

# Open Source Web-based Planning Analytics Data Warehouse (PADWare)

A Data Science Approach to Transportation Data Challenges

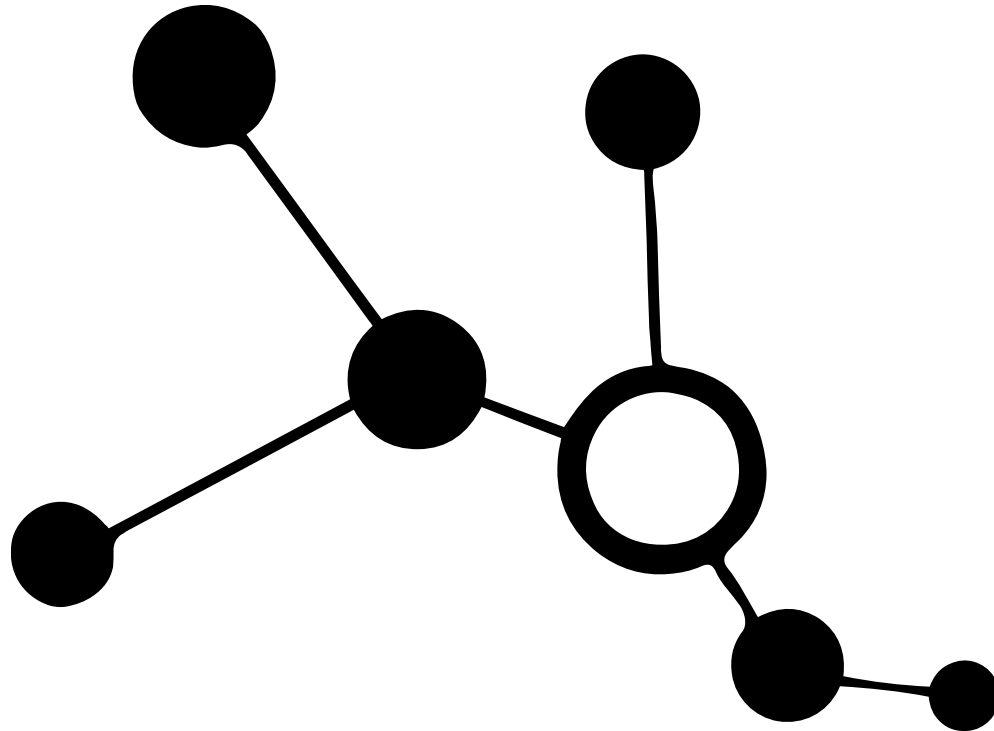
*Dr. Catherine T. Lawson  
Albany Visualization And Informatics Lab (AVAIL)  
Lewis Mumford Center  
University at Albany, SUNY*



We have many challenges ahead with performance measures and monitoring.

