


Big Data: Untapped potential for public value

Jane Wiseman

CEO
Institute for Excellence in Government

Innovations in Government Fellow
Ash Center for Democratic Governance and Innovation
Harvard Kennedy School

 Jane.Wiseman@gmail.com

 www.linkedin.com/in/jane-wiseman/

 [@JaneMWiseman](https://twitter.com/JaneMWiseman)

innovadores

Laboratorio
de Gobierno

públicos

Mes de la Innovación
Santiago, Chile
June 26th, 2019

Laboratorio
de Gobierno



Laboratorio
de Analítica
Pública®



4 insights about Big Data for Government

1: Data is everywhere

2: Most data is not yet analyzed

3: Leading governments are beginning to demonstrate value

4: Pathway to success is open to all



1: Data is everywhere

You're already using big data (or your friends and family are)



Uses big data to tell me how I compare to others of my age/gender



Uses artificial intelligence (AI) and natural language processing (NLP) to understand what we ask for and then deliver it. Machine learning (ML) helped improve early models.



Uses IOT (Internet of things) sensors and wifi to gather data on temperature, humidity, weather, activity patterns to predict HVAC needs.

2.5 quintillion bytes of data are created a day!



Every day:

- 5 billion internet searches a day
- 3.5 billion Google searches, or 40K/second
- Half of searches are on mobile phones
- 1.5 billion use FB daily (2B total users)

Every minute:

- 14 million text messages
- 4.3 million YouTube videos
- 1 million Tinder swipes
- 500K FB comments posted

Note: US/UK 1 billion = Chile 1,000 millions



2: Most data is not analyzed

We are just beginning to address the potential for big data



- **90% of all data** ever created was created in the past 2 years, only 1% has been analyzed (McKinsey)
- **4 out of 5 government leaders** say decision-making is either somewhat or rarely data-driven (PwC)

Most companies aren't yet data-driven

- **Only 1 in 10 companies** have fully implemented their analytics programs:
 - **74%** of firms say they want to be data-driven, but **only 29%** are connecting analytics to action ([Forrester](#))
 - **72%** of marketers are still focused primarily on knowledge gathering rather than making actionable use of their data. ([Forbes](#))



3: Leading governments are beginning to demonstrate value

Data analytics successes – selected examples



- NYC fire inspectors able to identify and manage the **riskiest buildings first**
- Chicago **finds illness-causing health code violations 7 days faster**
- New Orleans **gives out free smoke alarms** where risk is high
- LA gets **firefighters to the scene faster** with mapping

The financial value of using data in govt.



- **\$5 for every \$1** in cost at HHS OIG, including \$1.3 B takedown of medical fraud.
- Boston **saves \$5 million** and 20,000 pounds of carbon emissions **with an algorithm** for smart school bus routing.
- San Diego saves \$2.4 million, use **60% less energy** w smart streetlights
- Syracuse **predicts broken water pipes** and saves \$800K per year

Breaking down silos to put people first



MA connected **23 agency data sets** in a first ever privacy-protected opioid data platform. Findings:

- 120x OD risk for formerly incarcerated individuals
- 30x risk for homeless
- 27x risk for women who deliver a child w NAS
- 6x risk for mental health

Breaking down silos to put people first



Allegheny Co. connected HHS and Medical Examiner data on and found **2/3 of OD deaths had prior county services:**

- 34% had received mental health services
- 28% had received substance use services
- 18% had been incarcerated in past year, 49% had ever been
- 51% were enrolled in Medicaid in the 90 days prior to the OD

The value of digital and data transparency



Starbucks now steams each latte separately so that we can **see the process** of our drink being made

Apple voice response system has pre-recorded typing sounds because even the **illusion of work on our behalf improves customer service**

Kayak airline search shows us the names of airlines it searches because **customer satisfaction improves** when we see the work being done on our behalf

The value of government transparency



In an experiment, Buenos Aires **increased trust in government 67%** by showing members of the public its **open data** portal

