

# Data, Design, and Engagement

Lessons from 30+  
Data for Social Good Projects

# About me

**Philosophy**

*then*

**Software Engineering**

*then*

**Data Science**

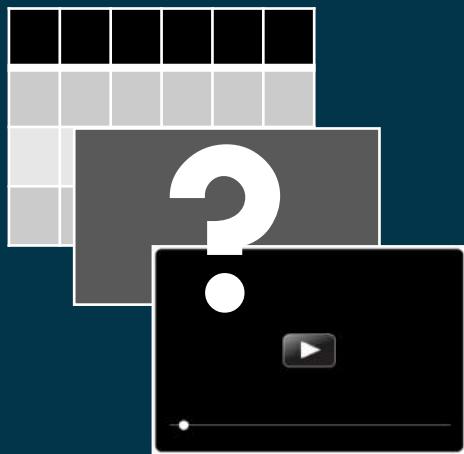
Co-founder of **DrivenData**



**Peter Bull**

Data Scientist for Social Good

**@pjbull | @drivendataorg**



## Active Competitions



### ALCF Concept to Clinic Challenge

UNTIL EARLY 2018 **\$100,000**

Early detection of abnormal lung tissue saves lives. This is a new type of data challenge where competition-tested algorithms are just the beginning. Help build open software that brings

[LET'S GO! →](#)



### N+1 fish, N+2 fish

2 WEEKS LEFT **\$50,000**

Cod, haddock, flounder - these iconic fish have supported New England's fishing fleets for generations. When they're not swimming around the depths of the North Atlantic Ocean,

dmytro  
CURRENT LEADER

[COMPETE →](#)



### DengAI: Predicting Disease Spread

2 MONTHS, 1 WEEK LEFT

Using environmental data collected by various U.S. Federal Government agencies—from the Centers for Disease Control and Prevention to the National Oceanic and Atmospheric

Divyanshu Suri

[COMPETE →](#)



### Random Walk of the Penguins

COMPETITION HAS ENDED **\$16,000**

Data on penguin populations are limited primarily due to the fact that most monitored colonies are nearby permanent research stations. This means that any remote populations are

ambarish  
1ST PLACE

[RESULTS →](#)



### America's Next Top (Statistical) Model

COMPETITION HAS ENDED **\$1,000**

US presidential elections come but once every 4 years, and this one's a big one. There are lots of people trying to predict what will happen. Can you top them? In this challenge, you'll predict

talimmeasure  
1ST PLACE

[RESULTS →](#)



### Senior Data Science: Safe Aging with SPHERE

COMPETITION HAS ENDED **\$10,000**

Contribute to open, cutting-edge research on the use of wearables in promoting health and independence for seniors. Passive monitoring can help detect when something is wrong, all

Daniel\_FG  
1ST PLACE

[RESULTS →](#)

# DRIVEN DATA



### Warm Up: Predict Blood Donations

4 MONTHS, 2 WEEKS LEFT

Can you predict whether a donor will return to donate blood given their donation history? Good data-driven systems for predicting donations and supply needs can improve the entire

rcking2  
CURRENT LEADER

[COMPETE →](#)



### United Nations Millennium Development Goals

4 MONTHS, 3 WEEKS LEFT

The UN's Millennium Development Goals provide the big-picture perspective on international development. We can use these indicators to help understand where to

hristo.buyuklie...  
CURRENT LEADER

[COMPETE →](#)



### Pump it Up: Data Mining the Water Table

5 MONTHS, 1 WEEK LEFT

Can you predict which water pumps are faulty? Using data from Taarifa and the Tanzanian Ministry of Water, can you predict which waterpoints are functional, which need some repairs,

nxox  
CURRENT LEADER

[COMPETE →](#)



### From Fog Nets to Neural Nets

COMPETITION HAS ENDED **\$15,000**

Your challenge is to develop a model that will predict the yield of Dar Si Hmad's fog nets for every day during an evaluation period, using historical data about meteorological conditions and

ulery  
1ST PLACE

[RESULTS →](#)



### Naive Bees Classifier

COMPETITION HAS ENDED **\$5,000**

Metis wants to know: can you identify a bee as a honey bee or a bumble bee? These bees have different behaviors and appearances, but given the variety of backgrounds, positions, and image

ea  
1ST PLACE

[RESULTS →](#)



### Keeping it Fresh: Predict Restaurant Inspections

COMPETITION HAS ENDED **\$5,000**

The City of Boston regularly inspects every restaurant to monitor and improve food safety and public health. As in most cities, health inspections are generally random, which can increase

LilianaMedina  
1ST PLACE

[RESULTS →](#)



### Reboot: Box-Plots for Education

1 YEAR, 5 MONTHS LEFT

We're rebooting our first prized competition for fun and education! Budgets for schools and school districts are huge, complex, and unwieldy. It's no easy task to digest

marieligh  
CURRENT LEADER

[COMPETE →](#)



### Countable Care: Modeling Women's Health Care Decisions

COMPETITION HAS ENDED **\$5,000**

Recent literature suggests that the demand for women's health care will grow over 6% by 2020. Given how rapidly the health landscape has been changing over the last 15 years, it's

giba  
1ST PLACE

[RESULTS →](#)



### Box-Plots for Education

COMPETITION HAS ENDED **\$7,500**

Budgets for schools and school districts are huge, complex, and unwieldy. It's no easy task to digest where and how schools are using their resources. Education Resource

quocle  
1ST PLACE

[RESULTS →](#)



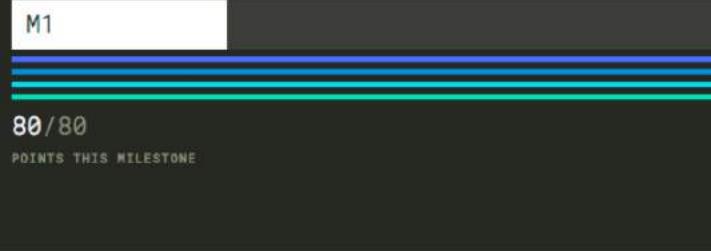
# CONCEPT TO CLINIC

A NEW KIND OF ML COMPETITION



2017-08-04

2017-10-27



## Point Tally

## Prediction

MVP	
@tdraebing	20
@reinakano	5
@musale	5
@anushthakalia	3
@foo-bar-baz-qux	3

## Interface frontend

MVP	
@musale	5
@WGierke	3

## Interface backend

MVP	
@WGierke	5
@poppingtonic	5

concept-to-clinic / concept-to-clinic

Unwatch 22 ⚡ Unstar 131 Fork 62

Code Issues 32 Pull requests 11 Settings Insights

Filters is:issue is:open Labels Milestones New issue

32 Open 22 Closed

Author Labels Projects Milestones Assignee Sort

**New API endpoint to fetch dataset directory** #77 opened 3 days ago by rahit 0 of 3

**140MB fresh repo size (without much code or data, not using LFS)** community minor official structure-or-tooling #59 opened 10 days ago by WGierke 5 of 5

**Create a template for algorithm documentation** community docs minor official #58 opened 12 days ago by tdraebing

**Document how to update documentation** community docs medium official #41 opened 26 days ago by pjbull 0 of 5 1-mvp

**Show "X candidates found" on the "Detect and select" page** feature interface-frontend minor official #37 opened 26 days ago by lamby 1-mvp

**Accept/reject nodule candidates interface** feature interface-frontend official #36 opened 26 days ago by lamby 1-mvp

**Show list of candidates nodules on "Detect and select" page** feature interface-frontend major official #35 opened 26 days ago by lamby 1-mvp

**Add ability to set right/left lung for candidate nodule** feature interface-frontend major official #34 opened 26 days ago by lamby 1-mvp

**Display metadata about an image prior to selection** feature interface-frontend major official #33 opened 26 days ago by lamby 1-mvp

# Cookiecutter Data Science

*A logical, reasonably standardized, but flexible project structure for doing and sharing data science work.*



Deon is a command line tool that allows you to easily add an ethics checklist to your data science projects.