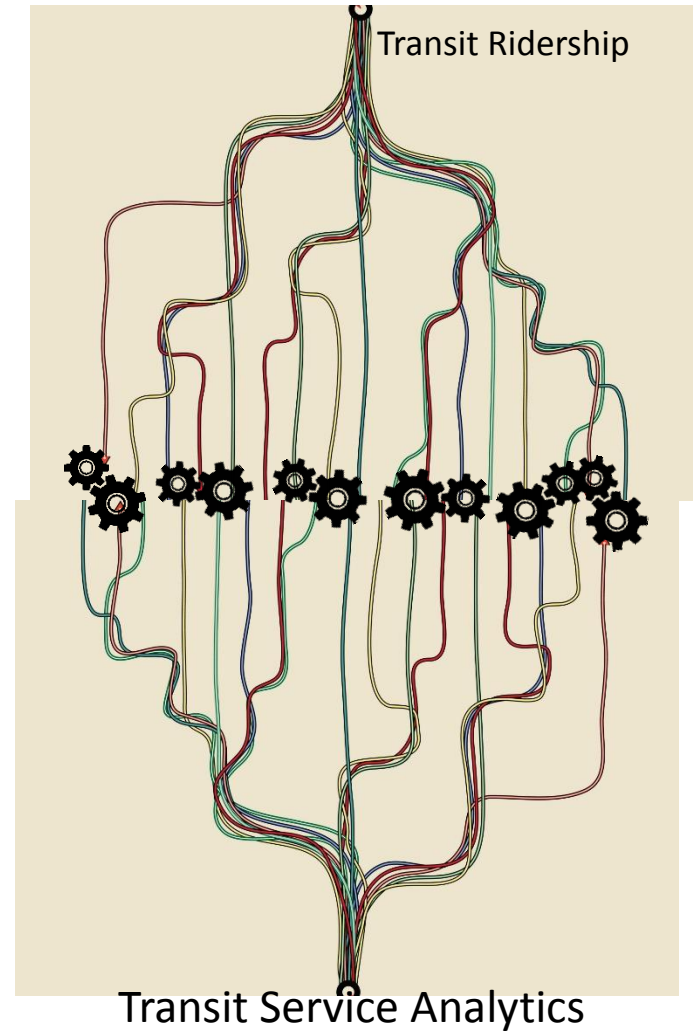
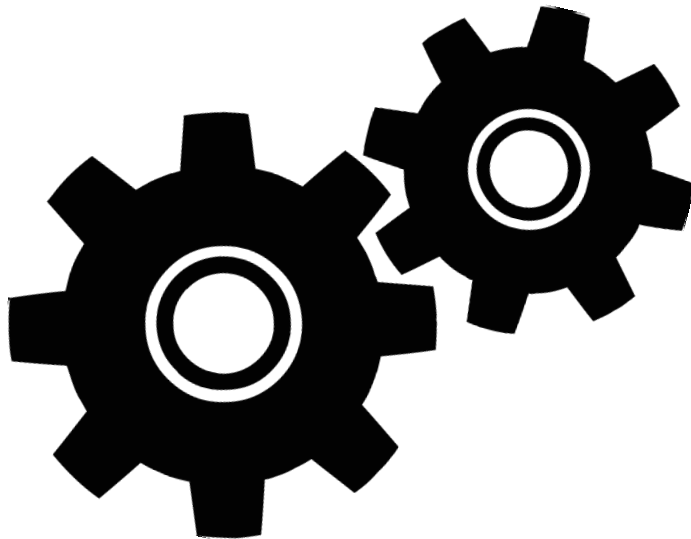
Could we utilize data science?

Some of us already do --- What if all of us did?



What would it look like?

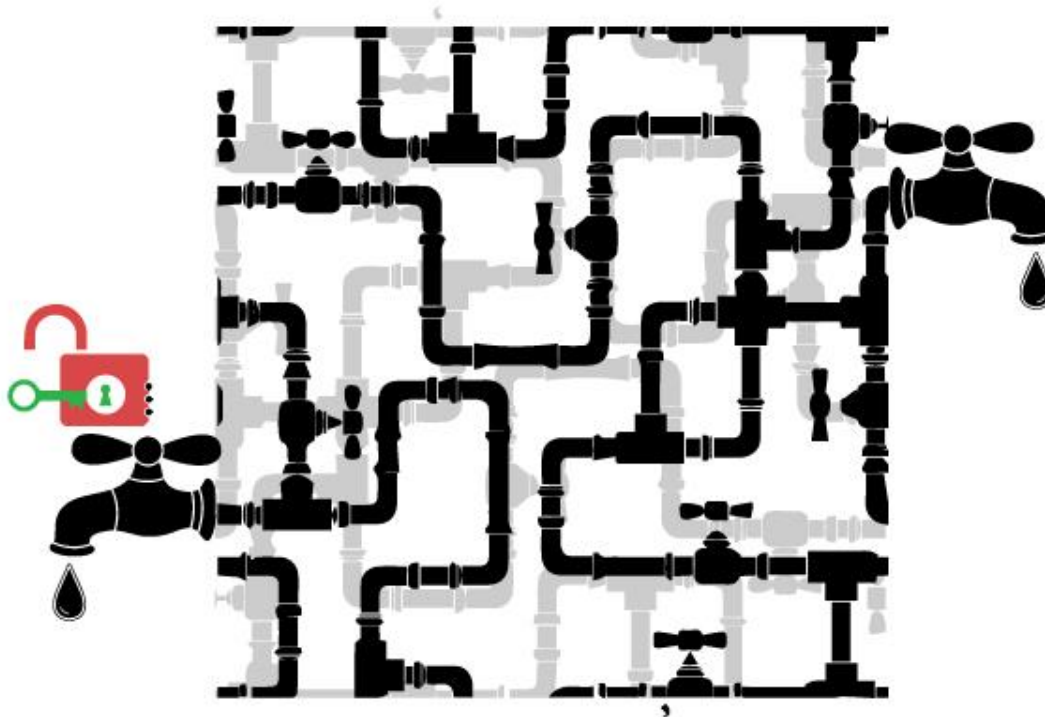
Data Scientists take traditional and emerging datasets and re-weave them into data-strands, and then into user-friendly analytics dashboards --- by first building tools for each data-strand that are designed to be integrated back together.