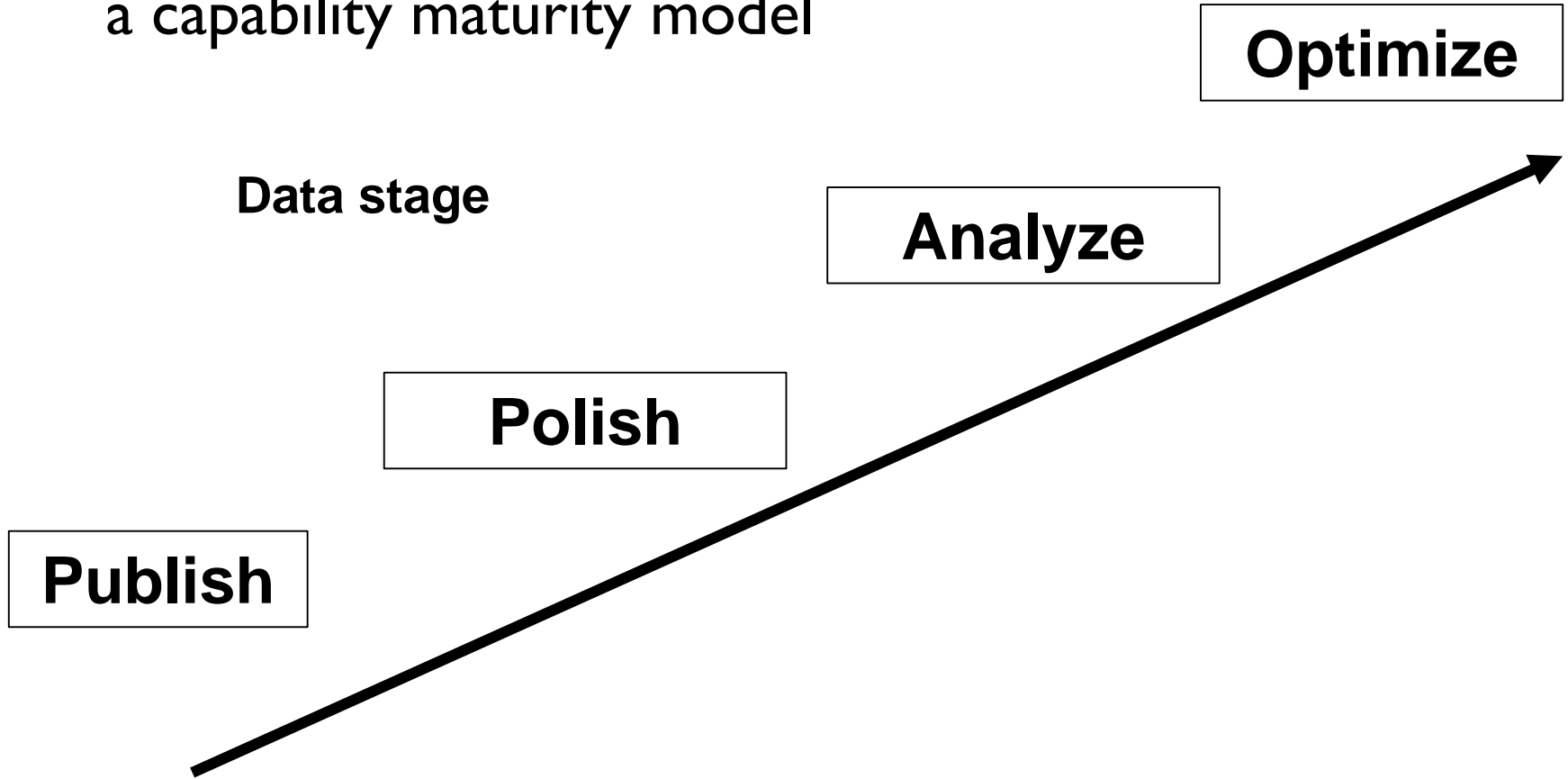
Snow plowing: **Public satisfaction went from 50% to 62%** based on open data efforts and “showing the work” via a Tweet-a-long in KCMO



4: The pathway to data maturity

Data-driven Government

a capability maturity model



Data-driven Government

a 4-stage capability maturity model

Publish

Govt is publishing open data and is beginning to develop the tools and skills to use data for decisions