BILL & MELINDA  
GATES foundation





FANDANGO<sup>®</sup>  
MOVIECLIPS





Bloomberg  
Business

# 404

Error! Error! Error!

**404. PAGE NOT FOUND**

Unfortunately, this page does not exist. Please check your URL or return to the [Home Page](#).

# Data for Impact Field Guide

## **Finding a Project**

Identifying and scoping the project is the hardest work and the biggest need

## **Launching a Project**

The bar for data ethics is higher in the social sector

## **Running a Project**

We need a process that builds trust between technologists and the social sector

## **Wrapping up a Project**

The capacity gap is wide and jeopardizes handoffs

# Data for Impact Field Guide

## **Finding a Project**

Identifying and scoping the project is the hardest work and the biggest need

## **Launching a Project**

The bar for data ethics is higher in the social sector

## **Running a Project**

We need a process that builds trust between technologists and the social sector

## **Wrapping up a Project**

The capacity gap is wide and jeopardizes handoffs

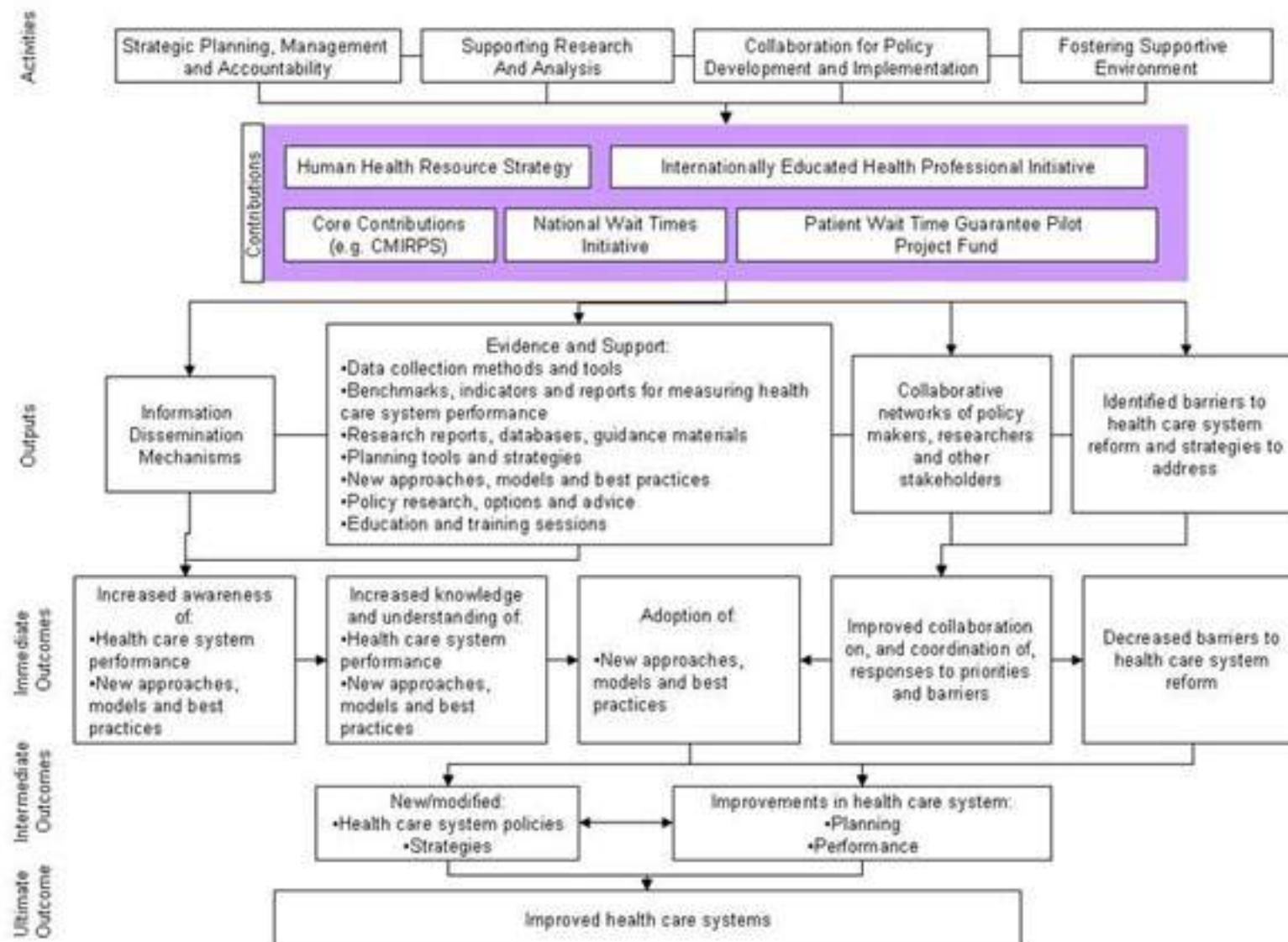
A question for social sector organizations:

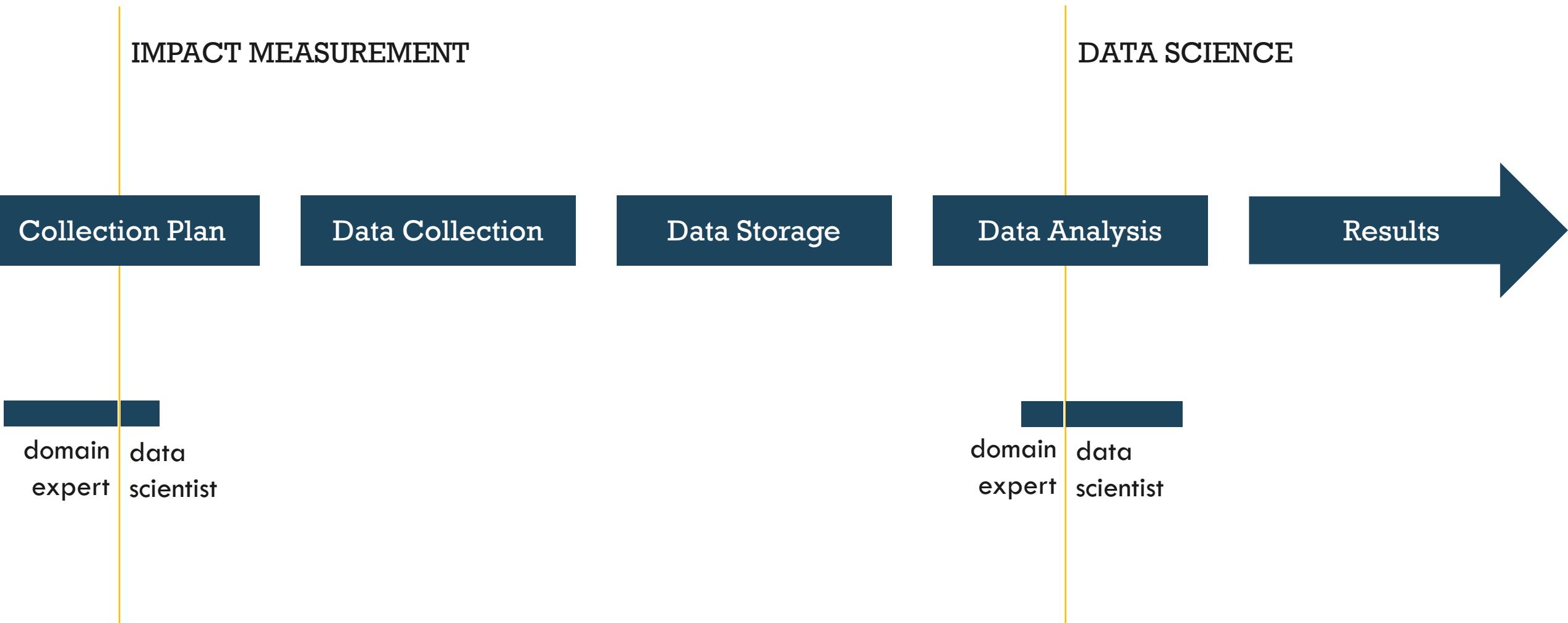
How can you use data more effectively?