# Application Programming Interface (APIs)

APIs are the data “pipes.”

APIs allow data scientists to build user-friendly software to analyze and visualize complex data.

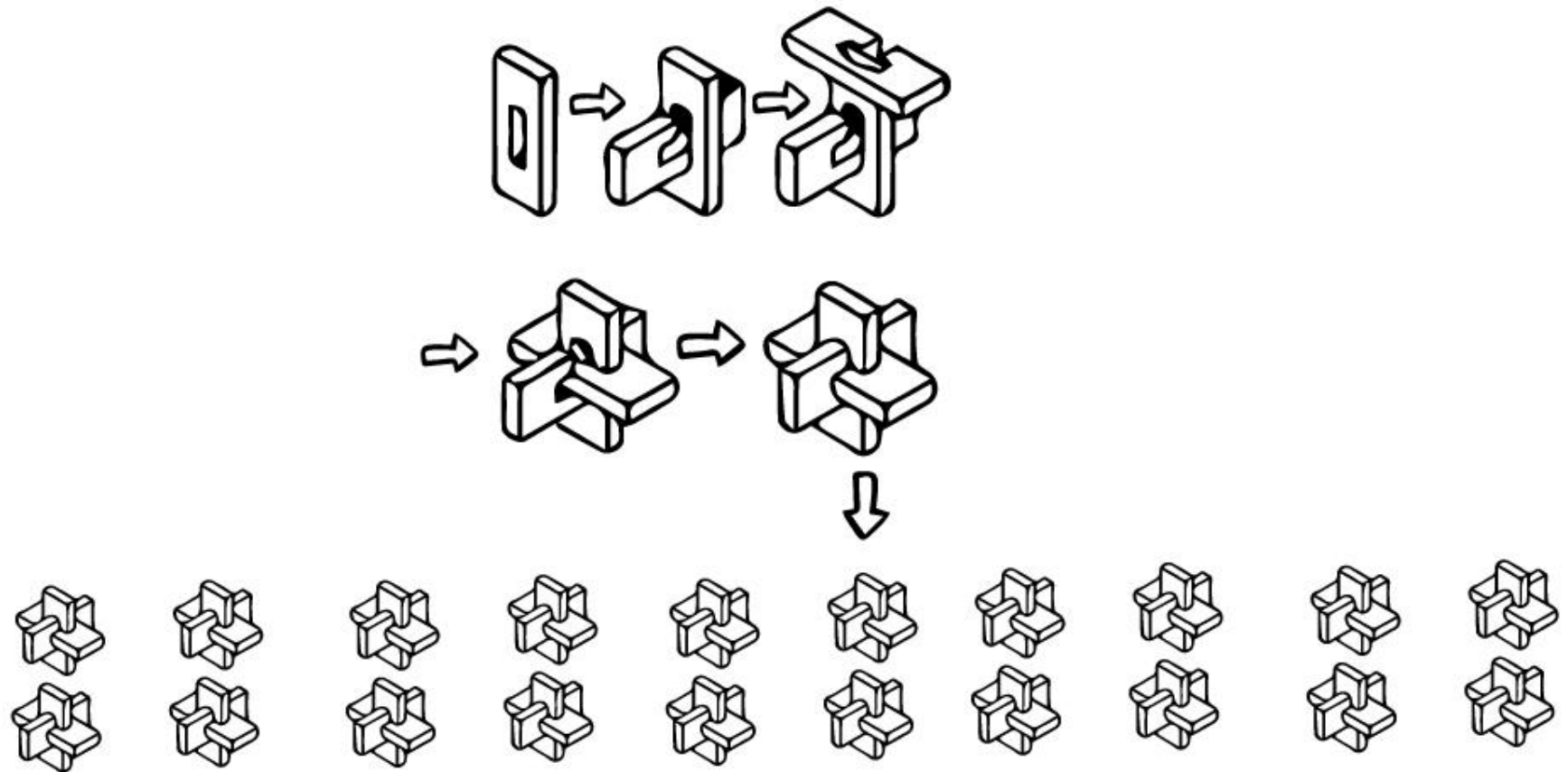


