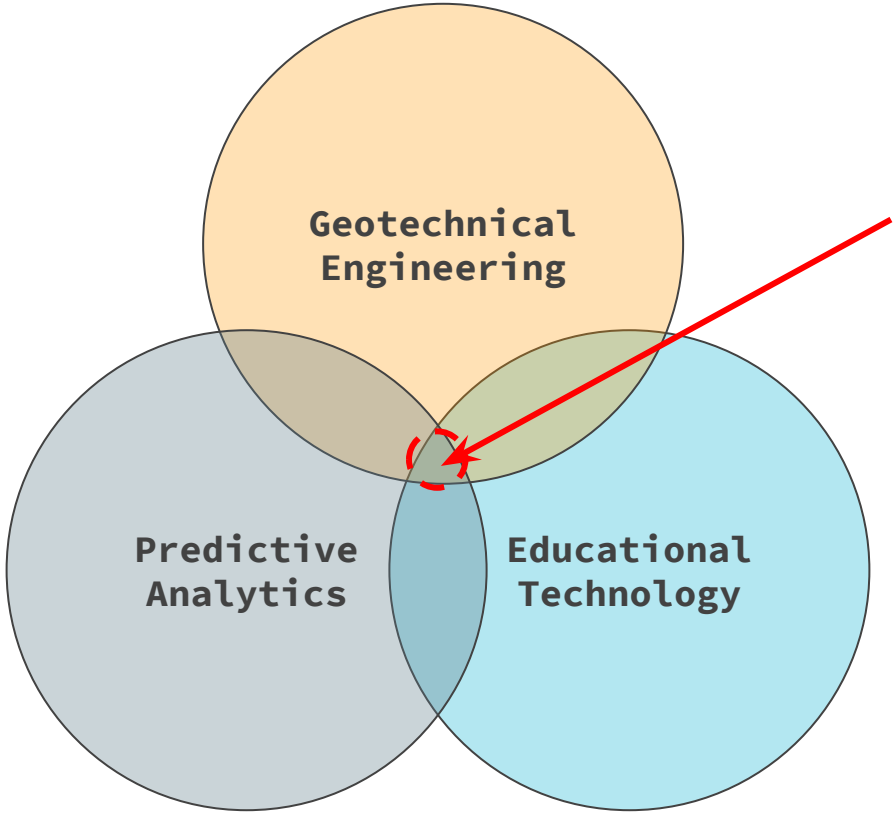
Advanced Data Analytics In Geotechnical Engineering

Nick Machairas

Topics

1. Advanced Data Analytics
2. Modern Data Management
3. Applications of Data Warehousing and Advanced Data Analytics in Geotechnical Engineering



Augment our engineering and decision making skills with powerful predictive techniques to deliver insights in forms that are comprehensible by all (preferably interactive).

What is “Advanced Data Analytics”?

Advanced Data Analytics

The field includes:

A. **Artificial Intelligence (AI) & Machine Learning (ML)**

B. **Efficient Automation**

Can have either ‘A’, ‘B’, or both

A.I.: The Fourth Industrial Revolution

"AI is the new electricity. Just as the Industrial Revolution freed up a lot of humanity from physical drudgery, I think AI has the potential to free up humanity from a lot of the mental drudgery"

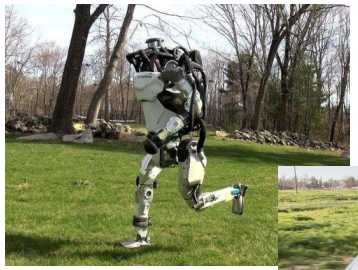
- Andrew Ng

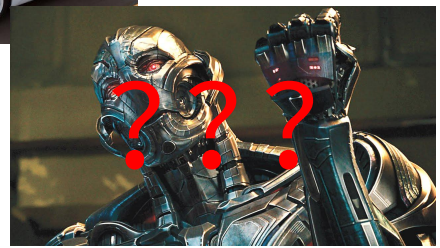
"The last 10 years have been about building a world that is mobile-first. In the next 10 years, we will shift to a world that is AI-first."

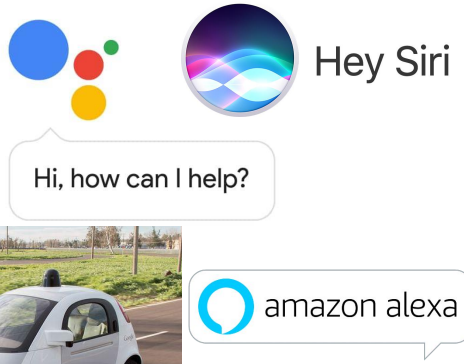
- Sundar Pichai

Applications of Artificial Intelligence

- Speech Recognition
- Handwriting Recognition
- Machine Translation
- Robotics
- Recommendation Systems
- Email Spam Detection and Sorting
- Face Detection
- Medical Applications (cancer detection)
- Adversarial Search
- Natural Language Processing and Information Extraction
- Autonomous Driving







From Artificial Intelligence to Machine Learning

How are Artificial Intelligence and Machine Learning related?

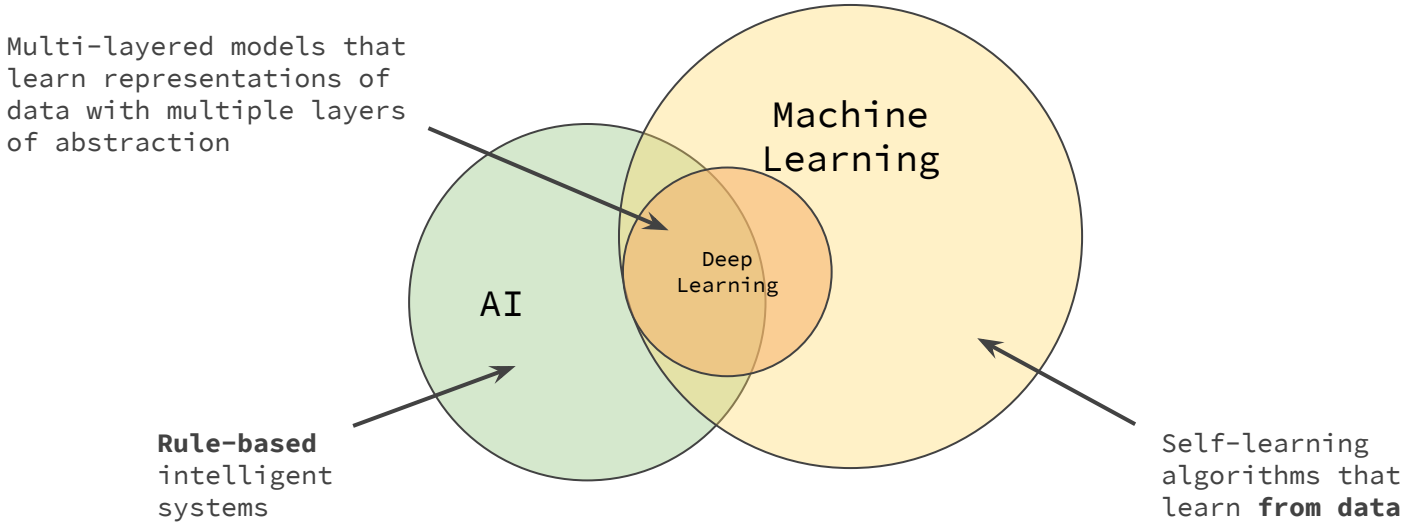
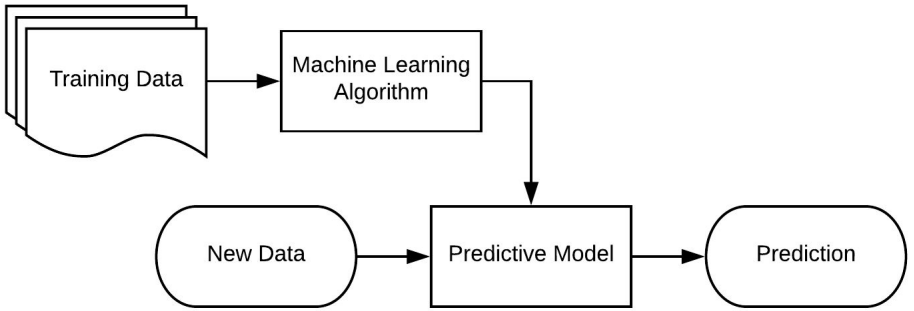


Figure adopted from Sebastian Raschka.