Polish

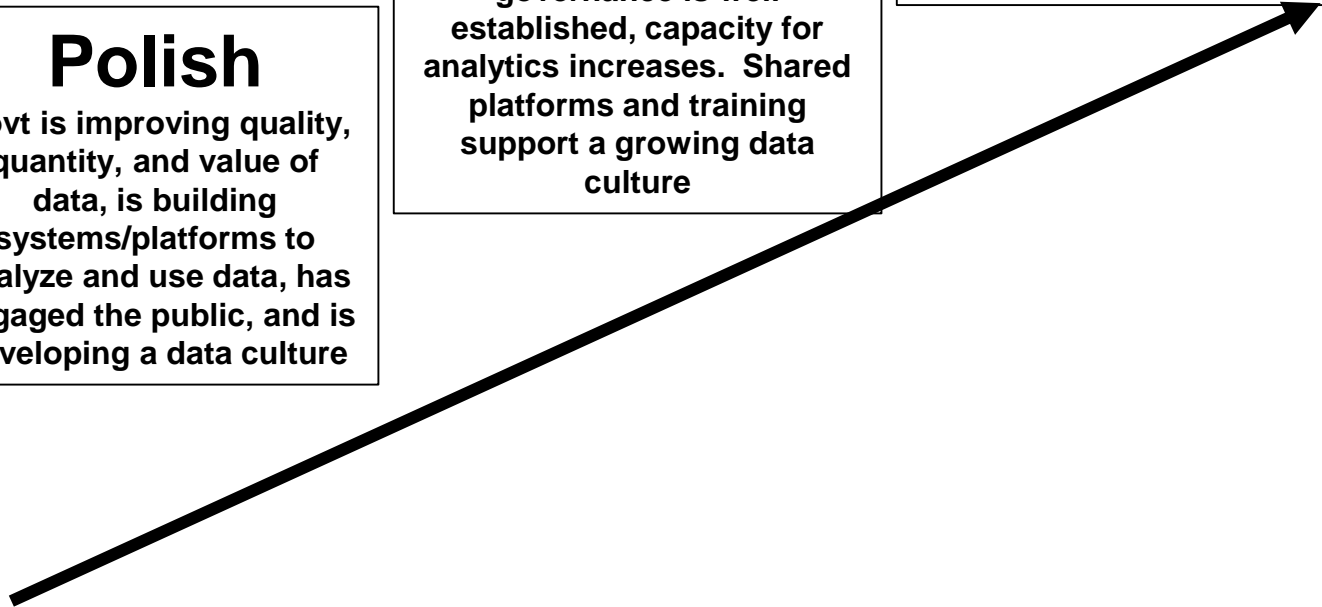
Govt is improving quality, quantity, and value of data, is building systems/platforms to analyze and use data, has engaged the public, and is developing a data culture