Answer:

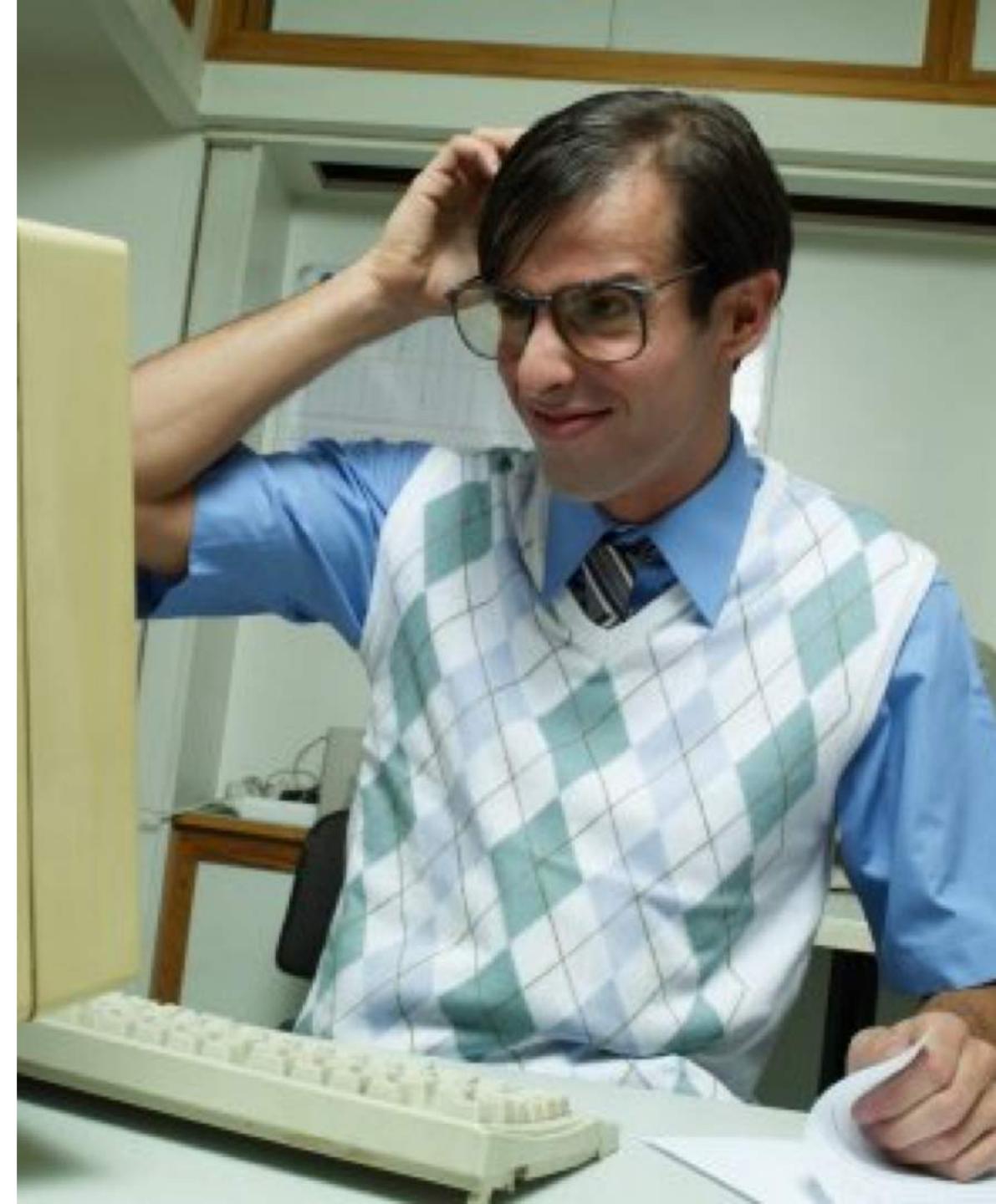
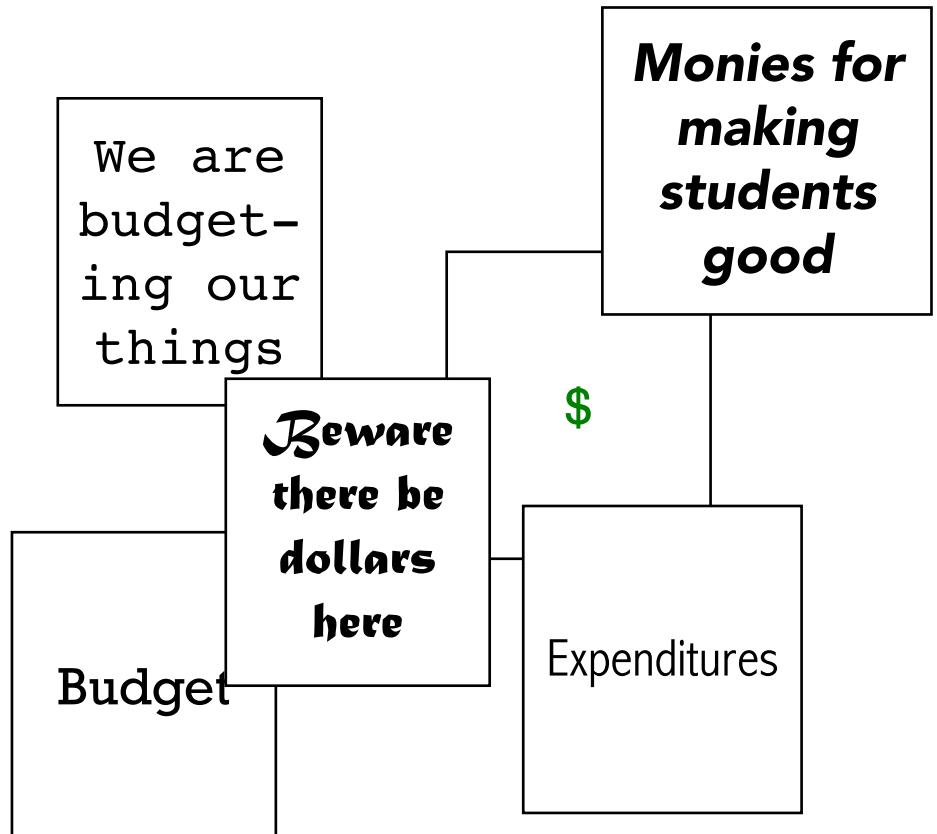
Measure Impact!