Data Scientists can turn on different faucets for different users with different levels of access

# Open Data

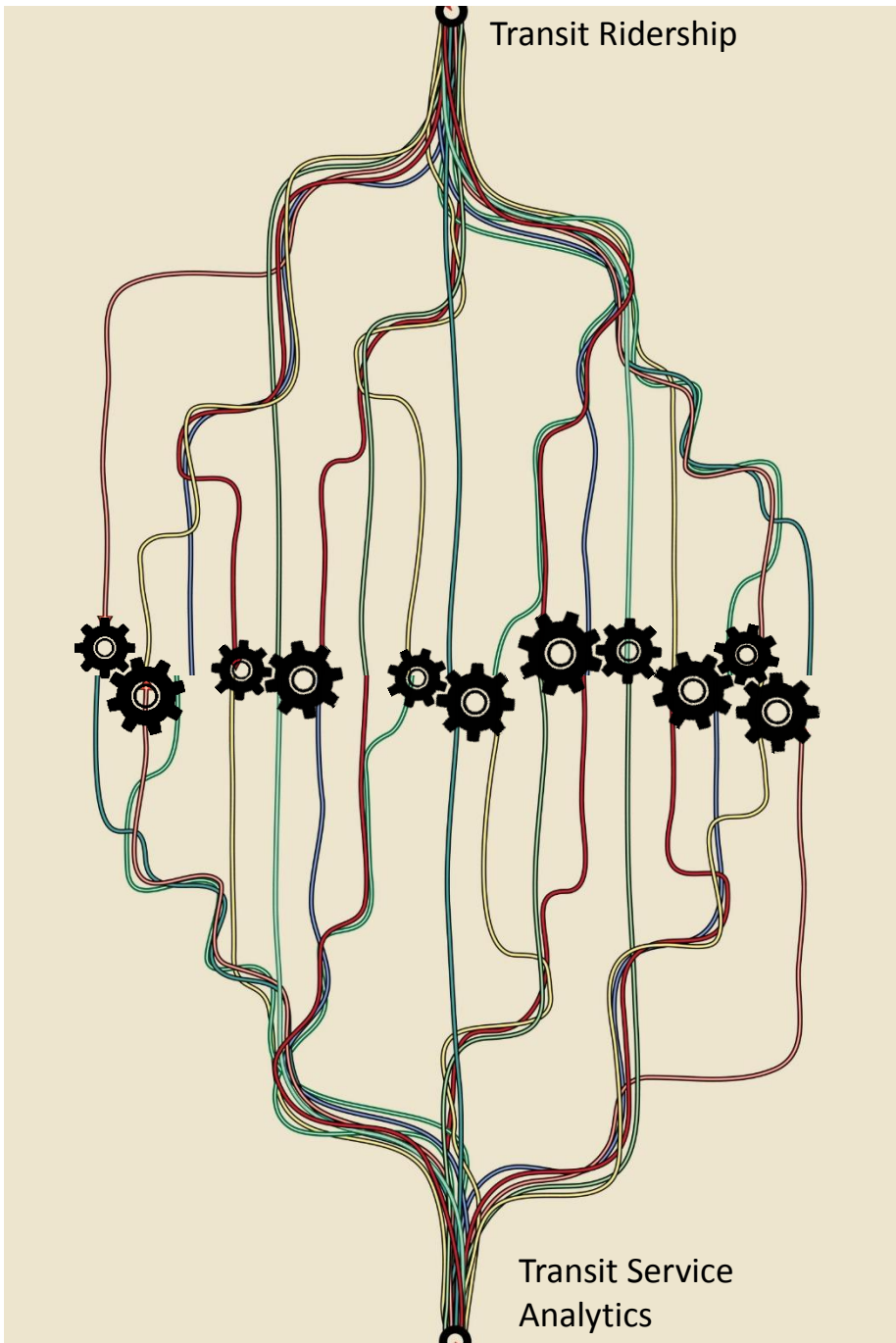
Allows the assembly of data elements using a consensus process  
like building blocks