What is Machine Learning

Building intelligent machines to transform data into knowledge



The Essence of Machine Learning:

- A pattern exists
- We cannot pin it down mathematically
- We have data on it

- Yaser Abu-Mostafa, Learning from Data, 2012

Types of Machine Learning

Types of Machine Learning:

- Supervised learning
- Unsupervised learning
- Reinforcement learning

	Continuous	Categorical
Supervised	Regression	Classification
Unsupervised	Dimension Reduction	Clustering

Efficient Automation

Advanced Data Analytics includes:

A. Artificial Intelligence (AI) & Machine Learning (ML)

B. Efficient Automation

a. Using custom tools for large tasks

b. Implementing checks, warnings and triggers

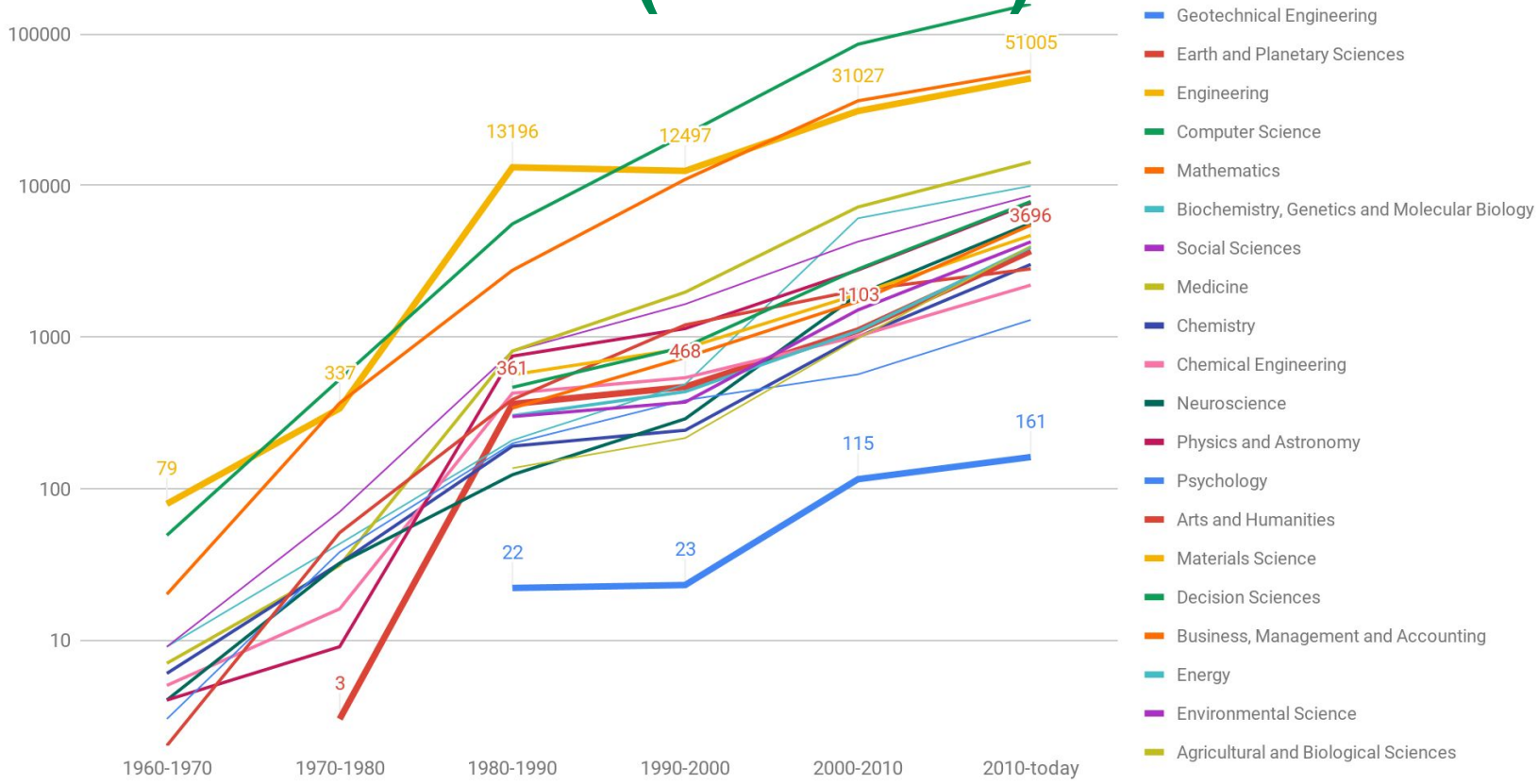
c. Bring in and process IoT data

d. Advanced Numerical Modelling (beyond building the model)

e. Don't be afraid to break away from the spreadsheet, embrace basic coding/scripting

What is the status of AI/ML in Geotechnical Engineering?

Published Work (AI or ML)



Modern Data Management

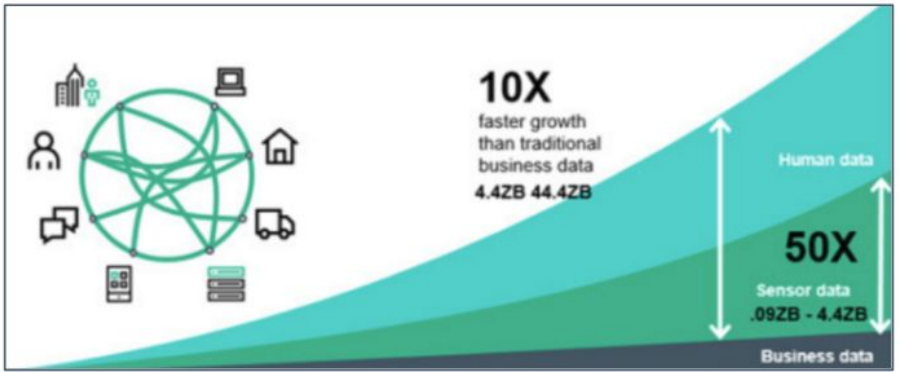
A Familiar System… Spreadsheets

- Intuitive table-type structure with columns and rows
- Manageable for few data but slow and cumbersome for a lot of data
- Difficult for another user to “decipher” complicated analyses
- Good to quickly produce charts
- Limited protection and data integrity constraints
- Limited options to “plug-in” to other applications

	A	B	C
1		Data Set 1	Data Set 2
2	Experimental K	0.341	0.246506896
3			
4		Data Set 1	Data Set 2
5	Wavelength (nm)	Eave(λ) [W/m2]	Eave(λ) [W/m2]
6	290	0.009990362	0.006295514
7	291	0.012906104	0.008377919
8	292	0.016328124	0.010896005
9	293	0.018552996	0.012691824
10	294	0.020625292	0.014456731
11	295	0.023820656	0.017050798
12	296	0.027690493	0.02020932
13	297	0.03299268	0.024495629
14	298	0.038026407	0.02870029
15	299	0.042884443	0.032837552
16	300	0.048955388	0.037998854
17	301	0.056658503	0.044572408
18	302	0.064565623	0.051364415
19	303	0.073376379	0.059025038
20	304	0.084335325	0.068463725
21	305	0.095932937	0.078532196
22	306	0.101638012	0.083810062

Data is everywhere and exponentially growing

- 90% of the data in the world today has been created in the last two years alone.
- 50X growth from 2010 to 2020
- By 2020:
  - 4.4ZB of Business Data
  - 44.4ZB of Human & Machine Data



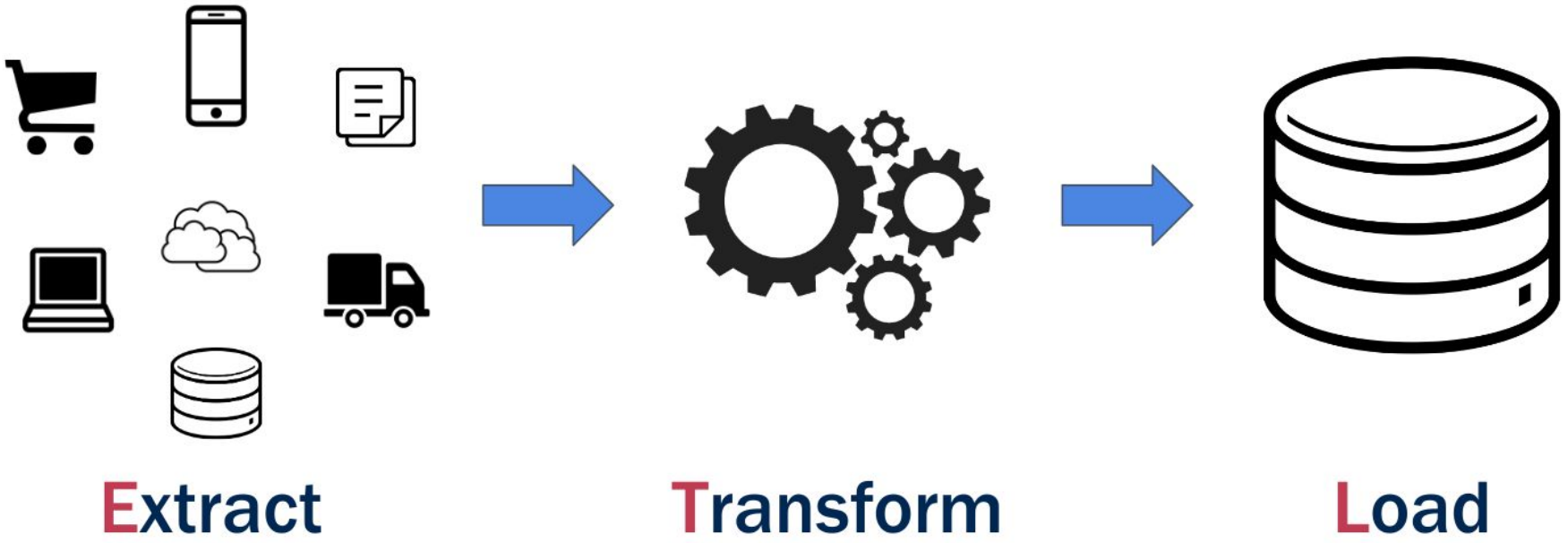
Source: InsideBIGDATA Guide to The Intelligent Use of Big Data on an Industrial Scale

Great benefits for those who manage to efficiently store, retrieve and analyze data.