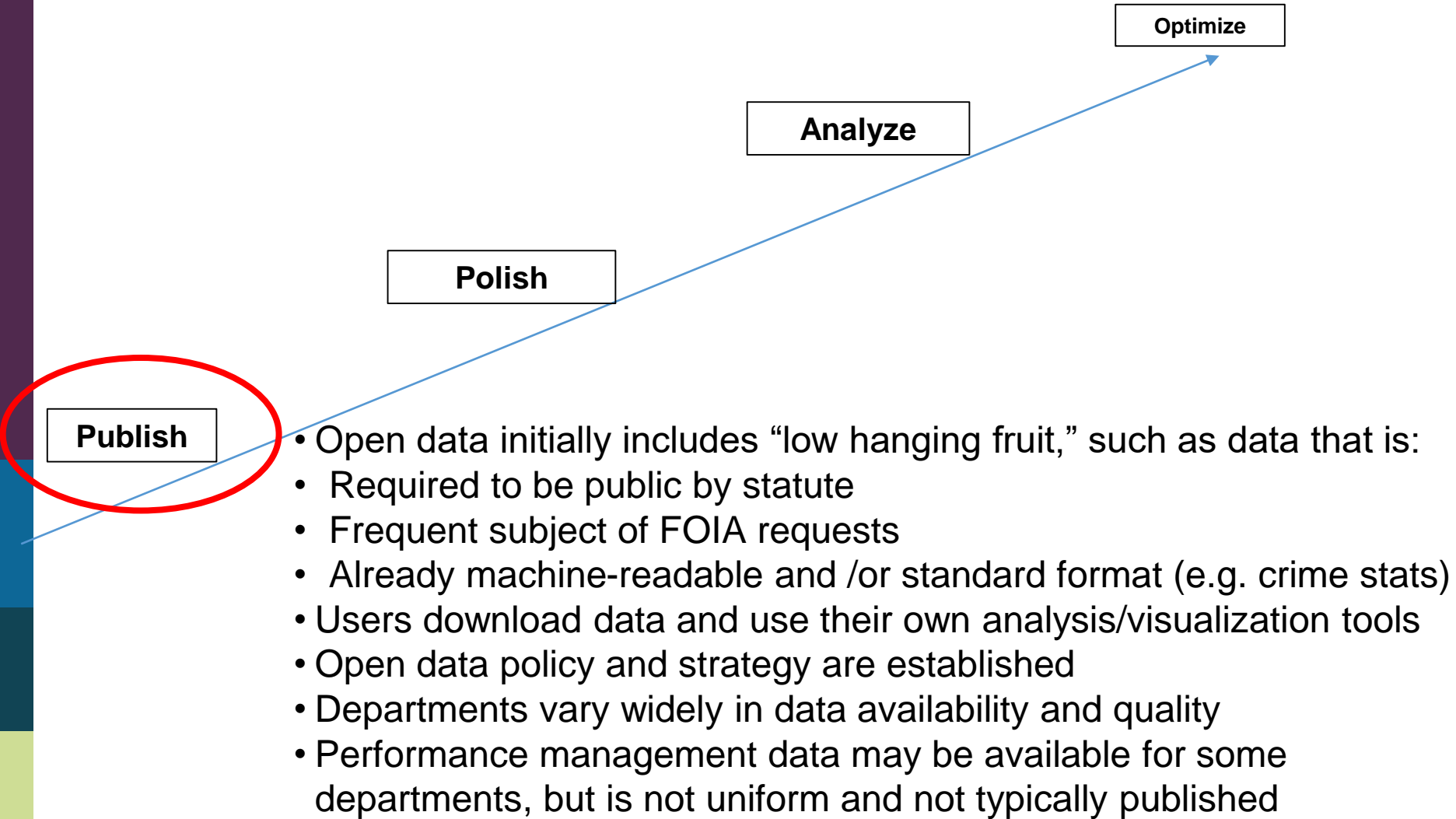
Analyze

Leaders demonstrate commitment to data. Data governance is well established, capacity for analytics increases. Shared platforms and training support a growing data culture

Optimize

Leaders and managers at all levels are committed to using data. A community of data champions provides mutual support. Successes are documented. Training supports enterprise-wide data culture



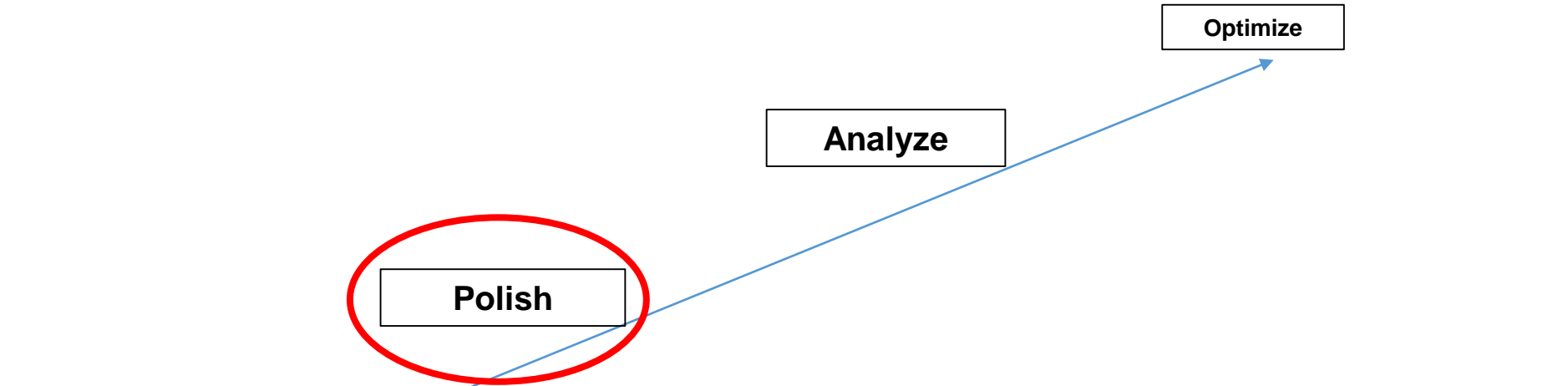


Publish

- Open data initially includes “low hanging fruit,” such as data that is:
- Required to be public by statute
- Frequent subject of FOIA requests
- Already machine-readable and /or standard format (e.g. crime stats)
- Users download data and use their own analysis/visualization tools
- Open data policy and strategy are established
- Departments vary widely in data availability and quality
- Performance management data may be available for some departments, but is not uniform and not typically published