Figure 2: Health Care Policy Contribution Program – Logic Model





# Education Resource Strategies



Use	Sharing	Reporting	Student_Type	Position_Type	Object_Type	Pre_K	Operating_Status	Object_ID
sation	Instruction	School Reported	School	NO_LABEL	Teacher	NO_LABEL	NO_LABEL	PreK-12 Operating
	NO_LABEL	NO_LABEL	NO_LABEL	NO_LABEL	NO_LABEL	NO_LABEL	Non-Operating	CONTR
sation	Instruction	School Reported	School	Unspecified	Teacher	Base Salary/Compensation	Non PreK	PreK-12 Operating
nsensation	Instruction	School Reported	School	Unspecified	Substitute	Benefits	NO_LABEL	PreK-12 Operating
nsensation	Instruction	School Reported	School	Unspecified	Teacher	Substitute Compensation	NO_LABEL	PreK-12 Operating
enance	O&M	School Reported	School	Unspecified	Custodian	Benefits	NO_LABEL	PreK-12 Operating
erials & Supplies	Instruction	School Reported	School	Special Education	Non-Position	Supplies/Materials	NO_LABEL	PreK-12 Operating
	O&M	School on Central Budgets	Non-School	Unspecified	Coordinator/Manager	Benefits	NO_LABEL	PreK-12 Operating
sation	Instruction	School Reported	School	Unspecified	Teacher	Benefits	NO_LABEL	PreK-12 Operating
nsensation	Instruction	School Reported	School	Unspecified	Substitute	Benefits	NO_LABEL	PreK-12 Operating
	NO_LABEL	NO_LABEL	NO_LABEL	NO_LABEL	NO_LABEL	NO_LABEL	Non-Operating	EMPLOY
erials & Supplies	Instruction	School Reported	School	NO_LABEL	NO_LABEL	Supplies/Materials	NO_LABEL	PreK-12 Operating
sation	Instruction	School Reported	School	Unspecified	Teacher	Base Salary/Compensation	Non PreK	PreK-12 Operating
	NO_LABEL	NO_LABEL	NO_LABEL	NO_LABEL	NO_LABEL	NO_LABEL	Non-Operating	Person
nsensation	Instruction	School Reported	School	Unspecified	Substitute	Benefits	NO_LABEL	PreK-12 Operating
	NO_LABEL	NO_LABEL	NO_LABEL	NO_LABEL	NO_LABEL	NO_LABEL	Non-Operating	EQUIPM
erials & Supplies	Instruction	School Reported	School	NO_LABEL	NO_LABEL	Supplies/Materials	NO_LABEL	PreK-12 Operating
sation	Instruction	School Reported	School	Unspecified	Teacher	Base Salary/Compensation	Non PreK	PreK-12 Operating
	NO_LABEL	NO_LABEL	NO_LABEL	NO_LABEL	NO_LABEL	NO_LABEL	Non-Operating	Regular
nsensation	Instruction	School Reported	School	Unspecified	Substitute	Substitute Compensation	Non PreK	PreK-12 Operating
	NO_LABEL	NO_LABEL	NO_LABEL	NO_LABEL	NO_LABEL	NO_LABEL	Non-Operating	Salaries
erials & Supplies	Instruction	School Reported	School	Poverty	NO_LABEL	Supplies/Materials	NO_LABEL	PreK-12 Operating
	O&M	Leadership & Management	Non-School	Unspecified	Other	Base Salary/Compensation	Non PreK	PreK-12 Operating
O&M	School Reported	School	Unspecified	Non-Position	Rent/Utilities	NO_LABEL	PreK-12 Operating	ELECTRI
	NO_LABEL	NO_LABEL	NO_LABEL	NO_LABEL	NO_LABEL	NO_LABEL	Non-Operating	Non-Ce
	NO_LABEL	NO_LABEL	NO_LABEL	NO_LABEL	NO_LABEL	NO_LABEL	Non-Operating	Purchas
sensation	Instruction	School Reported	School	Unspecified	Substitute	Base Salary/Compensation	NO_LABEL	PreK-12 Operating
sation	Instruction	School Reported	School	Poverty	Teacher	Base Salary/Compensation	Non PreK	PreK-12 Operating
	NO_LABEL	NO_LABEL	NO_LABEL	NO_LABEL	NO_LABEL	NO_LABEL	Operating, Not PreK-12	BONUS
sensation	Instruction	School Reported	School	Unspecified	Substitute	Substitute Compensation	Non PreK	PreK-12 Operating
sation	Instruction	School Reported	School	Poverty	Teacher	Base Salary/Compensation	Non PreK	PreK-12 Operating
	NO_LABEL	NO_LABEL	NO_LABEL	NO_LABEL	NO_LABEL	NO_LABEL	Non-Operating	Person
nsensation	Instruction	School Reported	School	NO_LABEL	NO_LABEL	EQUIPM	PreK-12 Operating	EMPLOY
sation	Instruction	School Reported	School	Unspecified	Substitute	Benefits	NO_LABEL	PreK-12 Operating
	O&M	Shared Services	Non-School	Unspecified	Teacher	Benefits	NO_LABEL	PreK-12 Operating
sation	Instruction	School Reported	School	Unspecified	Other	Other Compensation/Stipend	Non PreK	PreK-12 Operating
	NO_LABEL	NO_LABEL	NO_LABEL	NO_LABEL	NO_LABEL	NO_LABEL	Non-Operating	CONTRA
sation	Instruction	School Reported	School	Poverty	Teacher	Base Salary/Compensation	Non PreK	PreK-12 Operating
	O&M	Shared Services	Non-School	Unspecified	Other	Other Compensation/Stipend	Non PreK	PreK-12 Operating
	NO_LABEL	NO_LABEL	NO_LABEL	NO_LABEL	NO_LABEL	NO_LABEL	Non-Operating	Regular
sation	Instruction	School Reported	School	Unspecified	Substitute	Other Compensation/Stipend	Non PreK	PreK-12 Operating
	O&M	Shared Services	Non-School	Unspecified	Teacher	Other Compensation/Stipend	Non PreK	PreK-12 Operating
	NO_LABEL	NO_LABEL	NO_LABEL	NO_LABEL	NO_LABEL	NO_LABEL	Non-Operating	Extra Du
