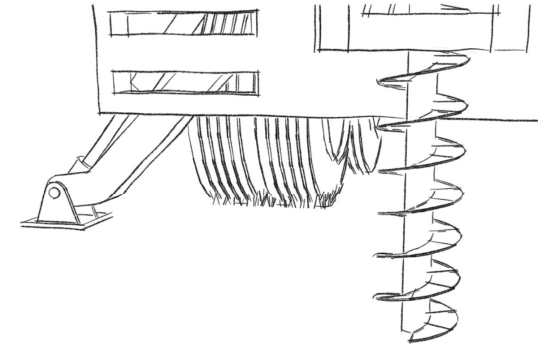


We need your support to keep this data open and growing.

- Help us get more data (Donating data, Connecting us with data owners)
- Be a Sponsor
- Help us spread the word
- Utilize our API for integrating historic data



Geosetta



Discussion



Ross.cutts@geosetta.org