Data Engineering

- Relational Database Management Systems (RDBMS)
- Data Warehouses
- NoSQL for Big Data Applications
- Combining multiple data sources
- Data Wrangling
- Extract, Transform & Load (ETL)

Extract, Transform and Load

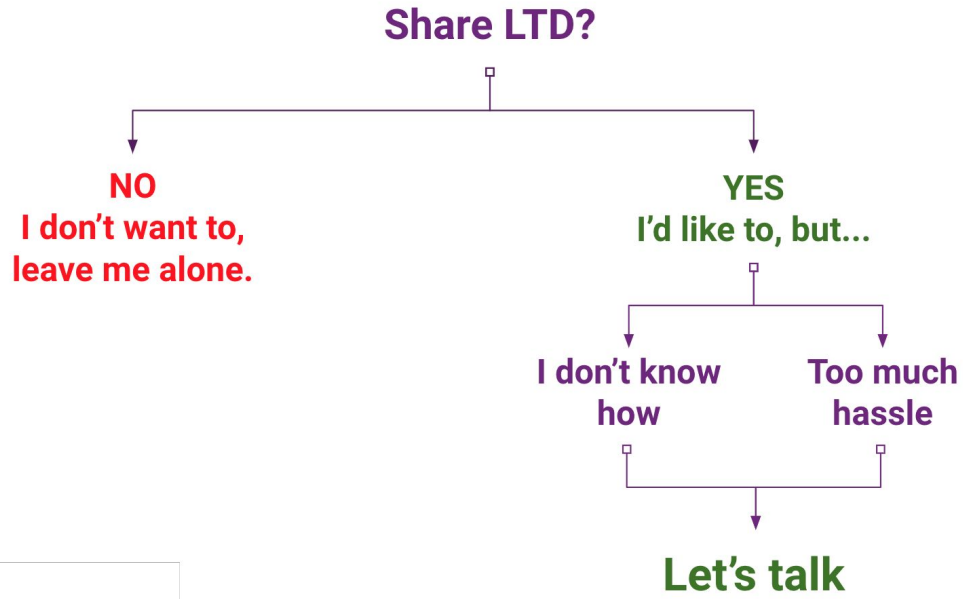


Extract, Transform and Load (cont.)

- Common challenges:
  - Retrieving and integrating data from multiple sources
  - Cleansing and transforming the data
  - Loading the data into appropriate data stores for analysis and reporting
- Enterprises spend 60%–80% of their resources developing, testing and maintaining their ETL processes
- ETL processes usually require dedicated monitoring and maintenance

How is the Geo Community Handling Data?

- Not good... but this is starting to change
- From experience on Pile Load Test Data:



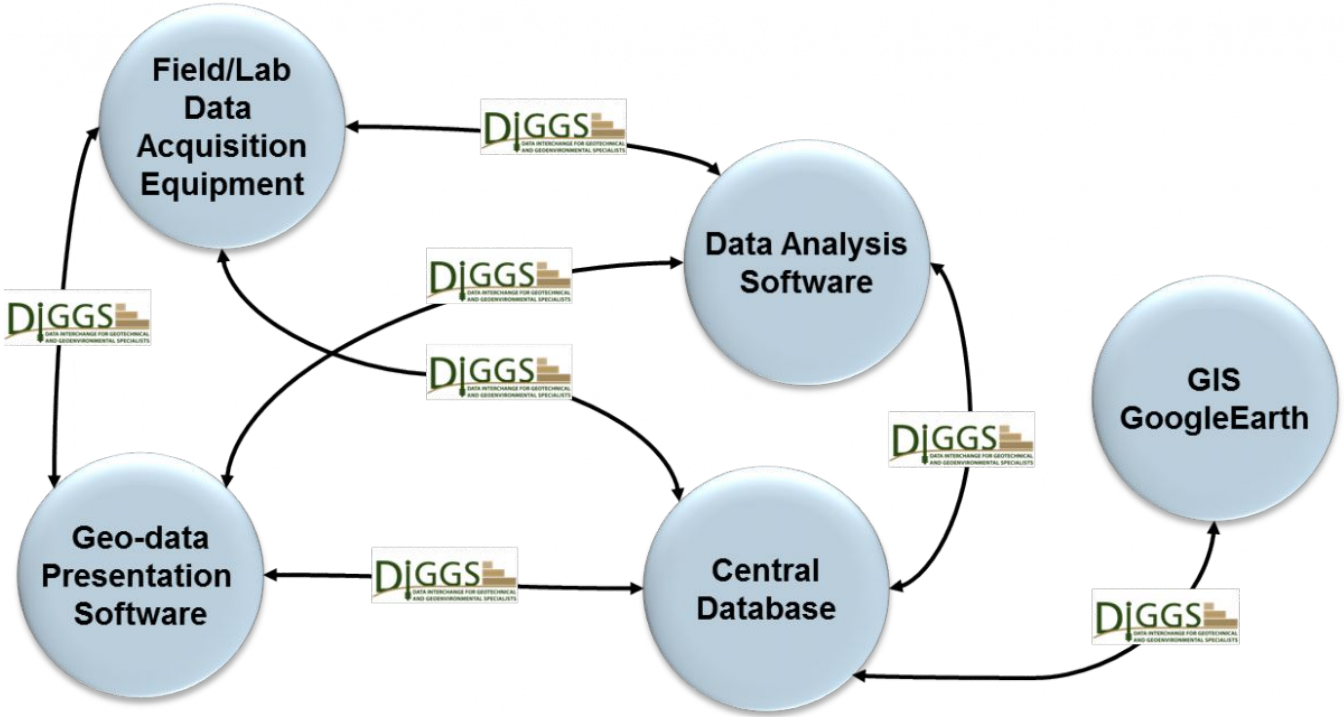
Handling Engineering Data is Not Easy...

... but that should never be an excuse!

- Isolated environment, no common consensus, everybody is basically speaking their own language.
- A standard for geotechnical data transfer exists but has not been adopted (DIGGS).
- Properly managing Geotechnical Data can get very complicated, especially when combined information from other elements and processes.

DIGGS (diggsml.org)

Data Interchange for Geotechnical and Geoenvironmental Specialists (DIGGS)