sation	Instruction	School Reported	School	Unspecified	Other	Other Compensation/Stipend	Non PreK	PreK-12 Operating
	O&M	Shared Services	School	Unspecified	Teacher	Benefits	NO_LABEL	PreK-12 Operating
sation	Instruction	School Reported	School	Poverty	Teacher	Base Salary/Compensation	Non PreK	PreK-12 Operating
	O&M	Shared Services	Non-School	Unspecified	Other	Other Compensation/Stipend	Non PreK	PreK-12 Operating
	NO_LABEL	NO_LABEL	NO_LABEL	NO_LABEL	NO_LABEL	NO_LABEL	Non-Operating	Purchas
O&M	Leadership & Management	Non-School	Unspecified	Coordinator/Manager	Base Salary/Compensation	NO_LABEL	PreK-12 Operating	SALARIE
sensation	Instruction	School Reported	School	Unspecified	Substitute	Base Salary/Compensation	NO_LABEL	PreK-12 Operating
O&M	School Reported	School	Unspecified	Other	Base Salary/Compensation	NO_LABEL	PreK-12 Operating	OTHER S
sation	Instruction	School Reported	School	Unspecified	Teacher	Benefits	NO_LABEL	PreK-12 Operating
lity Relations	Pupil Service	School Reported	School	Poverty	Non-Position	Supplies/Materials	NO_LABEL	PreK-12 Operating
sation	Instruction	School Reported	School	Unspecified	Teacher	Other Compensation/Stipend	Non PreK	PreK-12 Operating
sation	Instruction	School Reported	School	Poverty	Teacher	Other Compensation/Stipend	Non PreK	PreK-12 Operating
nsensation	Instruction	School Reported	School	Unspecified	Substitute	Substitute Compensation	NO_LABEL	PreK-12 Operating
Tutoring	Instruction	School Reported	School	Poverty	Teacher	Other Compensation/Stipend	NO_LABEL	PreK-12 Operating
sation	Instruction	School Reported	School	Unspecified	Teacher	Other Compensation/Stipend	NO_LABEL	PreK-12 Operating
	O&M	School Reported	School	Unspecified	Teacher	Other Compensation/Stipend	NO_LABEL	PreK-12 Operating
sation	Instruction	School Reported	School	Unspecified	Custodian	Other Compensation/Stipend	Non PreK	PreK-12 Operating
	O&M	School Reported	School	Unspecified	Custodian	Other Compensation/Stipend	Non PreK	PreK-12 Operating
	NO_LABEL	NO_LABEL	Non-School	NO_LABEL	NO_LABEL	Other Compensation/Stipend	Non PreK	PreK-12 Operating
erials & Supplies	Instruction	School Reported	School	Unspecified	Teacher	Other Compensation/Stipend	NO_LABEL	PreK-12 Operating
sensation	Instruction	School Reported	School	NO_LABEL	NO_LABEL	Supplies/Materials	NO_LABEL	PreK-12 Operating
enance	Instruction	School Reported	School	Unspecified	Substitute	Substitute Compensation	NO_LABEL	PreK-12 Operating
	O&M	School Reported	School	Unspecified	Custodian	Base Salary/Compensation	NO_LABEL	PreK-12 Operating
	NO_LABEL	NO_LABEL	Non-School	NO_LABEL	NO_LABEL	Supplies/Materials	NO_LABEL	PreK-12 Operating
	NO_LABEL	NO_LABEL	NO_LABEL	NO_LABEL	NO_LABEL	NO_LABEL	Non-Operating	SUPPLIE
on Program Manag	ISPD	Leadership & Management	Non-School	Special Education	Other	Other Compensation/Stipend	NO_LABEL	PreK-12 Operating
sation	Instruction	School Reported	School	ELL	Teacher	Base Salary/Compensation	NO_LABEL	PreK-12 Operating
Tutoring	Instruction	School Reported	School	Unspecified	Teacher	Other Compensation/Stipend	Non PreK	PreK-12 Operating
sation	Instruction	School Reported	School	Unspecified	Teacher	Benefits	NO_LABEL	PreK-12 Operating
	O&M	School Reported	School	Unspecified	Teacher	Base Salary/Compensation	NO_LABEL	PreK-12 Operating
enance	Instruction	School Reported	School	Unspecified	Custodian	Base Salary/Compensation	Non PreK	PreK-12 Operating
	NO_LABEL	NO_LABEL	Non-School	NO_LABEL	NO_LABEL	Supplies/Materials	NO_LABEL	PreK-12 Operating
	NO_LABEL	NO_LABEL	NO_LABEL	NO_LABEL	NO_LABEL	NO_LABEL	Non-Operating	EMPLOY
	NO_LABEL	NO_LABEL	NO_LABEL	NO_LABEL	NO_LABEL	NO_LABEL	Non-Operating	CONTRA
	O&M	School Reported	School	Unspecified	Other	Benefits	NO_LABEL	PreK-12 Operating
	NO_LABEL	NO_LABEL	Non-School	NO_LABEL	NO_LABEL	NO_LABEL	Non-Operating	OTHER
	O&M	School Reported	School	Unspecified	Custodian	Base Salary/Compensation	Non PreK	PreK-12 Operating
	NO_LABEL	NO_LABEL	School	Unspecified	Teacher	Base Salary/Compensation	NO_LABEL	PreK-12 Operating
	NO_LABEL	NO_LABEL	School	NO_LABEL	NO_LABEL	Contracted Services	NO_LABEL	PreK-12 Operating
	NO_LABEL	NO_LABEL	School	NO_LABEL	Teacher	NO_LABEL	PreK-12 Operating	OTHER
	NO_LABEL	NO_LABEL	School	NO_LABEL	NO_LABEL	NO_LABEL	PreK-12 Operating	CONTR
	NO_LABEL	NO_LABEL	School	Unspecified	Substitute	Benefits	NO_LABEL	PreK-12 Operating
	NO_LABEL	NO_LABEL	School	Unspecified	Substitute	Base Salary/Compensation	NO_LABEL	PreK-12 Operating
	NO_LABEL	NO_LABEL	School	NO_LABEL	NO_LABEL	NO_LABEL	Non-Operating	SALARIE
	NO_LABEL	NO_LABEL	School	NO_LABEL	NO_LABEL	NO_LABEL	Non-Operating	WORKS