Publish example: City of San Francisco

- Annual published data strategy
- Each department appoints a data steward
- Start with a data inventory – baseline
- Commitment to publish data each year
- Publish first what is most in demand
- Next, high priority public interest

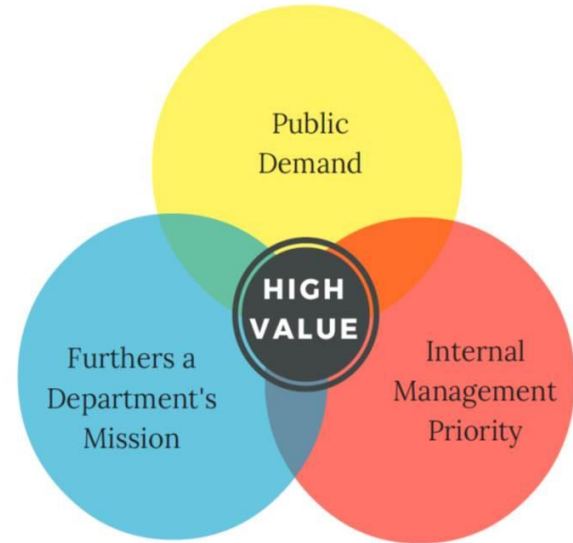


- Open data is regularly updated and improved
- Data quality and consistency assured via auditing/other feedback loops
- Platform is stable and supports large volumes of open data
- Transparency is increasingly achievable
- Formal feedback allows public input on data availability improvements
- Open data enables greater and more informed civic engagement
- Trained staff in central data office and in each department
- Data for quasi-public entities is incorporated
- APIs enable dev. of user-defined apps (transit, permits, housing, etc.)
- Performance mgt. data increasingly available in depts., shared public

Polish example: Boston data migration

- City of Boston data migration
- Sought input from the public, from government, academia
- Developed migration strategy, prioritized
- Transported only important data sets
- Discontinued publication of outdated data

PRIORITIZING DATASETS



Publish

Polish

Analyze

Optimize

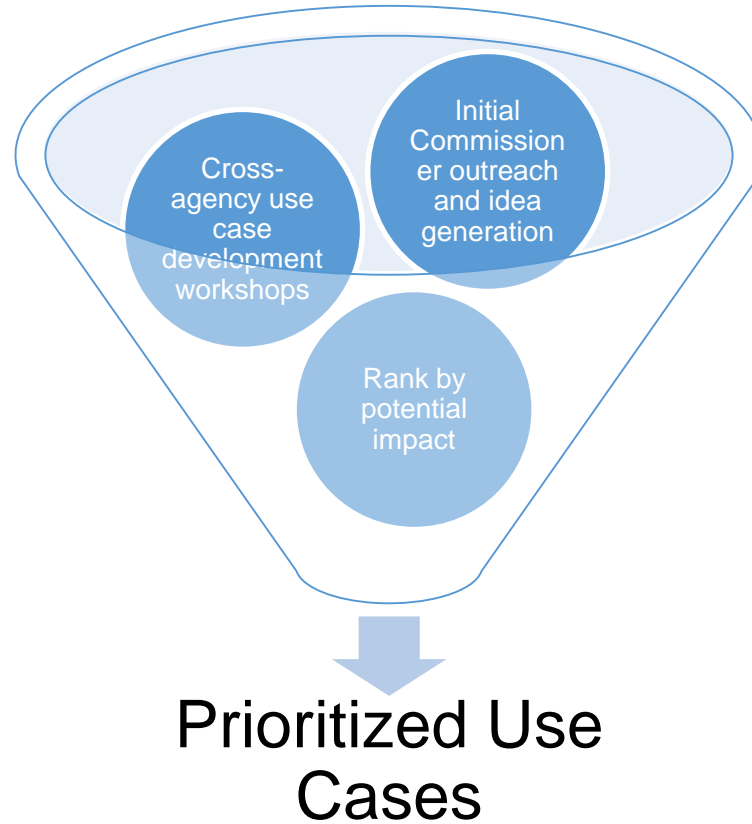
- Open data is well organized, easy for public, media, academia and govt to navigate, easy to extract and use for research and analysis
- Open data portal provides analysis and visualization tools
- Analytics officer in place with clear mission and resources
- Analytics used by data leaders in some departments
- Some cross-agency analytics projects
- Documented successes generate interest
- Growing data culture, growing data skills, training opportunities
- Open data and performance data are high quality and consistent across agencies allowing robust cross-enterprise analysis
- Performance and operational data integrated to analytics activities