Data Warehousing and Business Intelligence

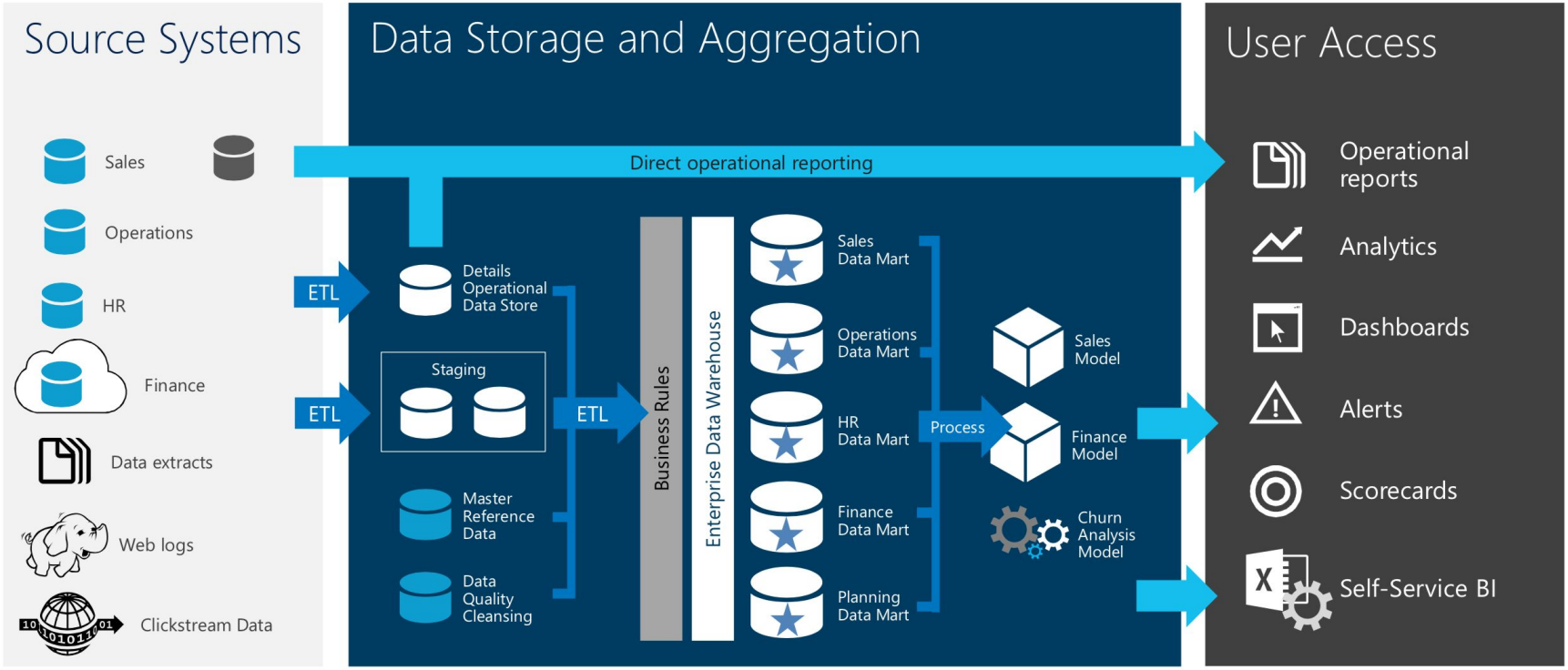


Figure adopted from Microsoft and Peter Meyers: Delivering a Relational Data Warehouse

Data Warehousing and ~~Business~~ Geotechnical Intelligence

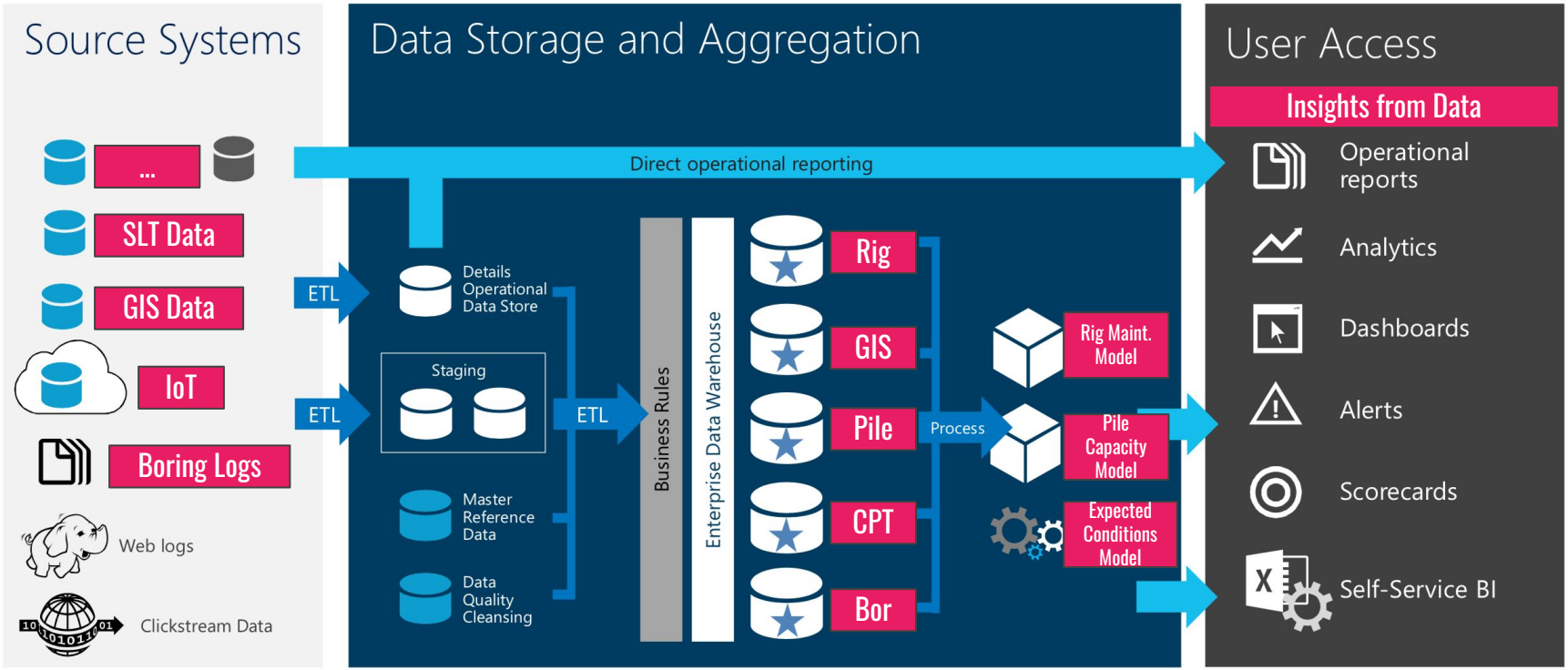


Figure adopted from Microsoft and Peter Meyers: Delivering a Relational Data Warehouse

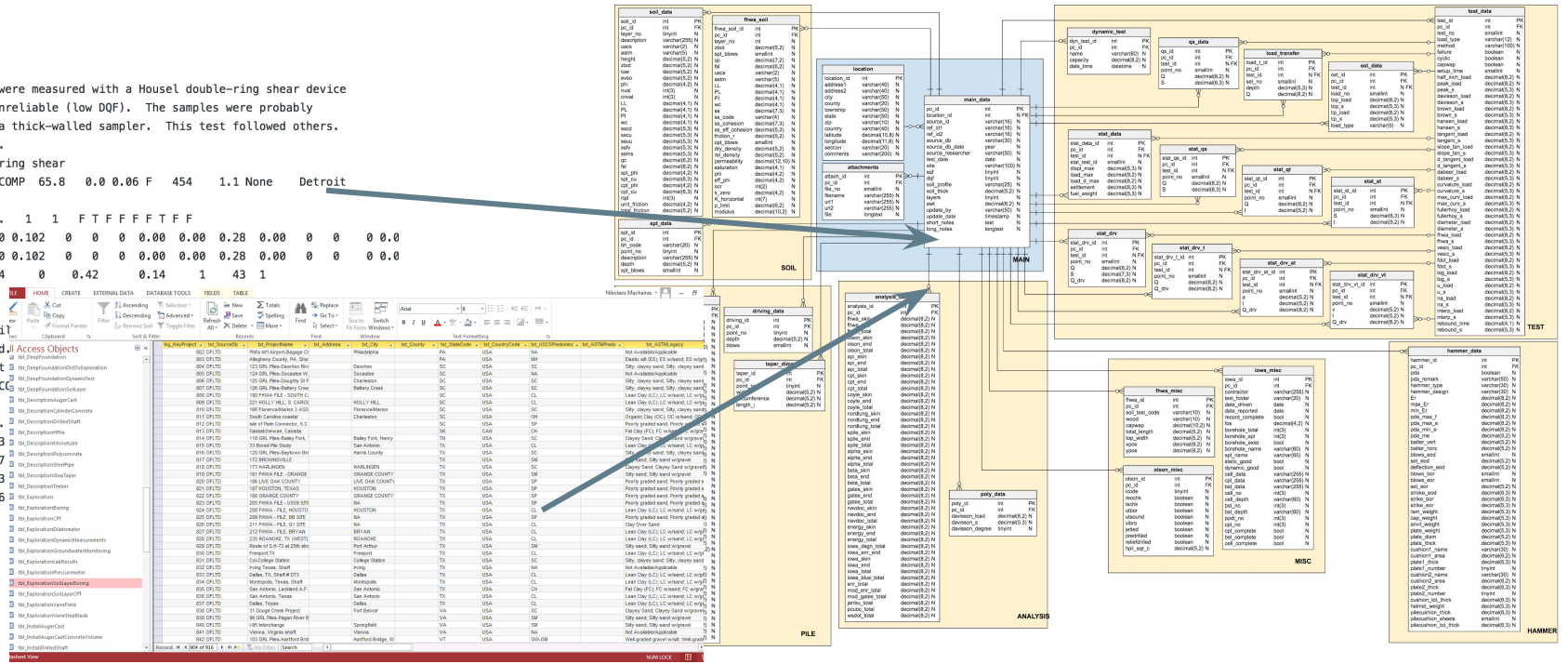
Benefits

- Combine all of your data sources
- Enter data once, use multiple times
- Access to information is controlled
- Immediate retrieval of archived data
- Enhance collaboration across the board
- Leverage large amounts of data to build highly accurate predictive models
- We must modernize our field and improve even fundamental design methodologies