# New Data

Regular *	TEACHER	Special Instruction	TCHR, SCNDY MATH	Certificated Employees Salaries And Wages	IN OTH CERTIFICATED PERSON	1	Disadvantaged Youth *	\$104,928.19	Title I - Disadvantaged Children/Targeted Assistance	BUILDING ALLOCATIONS
-----------	---------	---------------------	------------------	---	----------------------------	---	-----------------------	--------------	--	----------------------

## Predictions!

LABEL COLUMN	Function	Sharing	Object_Type	Student_Type	Position_Type
MOST LIKELY	Teacher Compensation	School Reported	Base Salary/Compensation	Poverty	Teacher
2 <sup>ND</sup> LIKELY	Food Services	School on Central Budgets	Benefit	Unspecified	Coordinator/Manager
3 <sup>RD</sup> LIKELY	NO_LABEL	NO_LABEL	Supplies/Materials	NO_LABEL	NO_LABEL

# Education Resource Strategies

Saving about  
300 staff-hours  
per year

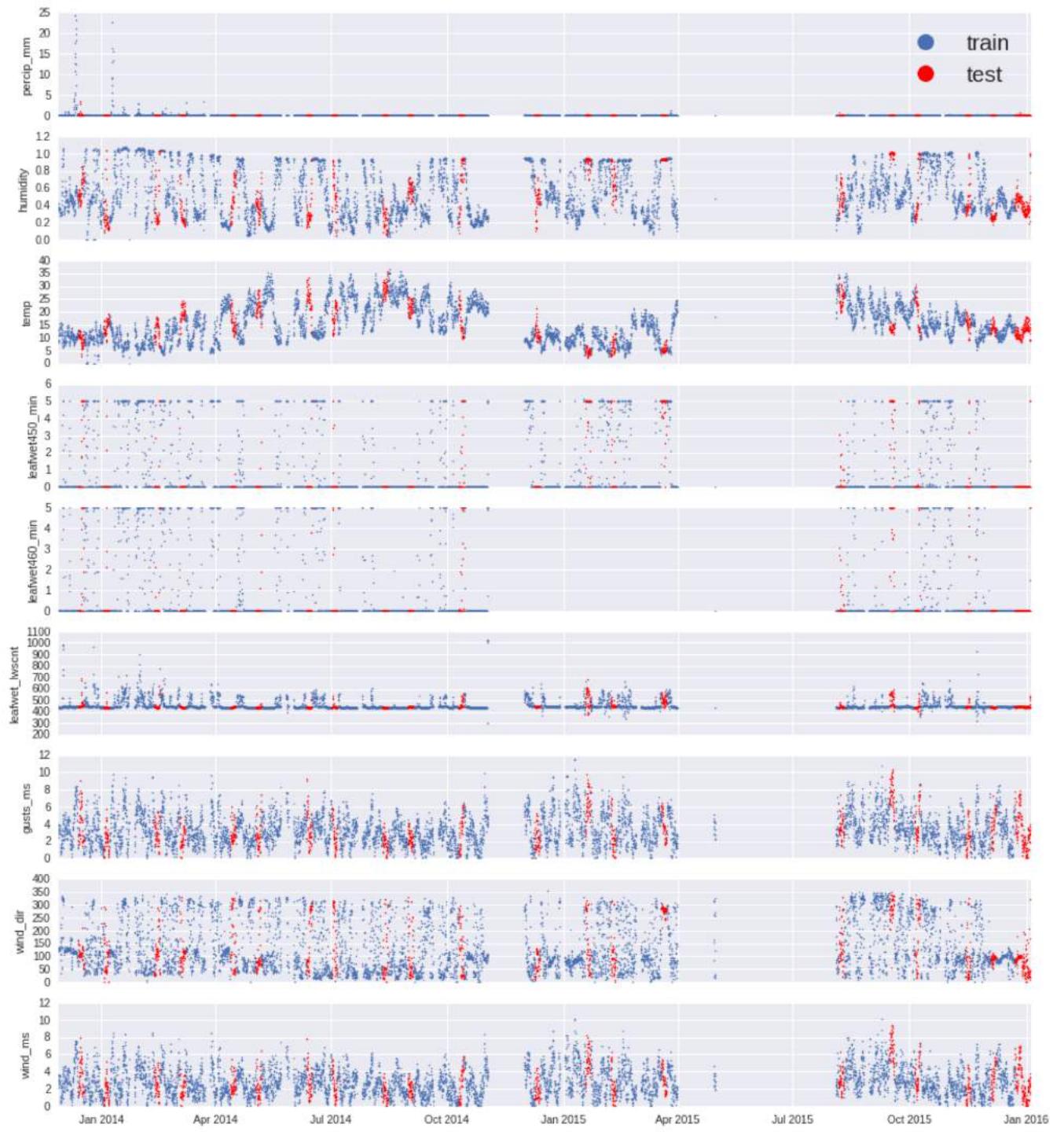


# Project Frame 1: Automation









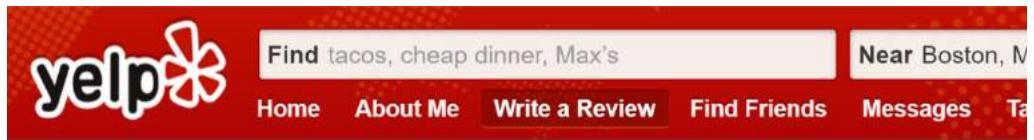


# Project Frame 2: Smarter Strategy



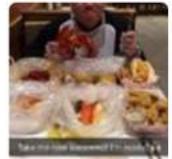


4 years of history  
8,000 locations  
315,000 violations



The screenshot shows the top navigation bar of the Yelp website. It features the Yelp logo with a stylized flower icon, a search bar with the placeholder "Find tacos, cheap dinner, Max's", and a location selector "Near Boston, M". Below the search bar are several menu links: Home, About Me, Write a Review, Find Friends, Messages, and a partially visible "Ta".

See the business you'd like to review?



1. Loui Loui



29 reviews

\$\$

Seafood, Salad, Sandwiches

101 Main St

Stoneham, MA 02180

(781) 433-2145



2. Pintxo Pincho Tapas Bar



189 reviews

\$\$

Tapas Bars

★ Write a Review

★ Write a Review

385 Main St

Woburn, MA 01801

(781) 932-1379



3. Giacomo's Ristorante



2137 reviews

\$\$

Italian

★ Write a Review

★ Write a Review

North End

355 Hanover St

Boston, MA 02113

(617) 523-9026

- **RATINGS**

- **REVIEWS**

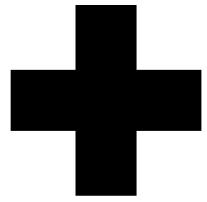
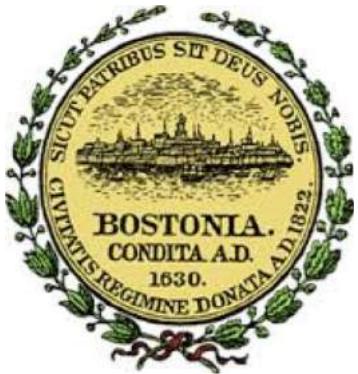
- Date of the review
- Free form text

- **CATEGORIES**

- Cuisine
- Format

- **NEIGHBORHOODS**

- **CHECKINS**



# Project Frame 3: Smarter Use of Scant Resources

# Data for Impact Field Guide

## **Finding a Project**

Identifying and scoping the project is the hard work the sector needs

## **Launching a Project**

The bar for data ethics is higher in the social sector

## **Running a Project**

We need a process that builds trust between technologists and the social sector

## **Wrapping up a Project**

The capacity gap is wide and jeopardizes handoffs

Social sector organizations exist for the public good, which demands a higher attention to data ethics.

# Data collection

**Data security:** Do we have a plan to protect and secure data (e.g., encryption at rest and in transit, access controls on internal users and third parties, access logs, and up-to-date software)?

## WHERE THINGS HAVE GONE WRONG

*Personal and financial data for more than 146 million people was stolen in Equifax data breach.*

# Modeling

**Explainability:** Can we explain in understandable terms a decision the model made in cases where a justification is needed?

## WHERE THINGS HAVE GONE WRONG

*Patients with pneumonia with a history of asthma are usually admitted to the intensive care unit as they have a high risk of dying from pneumonia.*

*Given the success of the intensive care, neural networks predicted asthmatics had a low risk of dying and could therefore be sent home. Without explanatory models to identify this issue, patients may have been sent home to die.*

# Deployment

**Redress:** Have we discussed with our organization a plan for response if users are harmed by the results (e.g., how does the data science team evaluate these cases and update analysis and models to prevent future harm)?