## Data Specifications + APIs = Universal Applicability



**E.G., General Transit Feed Specification (GTFS)**

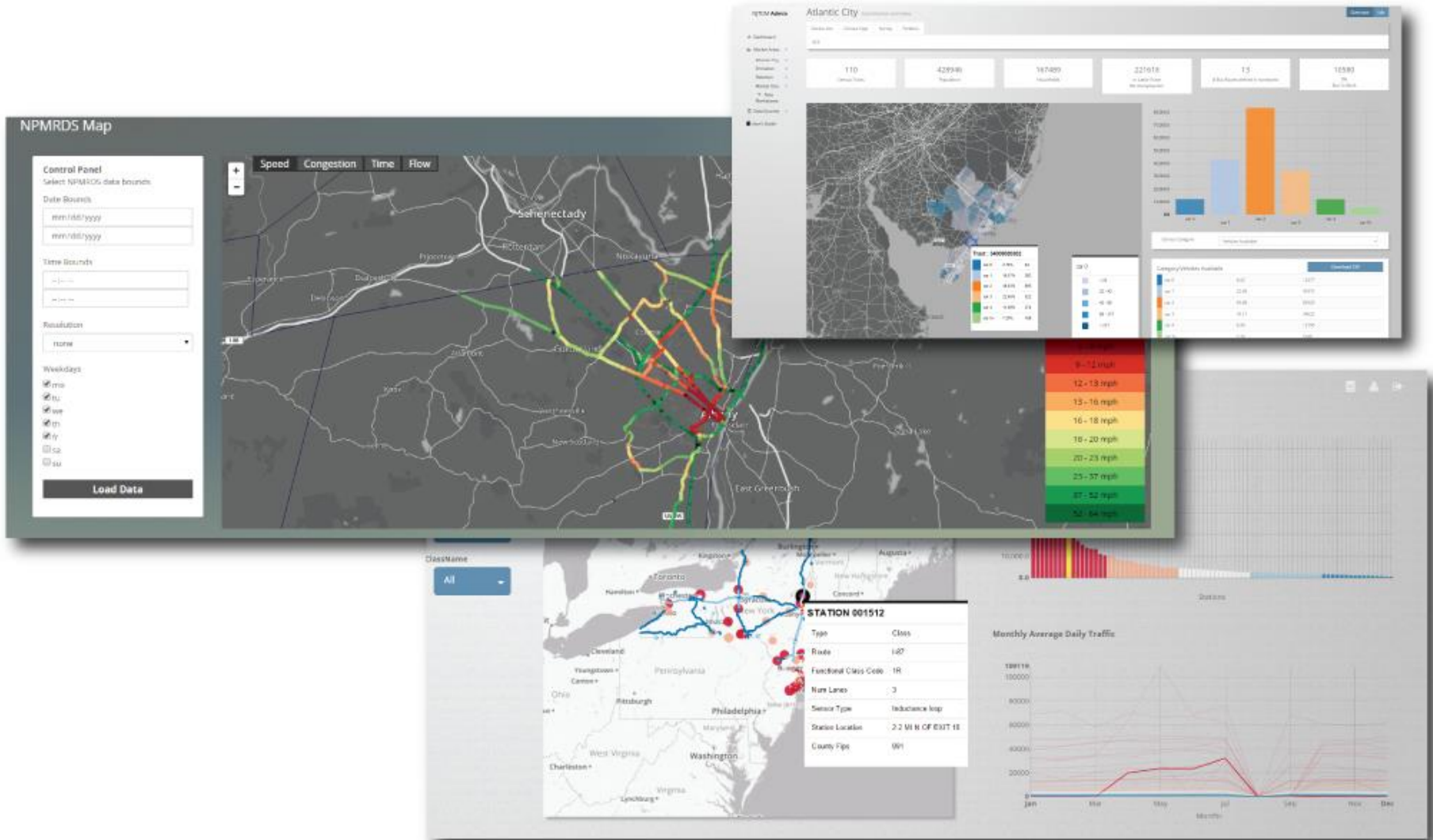




- What kinds of data-driven tools would help State DOTs and MPOs work more effectively and efficiently with traffic and transit operations?
- What kinds of data specifications could we collectively deploy to help all of us analyze networks for planning and operations?
- Can we pool together the development of data-driven analytics tools for researchers now, and for the workforce of tomorrow?
- Can Universities take the lead in forming research partnerships with multiple agencies to leverage data science to meet these challenges?

# Web-based Open Source Data Science Research Tools

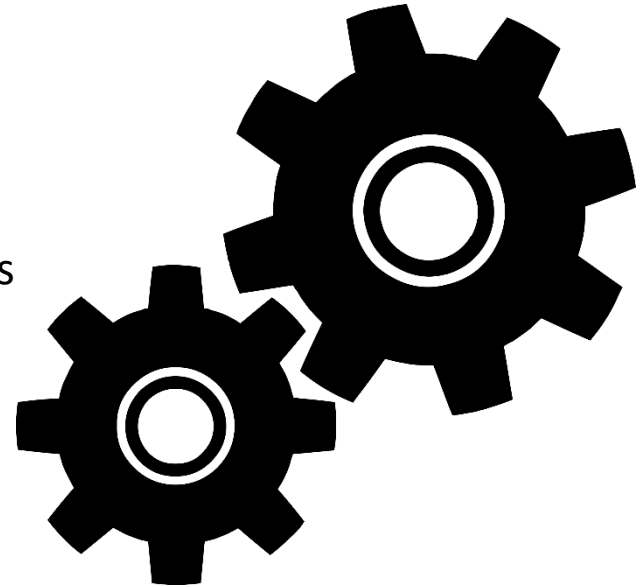
Open Sourcing the research products allows multiple teams to customize the software interface, develop new tools, and build upon legacy research



# Bringing Data Science to Transportation

## The Problem:

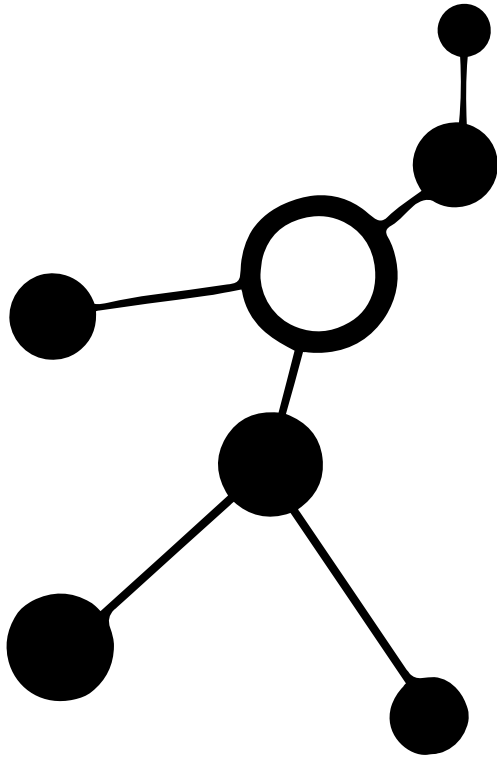
- There simply aren't enough data scientists.
- Programmers from Computer Science Departments come equipped with math, statistics and coding skills but aren't typically trained on how to apply computer science to domain specific problems, or how to design user interfaces.
- Web-developers typically don't focus on data.
- Domain specific researchers don't typically have computer science or design skills.



**One Data Scientist =  
Computer Programming Skills + Domain Specific Knowledge + User Interface Web-Design Skills**



# The Solution:



The Data Science Team Approach: building data science teams with transportation experts, computer programmers, graphic designers, plus interns.

Foster a national effort to create University partnerships using open source resources to develop web-based platforms (e.g., NYSDOT's Planning Analytics Data Warehouse (PADWare)).

Following the principles and practices of data science would create tremendous benefits and make it possible to meet the challenge of developing performance measures and monitoring strategies as a cooperative and collaborative community of transportation professionals and researchers.