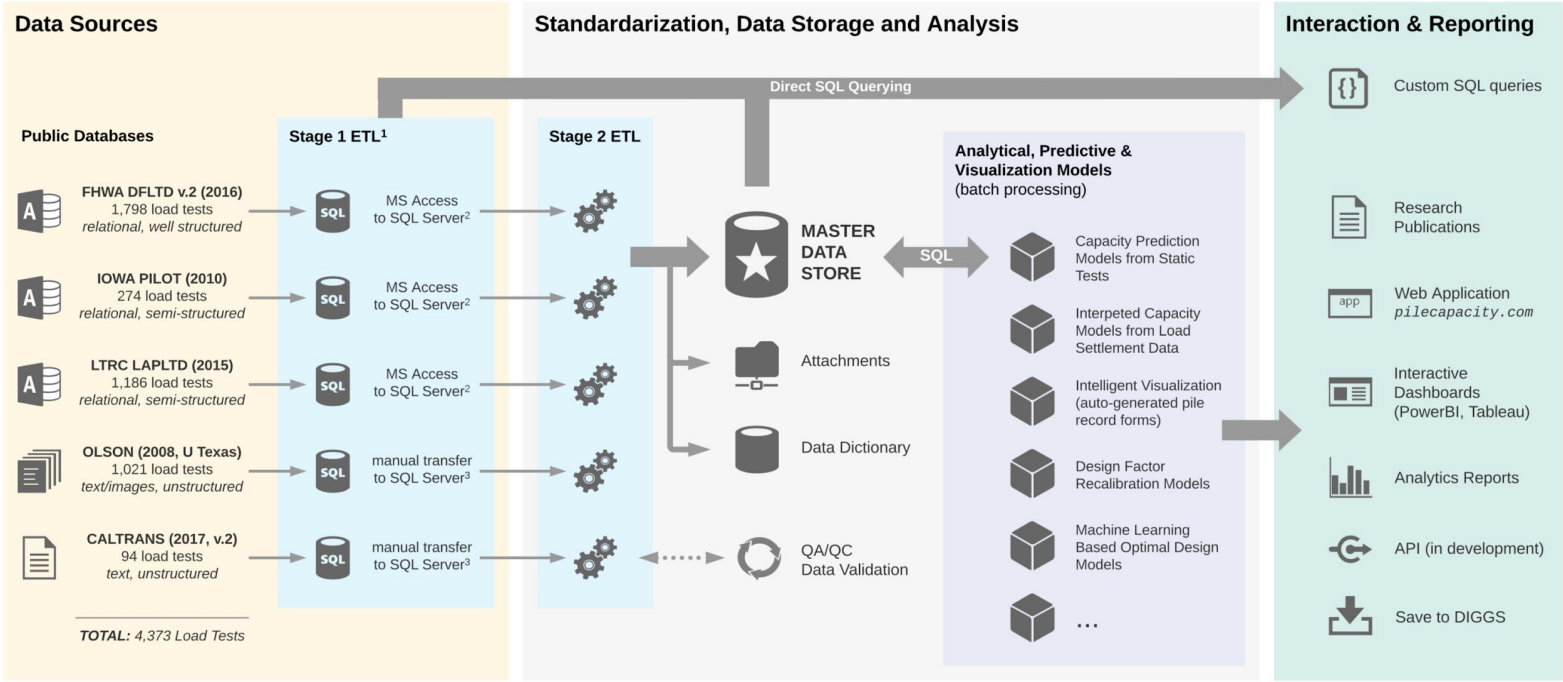
Examples of Modern Data Management and Predictive Analytics

NYU Pile Load Test Data Warehouse

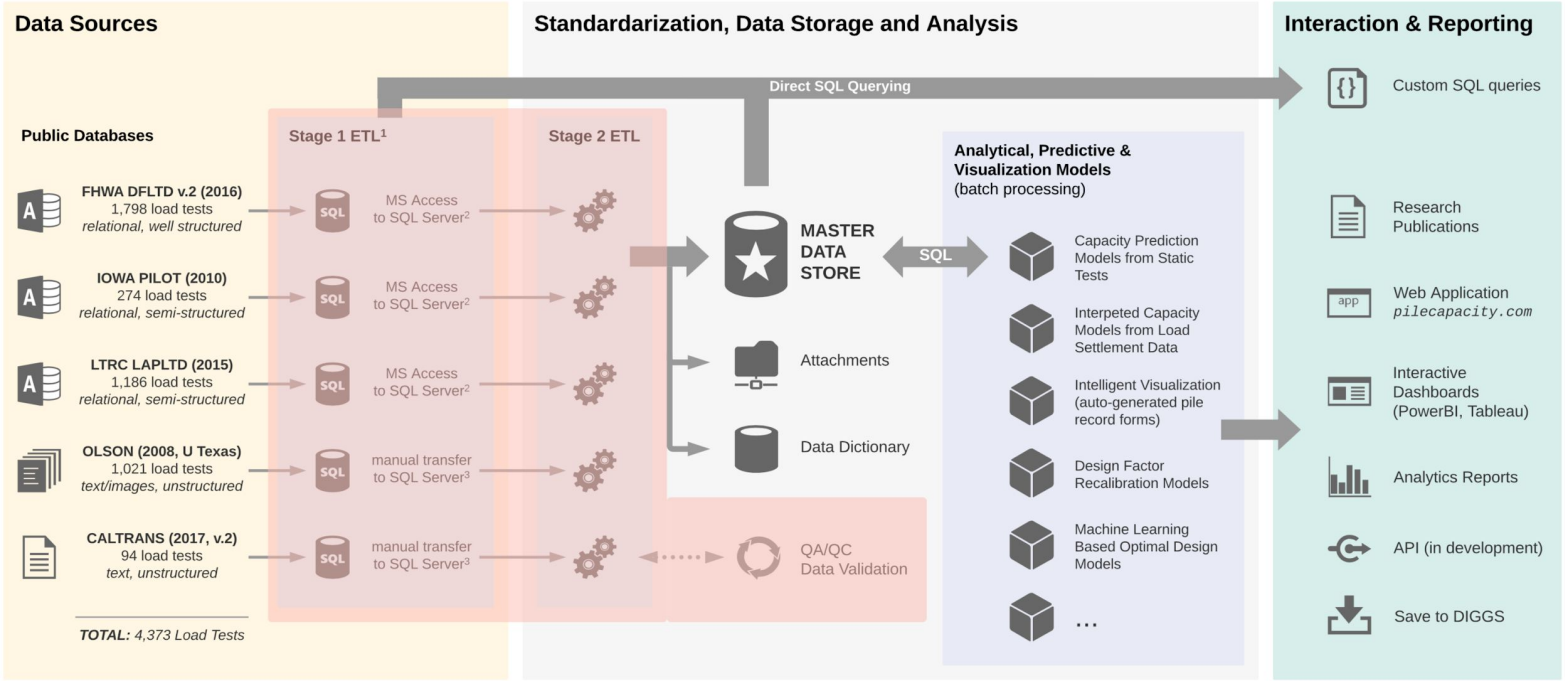
```
N Blank*
The strengths were measured with a Housel double-ring shear device
and are thus unreliable (low DQF).  The samples were probably
obtained with a thick-walled sampler.  This test followed others.
EWT is assumed.
Housel double ring shear
  1 STLP CIRC COMP  65.8    0.0 0.06 F   454     1.1 None     Detroit
 16.00   4.19
LAY     2   -5.   1   1   F T F F F F T F F
LAY  65.8  0.00 0.102   0   0   0 0.00  0.00  0.28  0.00   0   0    0 0.0
LAY  10.0  0.00 0.102   0   0   0 0.00  0.00  0.28  0.00   0   0    0 0.0
  0   84    84     0   0.42       0.14     1    43  1

N Blank*
This was the pil
Top layer cased
multiple retest
  2 STLP CIRC CO
  6.00   1.57
LAY     3    5.
LAY   9.0  0.23
LAY  14.5  1.07
LAY  10.0  1.43
  0    6    6
```