## WHERE THINGS HAVE GONE WRONG

*In Arkansas, an algorithm was used to allocate home care under a Medicaid waiver program to determine how frequently a caretaker would visit a patient.*

*However due to coding errors, people with diabetes or cerebral palsy were confronted with cuts to their care and with no clear mechanism for redress.*

*Eventually a court case was brought by the ACLU.*

# Data for Impact Field Guide

## **Finding a Project**

Identifying and scoping the project is the hard work the sector needs

## **Launching a Project**

The bar for data ethics is higher in the social sector

## **Running a Project**

We need a process that builds trust between technologists and the social sector

## **Wrapping up a Project**

The capacity gap is wide and jeopardizes handoffs

A memo to big tech

# The techlash against Amazon, Facebook and Google—and what they can do

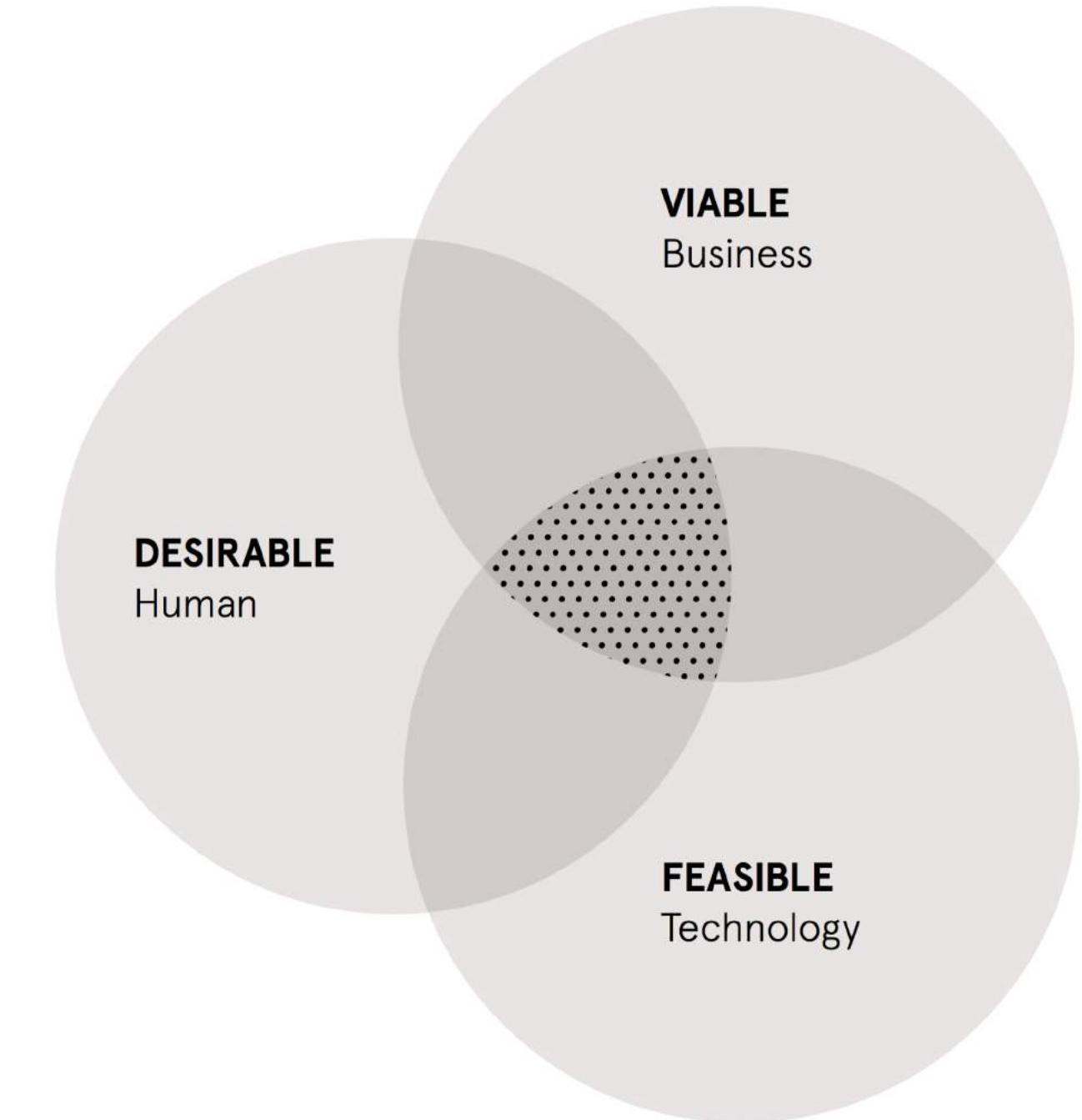
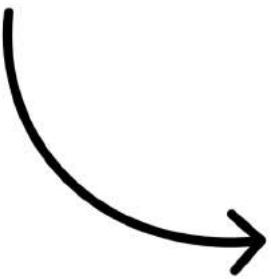
*Which antitrust remedies to welcome, which to fight*

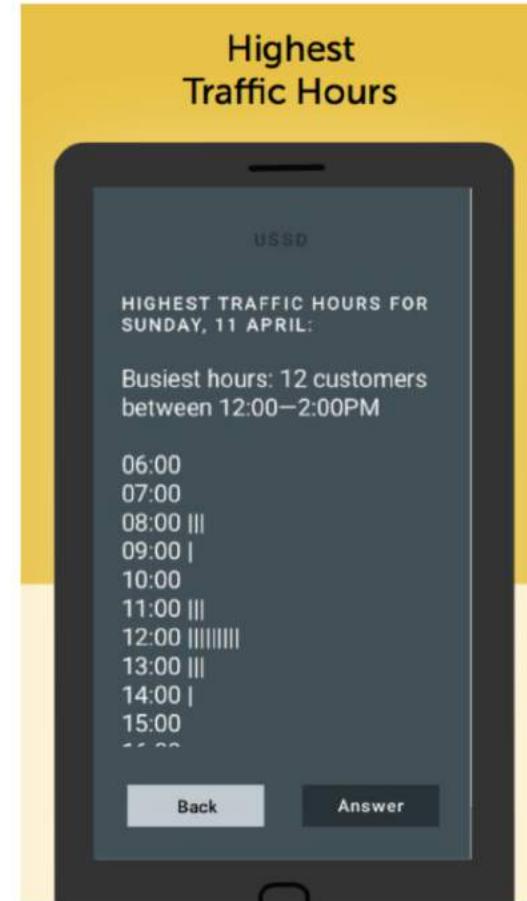
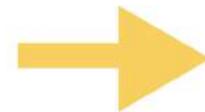
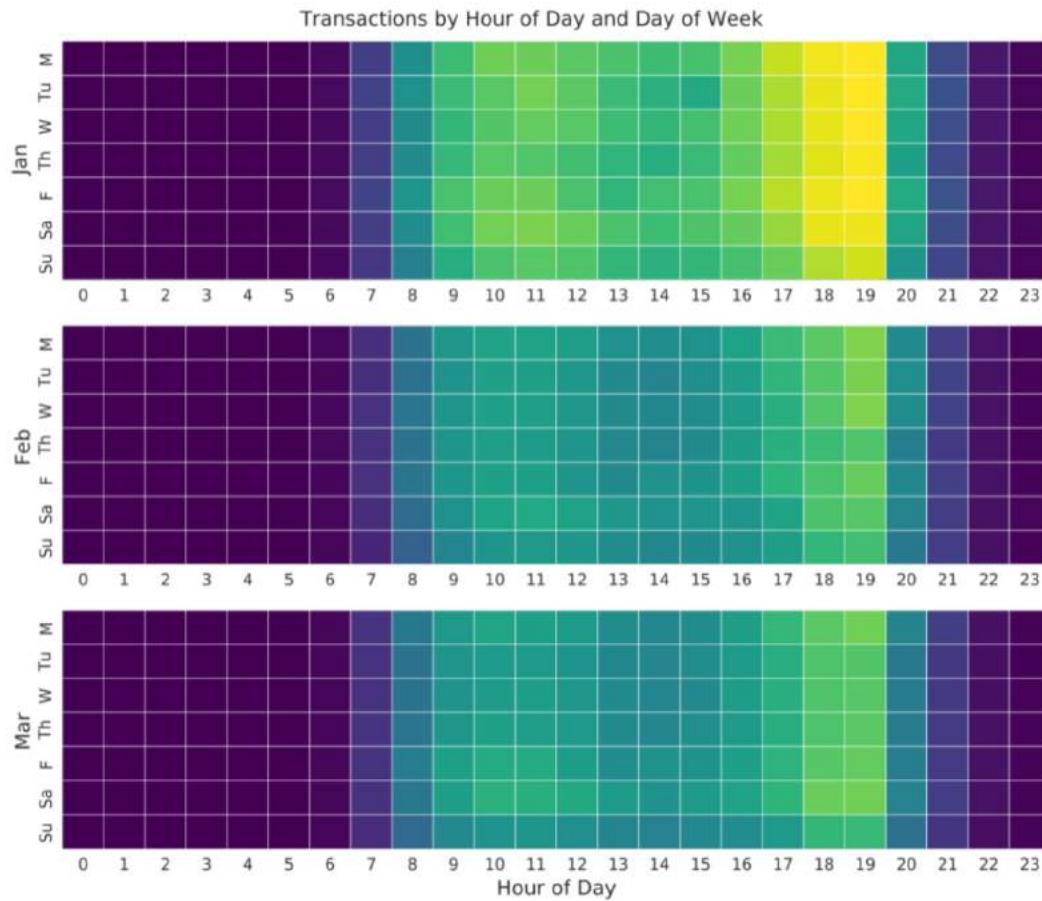