Analyze example: City of Chicago



- Prioritizing selection of analytics projects
 - Policy priority of mayor
 - Manageable scope
 - Data exists to solve problem
- Rodent abatement algorithm
 - Start with straight-forward problem
 - Document success
 - Generate momentum
- Additional predictions:
 - Restaurant inspections
 - Lead paint exposure for children
 - Beach closures for contamination

Use case priority setting process



Use case prioritization rubric

1. Data Readiness	Is data associated with this RQ available and/or ready to be analyzed?	<ul style="list-style-type: none"> • Available/ready • Partly available/ready • Not available/ready
2. Policy Alignment	Does this RQ align with City policy goals—from both the Mayor's Office and applicable departments?	<ul style="list-style-type: none"> • Yes • No
3. Operational Impact	As a pilot, what level of opportunity would this RQ provide for positive operational outcomes—i.e., a reduction in time/cost when compared to current operations?	<ul style="list-style-type: none"> • High • Medium • Low
4. Resident Impact	As a pilot, what level of opportunity would this RQ directly provide benefit to the residents of Chicago?	<ul style="list-style-type: none"> • High • Medium • Low
5. Level of Use	If this pilot were to become a fully operational tool in your department, what level of use would it receive?	<ul style="list-style-type: none"> • Weekly • Monthly • Quarterly • Annually
6. Potential for Replication	Would implementing this RQ as a pilot provide a model that can be reused for other operational areas of your department, or elsewhere in the city?	<ul style="list-style-type: none"> • Yes • Unsure • No
7. Operational Change	If this pilot were to become a fully operational tool in your department, how drastically would it alter current operations?	<ul style="list-style-type: none"> • (Open-ended response)

SDP team use case development process



3 main types of analytics

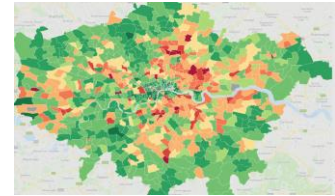
Descriptive
statistics



Predictive
analytics



GIS/mapping



Publish

Polish

Analyze

Optimize

- Analytics drive decision-making for all top priority policy areas
- Data analytics enables allocation of resources to highest public needs
- Govt leadership rely on data analytics for key decisions, taking a “problem-first” mentality, and looking for data to answer questions
- Analytics skills, data scientists exist in both central org and in depts
- Departments collaborate on inter-disciplinary data analytics projects
- Case studies of successes avail, some include playbook
- Performance mgt. optimized, robust civic engagement in providing customer feedback on performance, improves transparency and trust in govt.

Optimize example

- Who will it be?
- Chile?