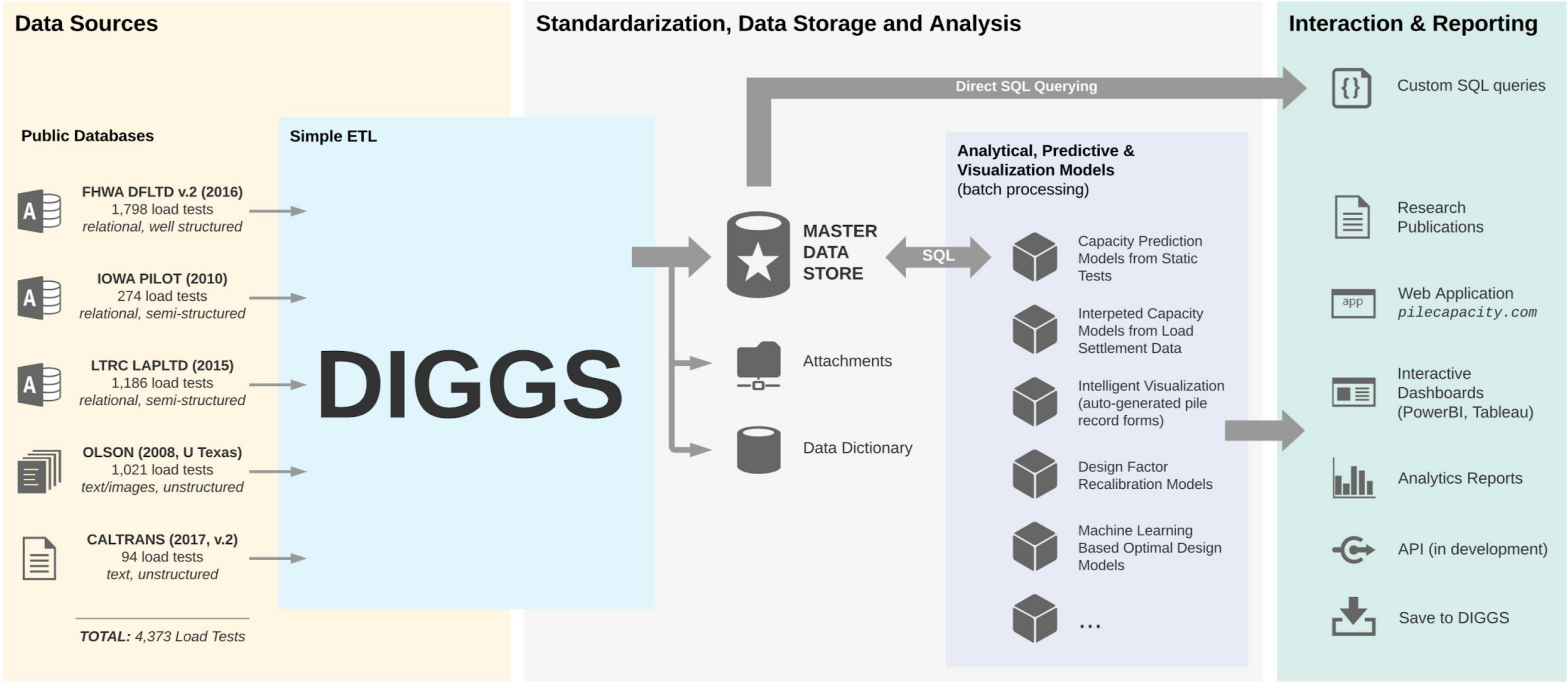
NYU Pile Load Test Data Warehouse



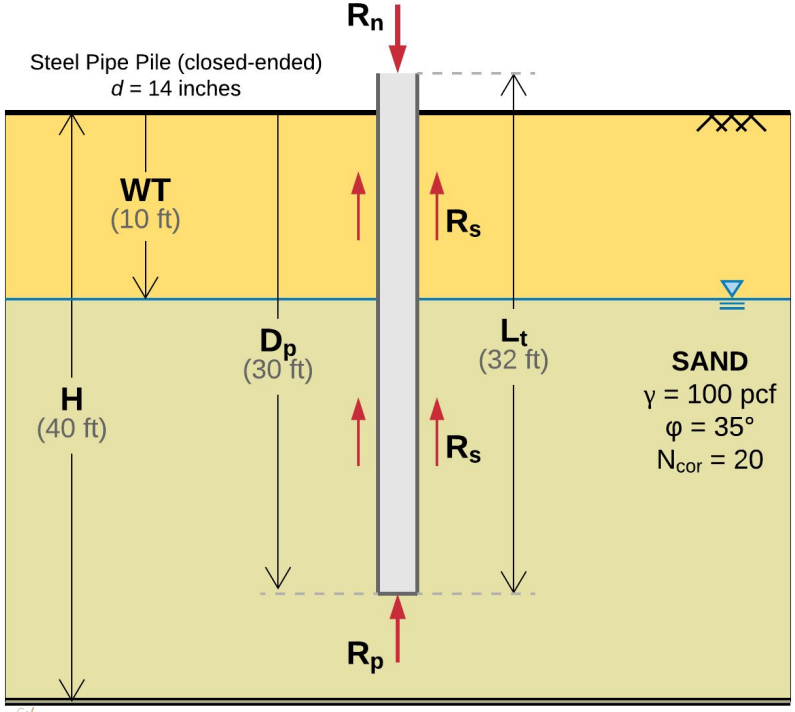
NYU Pile Load Test Data Warehouse



NYU Pile Load Test Data Warehouse

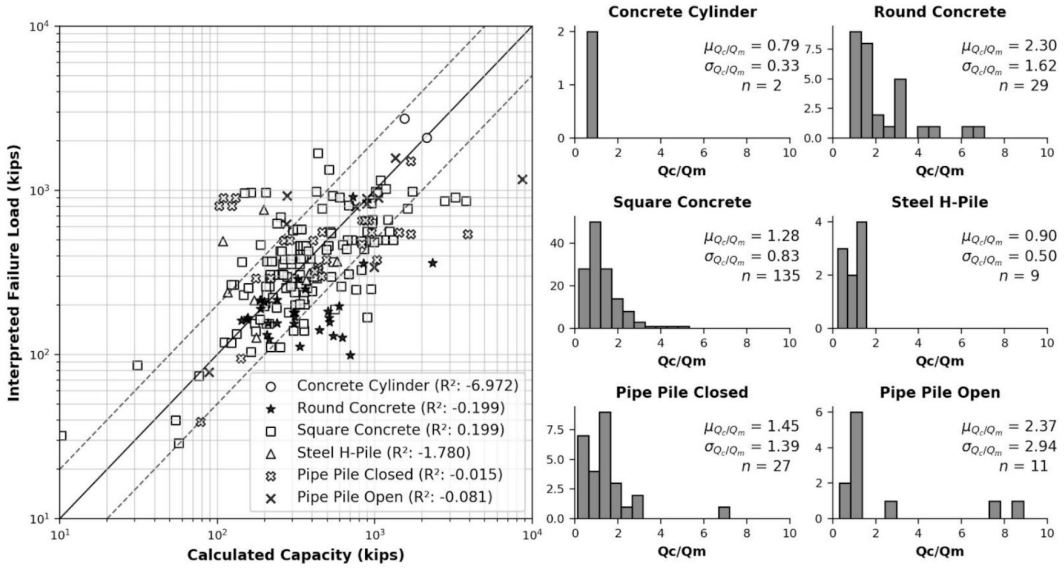


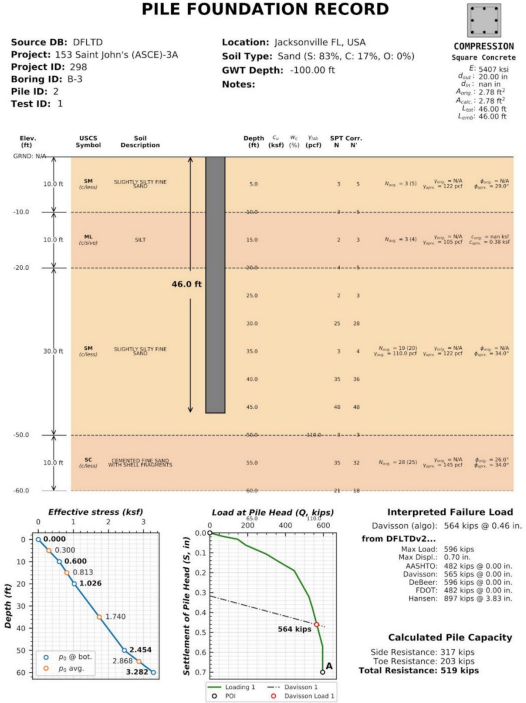
Algorithmic Implementations in Python



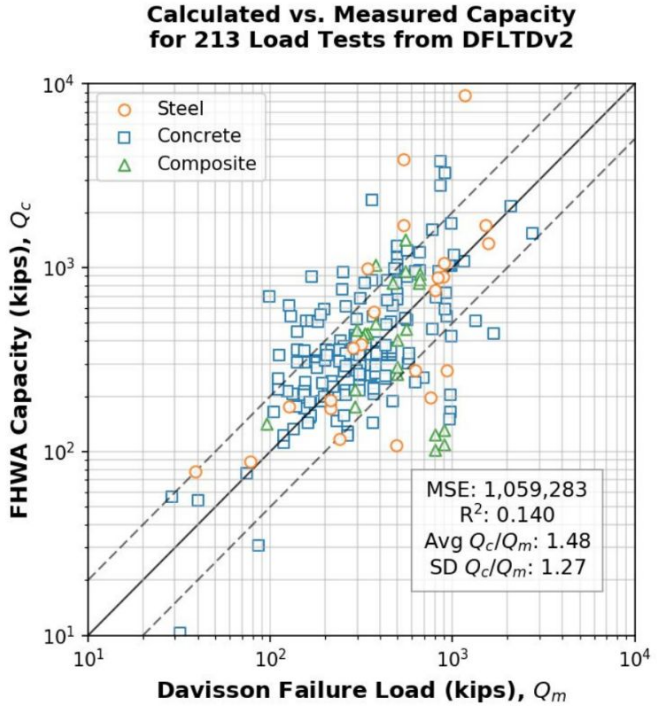

```
# Import the Project, SoilProfile and Pile classes
In [1]: from edafos.project import Project

In [2]: from edafos.soil import SoilProfile

In [3]: from edafos.deepfoundations import Pile

# Create the project object
In [4]: project = Project(unit_system='English', project_name='Example 1')

# Create a SoilProfile object with initial parameters
In [5]: profile = SoilProfile(unit_system='English', water_table=10)

# Add layer properties
In [6]: profile.add_layer(soil_type='cohesionless',
   ...:                   height=40,
   ...:                   tuw=100,
   ...:                   field_phi=35,
   ...:                   corr_n=20)
   ...:
Out[6]: <edafos.soil.profile.SoilProfile at 0x10d04d390>

# Attach the soil profile to the project
In [7]: project.attach_sp(profile)
Out[7]: <edafos.project.Project at 0x10d04d780>

# Create a pile
In [8]: pile = Pile(unit_system='English',
   ...:             pile_type='pipe-closed',
   ...:             length=32,
   ...:             pen_depth=30,
   ...:             diameter=14,
   ...:             thickness=0.75)
   ...:

# Attach the pile to the project
In [9]: project.attach_pile(pile)
Out[9]: <edafos.project.Project at 0x10d04d780>
```

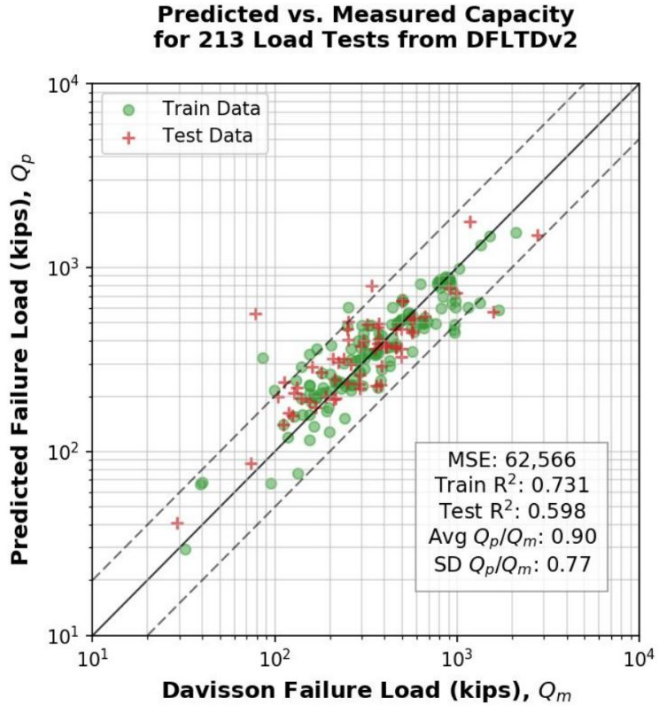
Evaluation of Driven Pile Capacity





Prediction of Driven Pile Capacity using Machine Learning





Prediction of Driven Pile Capacity (cont.)

PILE CAPACITY PREDICTOR (BETA)

This online tool features a *Support Vector Regressor* to predict the axial load capacity of pile foundations given soil type, average SPT-N values, pile type and open/closed end condition, pile cross sectional area, circumference and length. The process is outlined in:

> Machairas, N. P., and Iskander, M. G. (2018). "An Investigation of Pile Design Utilizing Advanced Data Analytics." *Proceedings of the International Foundations Congress and Equipment Expo 2018*, ADSC-The International Association of Foundation Drilling, DFI (Deep Foundations Institute), G-I (Geo-Institute of American Society of Civil Engineers), and PDCA (Pile Driving Contractors Association), March 5-10, 2018, Orlando, Florida.

DISCLAIMER

This tool is offered without any warranties about the accuracy of the predicted capacity. The predicted capacity is a result of approximation by scientific methodologies. The authors' sole intent is to further advance the field of Geotechnical Engineering and are not offering this online tool as a design aid. **Use to learn and experiment, do not design piles based on the numbers you get below.**