Ryan Olbrysh, AFP, EPA, Facebook, Amazon, Google

# Design Thinking

Start here





core activities of the

# Human Centered Data Scientist

1. Go to the field and observe data being generated.
2. Design with, not for, by iterating on prototypes.
3. Assess outcomes both quantitatively and qualitatively.
4. Be honest about what doesn't work and learn from it.

# Data for Impact Field Guide

## **Finding a Project**

Identifying and scoping the project is the hard work the sector needs

## **Launching a Project**

The bar for data ethics is higher in the social sector

## **Running a Project**

We need a process that builds trust between technologists and the social sector

## **Wrapping up a Project**

The capacity gap is wide and jeopardizes handoffs



# THE DATA CAPACITY GAP

140k – 180k

shortage of data  
scientists

\$118,700

average salary of  
data scientist

\$135,000

average salary  
executive director  
US nonprofit

Scale: Absolute - **Relative**



---

# Machine Learning: The High-Interest Credit Card of Technical Debt

---

**D. Sculley, Gary Holt, Daniel Golovin, Eugene Davydov,  
Todd Phillips, Dietmar Ebner, Vinay Chaudhary, Michael Young**

{dsculley, gholt, dg, edavydov}@google.com

{toddphillips, ebner, vchaudhary, mwyong}@google.com

Google, Inc

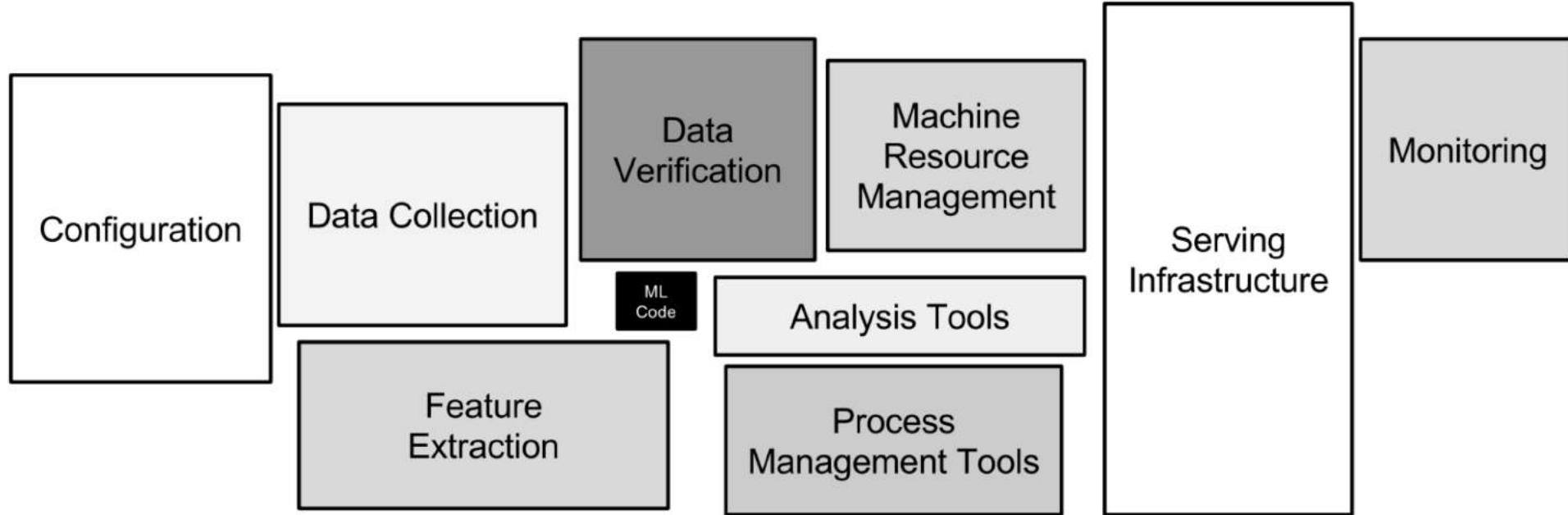


Figure 1: Only a small fraction of real-world ML systems is composed of the ML code, as shown by the small black box in the middle. The required surrounding infrastructure is vast and complex.

---

## Hidden Technical Debt in Machine Learning Systems

---

**D. Sculley, Gary Holt, Daniel Golovin, Eugene Davydov, Todd Phillips**

{dsculley, gholt, dgg, edavydov, toddphillips}@google.com

Google, Inc.

**Dietmar Ebner, Vinay Chaudhary, Michael Young, Jean-François Crespo, Dan Dennison**

{ebner, vchaudhary, mwyong, jfcrespo, dennison}@google.com

Google, Inc.

# Data for Impact Field Guide

## **Finding a Project**

Identifying and scoping the project is the hardest work and the biggest need

## **Launching a Project**

The bar for data ethics is higher in the social sector

## **Running a Project**

We need a process that builds trust between technologists and the social sector

## **Wrapping up a Project**

The capacity gap is wide and jeopardizes handoffs

# Thanks!

Any questions?