Moving along the maturity curve: examples

HHS OIG

GSA

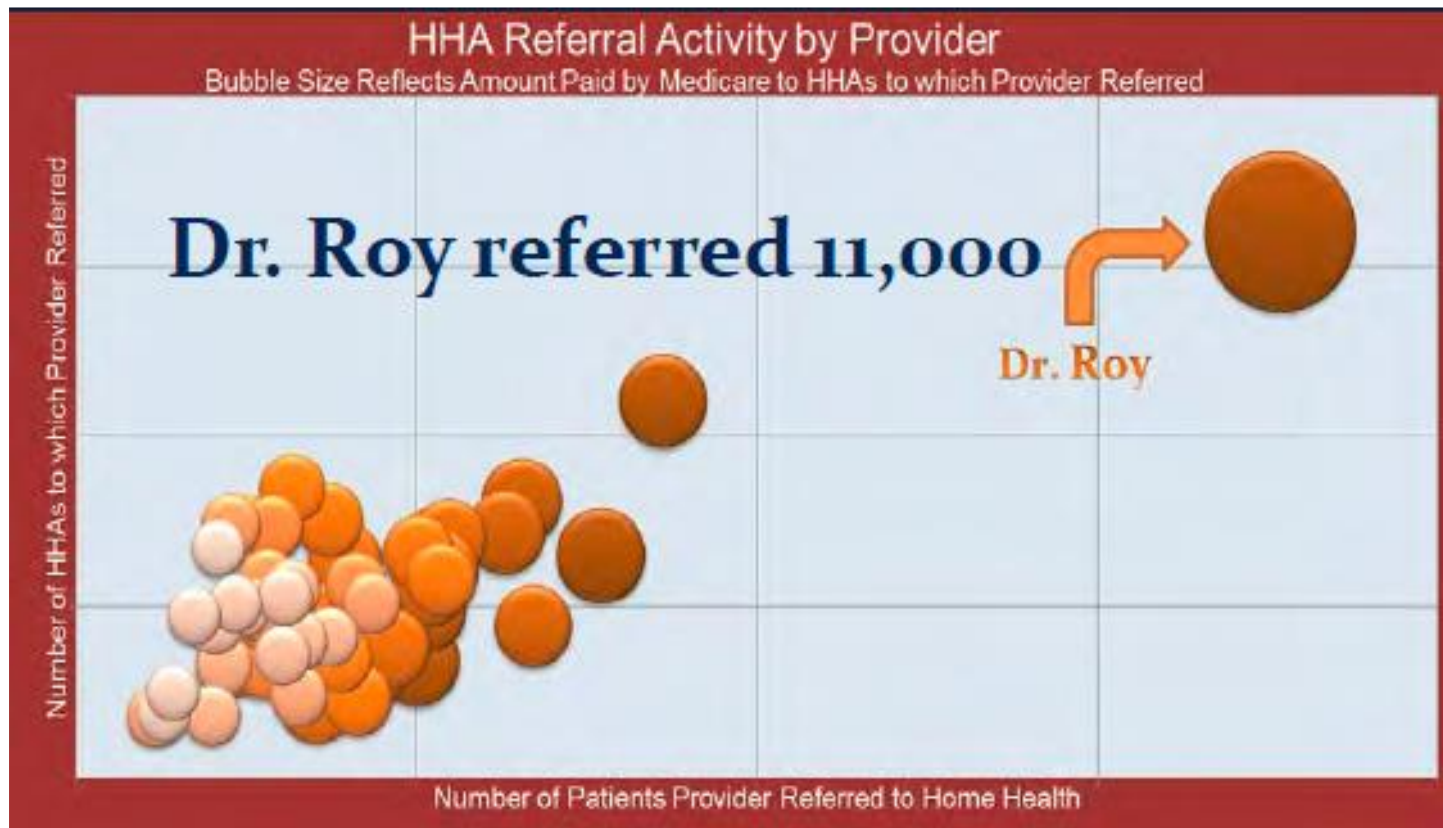
Boston, MA

New Orleans, LA

US Department of Health and Human Services, Office of the Inspector General

- CDO role:
 - Peer with other department leads
 - Leads strategic planning
 - Viewed as an enabler of agency success
- Achieved 2 \$1 billion takedowns of fraud
- \$5 returned for every \$1 of cost

Fraud detection: mapping outliers



Dr. Roy sentenced to 35 years, \$268 M restitution

- Doctors fraudulently recruited people on Medicare to sign up for home health care services.
 - Recruiters went to a homeless shelter and brought people to a car parked outside, and were paid \$50 per recruit.
 - Nurses falsified medical documents to make it appear as though they qualified for home health care services.
 - Once certified for home health care services, doctors and nurses falsified visit notes to make it appear as though nursing services were being provided.
- The fraud covered 11,000 Medicare beneficiaries from more than 500 different home health agencies.

General Services Administration: Enabler of others success

- Establishes platforms and tools
- Provides training
- Creates community of practice
- "Data science is a team sport"

Boston, a data-driven city



Boston's data team, by the numbers

- 28 staff (9 new last year)
- 60 analytics projects
 - Citywide: Performance, CityScore, Vision Zero, Boston 2030
 - Data warehouse to launch soon
 - Department-driven projects
- 13 city departments
- 5 summer fellows
- 626,387 page views of analytics site

Boston sample data projects



- Integrating **7 data sets** on hazards in buildings and road closures, **dashboard and map** gets firefighters to the scene safely
- **Saves \$5 million** and 20,000 pounds of carbon emissions **with an algorithm** for smart school bus routing.
- Energy monitoring saves \$1 million per year, including \$40K in central library

New Orleans



- **Blightstat** project
 - Reduced backlog by 18,000
 - Improved efficiency by 30%
- New Orleans **gives out free smoke alarms** where risk is high
 - Project has saved lives already
 - Uses housing data, vulnerability data to prioritize
- **Staffing model**
 - Tied to performance management
 - Integrated to other city staff

Thank You!

@JaneMWiseman



<https://scholar.harvard.edu/janewiseman>