SOIL PROPERTIES

Select Predominant Soil Type: Sand

Select Average SPT-N count: **50**

PILE PROPERTIES

Select Pile Type: Steel

Open Ended? No

Select cross sectional area (in2): **16**

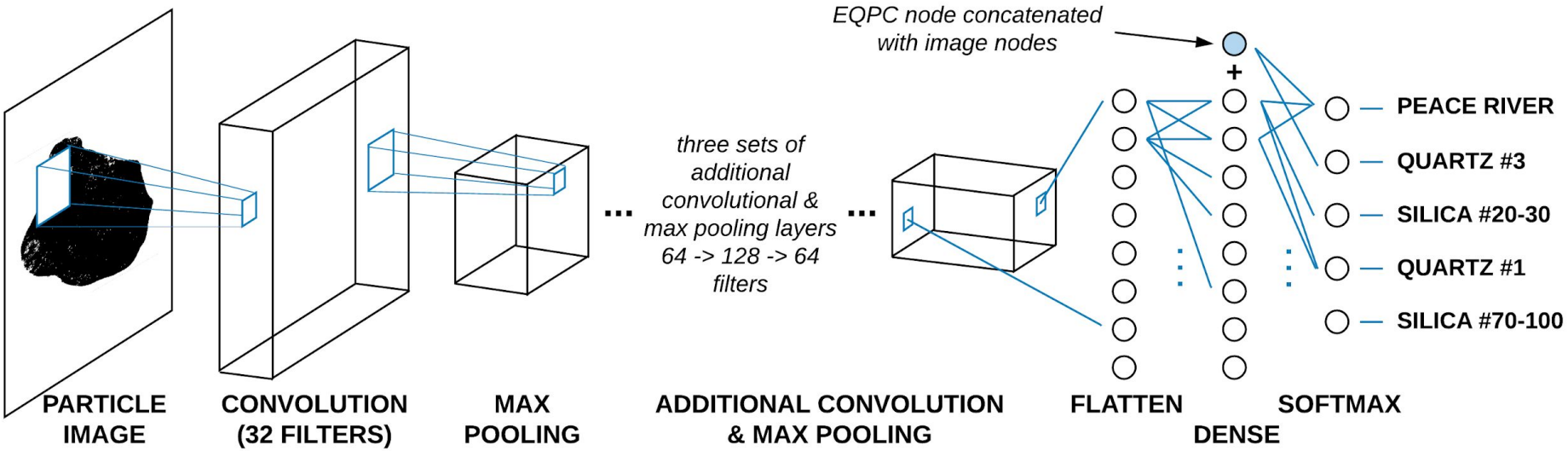
Select circumference (in): **44**

Select length (ft): **22.50**

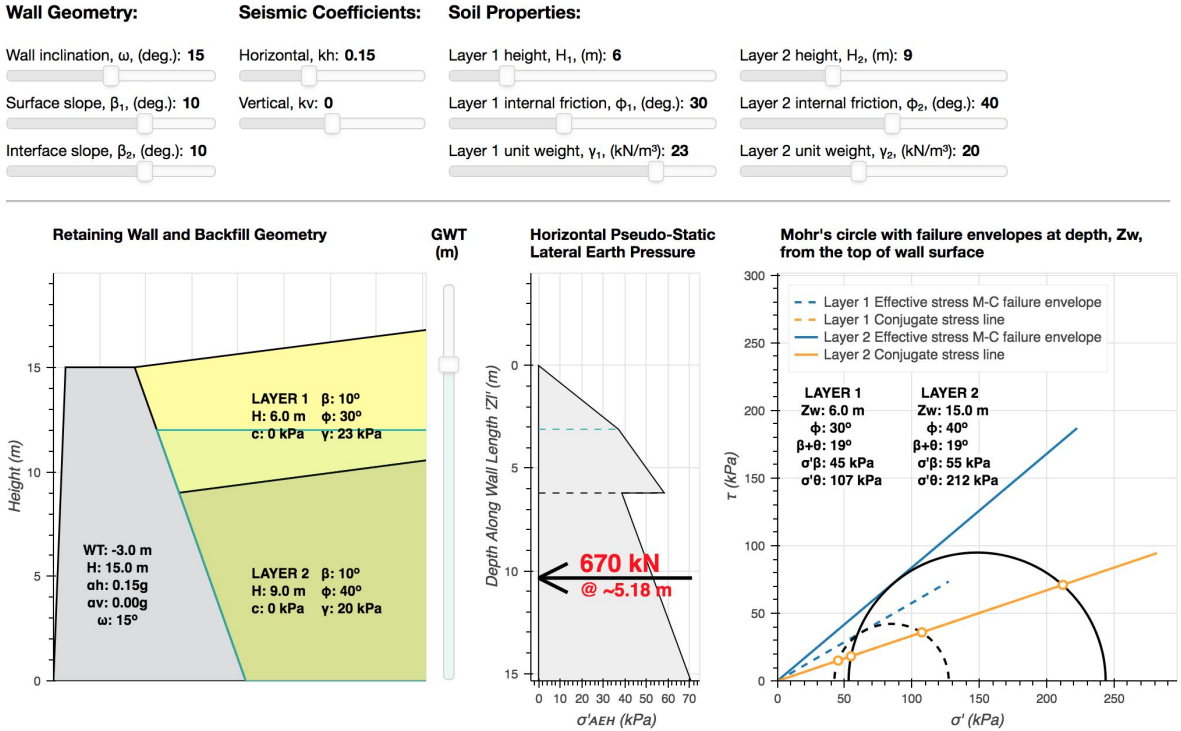
RESULT

457.75 kips

Deep Learning: Soil Particle Classification

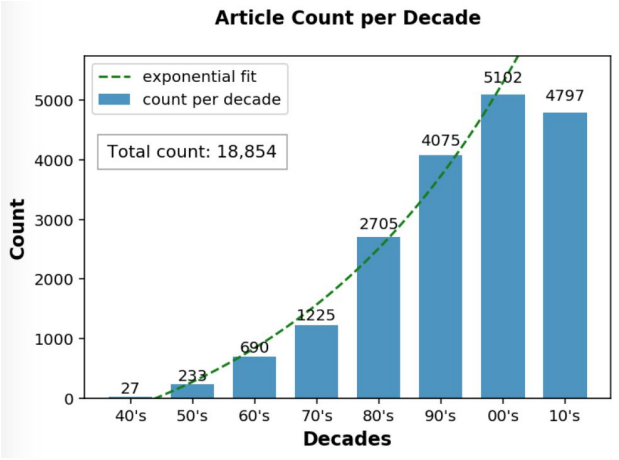


Seismic Earth Pressure Calculator (wp.nyu.edu/sep)



Text Analytics

Probabilistic Topic Modeling on a Large Corpus of Geotechnical Engineering Journal Articles



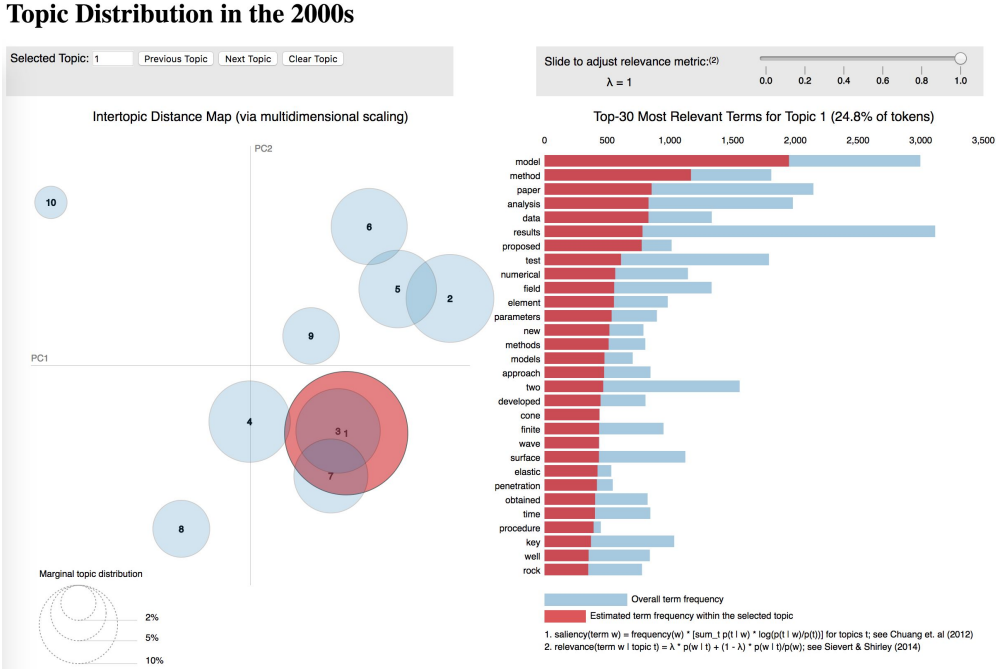
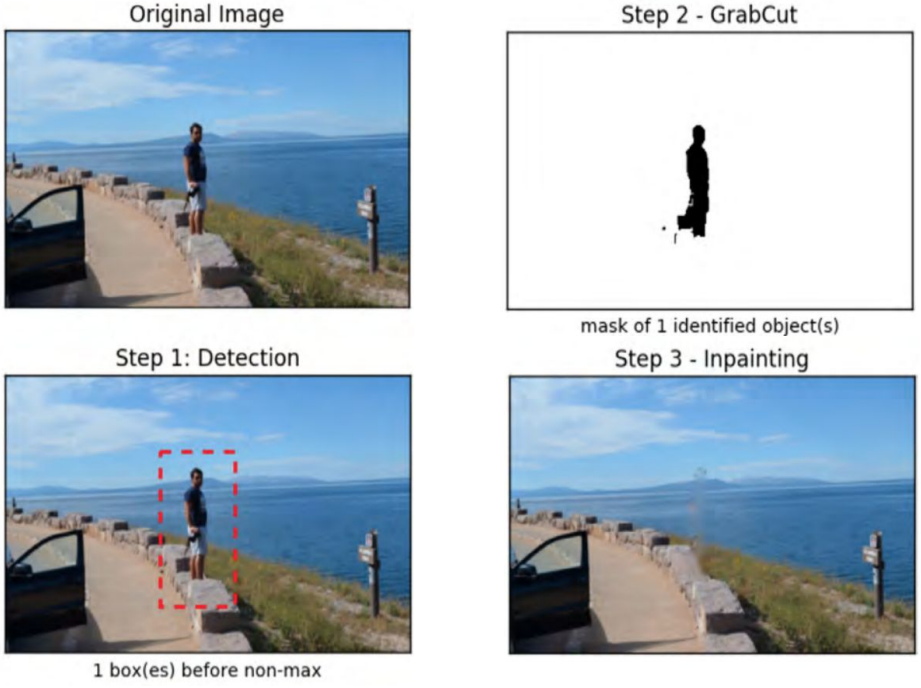


Image Analysis

Removing Human Artifacts from Images Without User Intervention



Predict Expected Geotechnical Conditions

- Build a Data Warehouse combining all of the available data sources
  - GIS data
  - Boring Logs
  - Satellite Data
  - Survey data
  - Geologic Maps
- Reinforcement Learning:
  - Develop predictive models from satellite, GIS, survey and geology
  - Validate against boring logs and improve the models
  - Predictive models will learn on their own and get better as more data is added to the Warehouse

Closing Remarks

Closing Remarks

- Many design approaches are deterministic. However, Data Management and Machine Learning can lead to probabilistic methods of analysis:
  - To improve design methods
  - To optimize design workflows and reduce costs
- Pairing traditional methods with predictive models will transform how we investigate, design and build
- Top-down implementation: executives are responsible for pushing adoption and implementation
- Basic coding skills are a must, provide incentives to your employes to get started

Thank you.

References

- Russell, S., and Norvig, P. (2010). Artificial Intelligence: A Modern Approach, 3e. Pearson Education.
- Raschka, S. (2015). Python Machine Learning. Packt Publishing.
- Abu-Mostafa, Y. S., Magdon-Ismail, M., and Lin, H.-T. (2012). Learning from Data: A Short Course. AMLBook.com.
- Machairas, N. P., Highley, G. A., and Iskander, M. G. (2018). “Evaluation of FHWA Pile Design Method Against the FHWA Deep Foundation Load Test Database Version 2.0.” Transportation Research Record: Journal of the Transportation Research Board.
- Machairas, N. P., and Iskander, M. G. (2018). “An Investigation of Pile Design Utilizing Advanced Data Analytics.” IFCEE 2018