Canaries in the Coal Mine: Understanding and Acting on Data in Context

Henry Amoroso, JD
Seton Hall University, Associate Professor of Law and Economics
HJA Strategies, Principal
Lackawanna Luzerne Indicators Regional Forum
Thursday, May 22nd, 2014

What we do

Henry Amoroso, Seton Hall/HJA Strategies

The Data "Revolution"

- State and local governments were projected to spend $58 billion on information technology in the 2013 budget year
- There's almost no more fashionable word in the policy world than data
- But what is data anyway…?

What is “Data” Anyway?

My Main Goals:

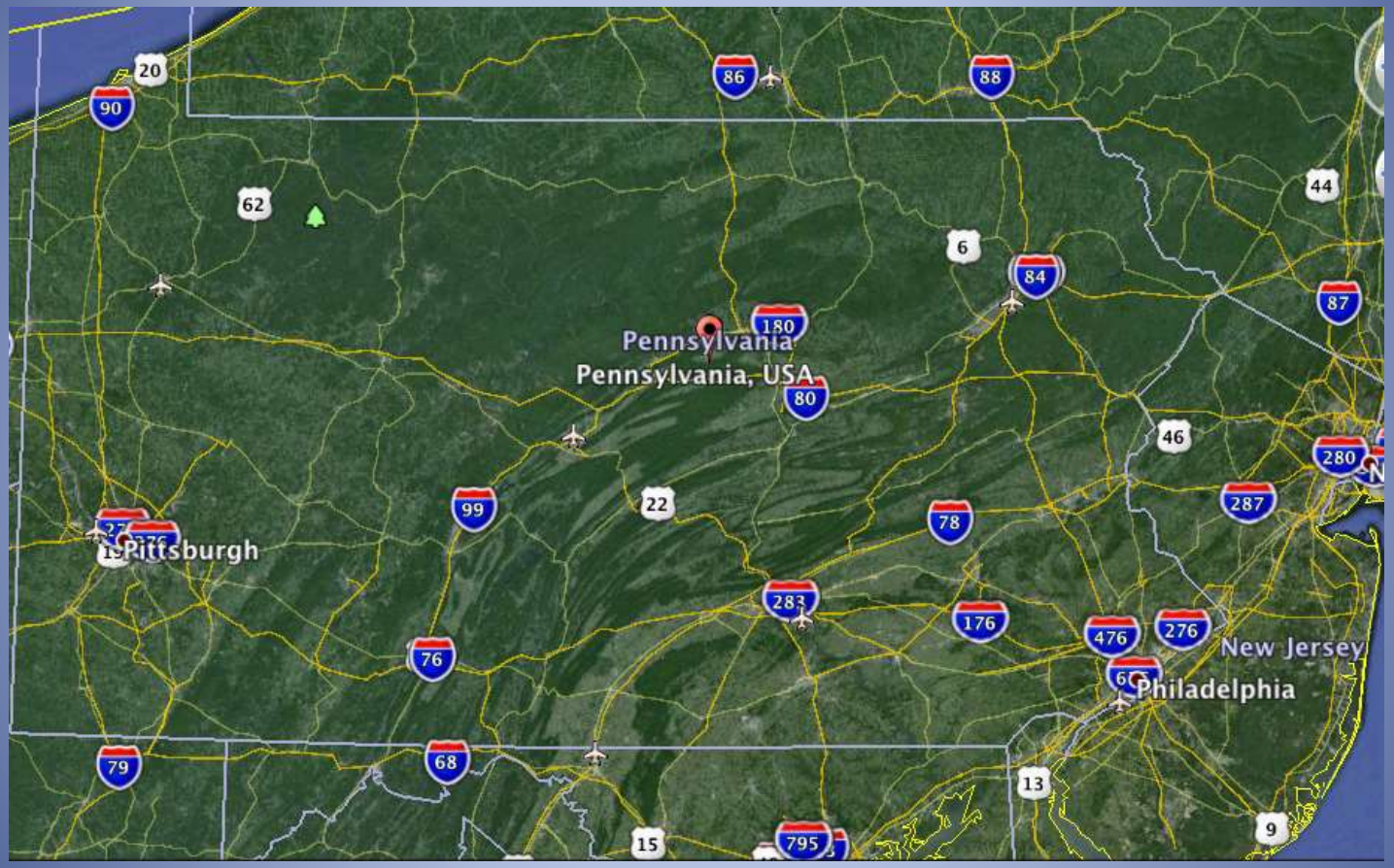
- Explain that the true definition of data is a lot simpler, familiar, and more accessible than it appears
- Recalibrate our idea of data and its role in government
- Underline the importance of *executing* on vital indicators

Five Rules:

1. Data-oriented governing is not new
2. Data does not and cannot do anything by itself
3. All forms of data are always situated in a particular context
4. Data does not exist unless we *know* data exists
5. Data must be taught to speak—*contextualized!*

1. Data Oriented Governing is Not New

1. Data Oriented Governing is Not New



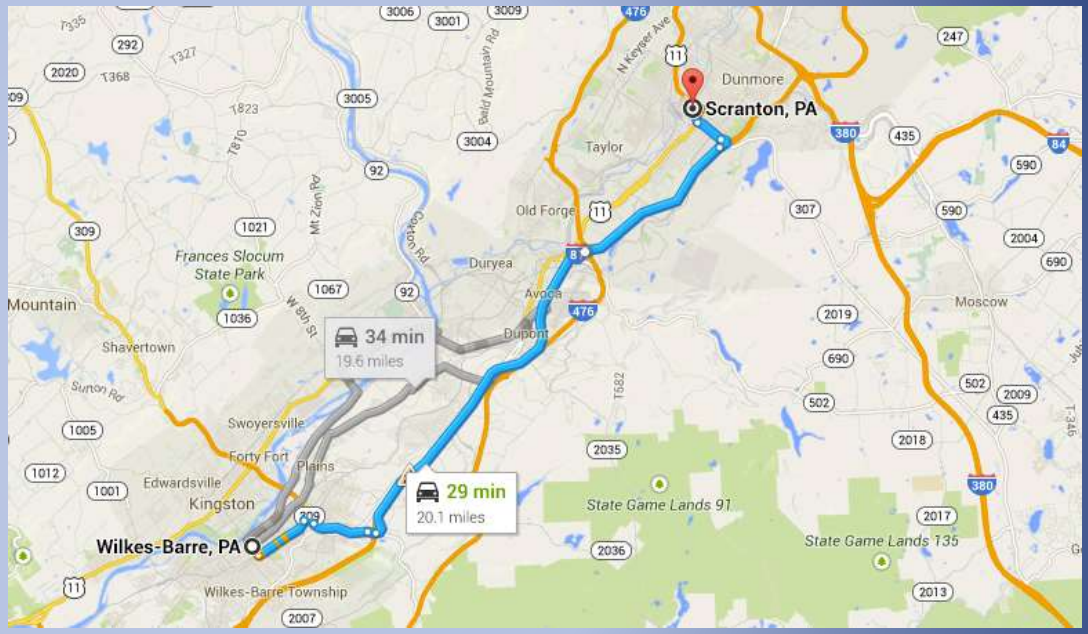
That's better.

1. Data Oriented Governing is Not New

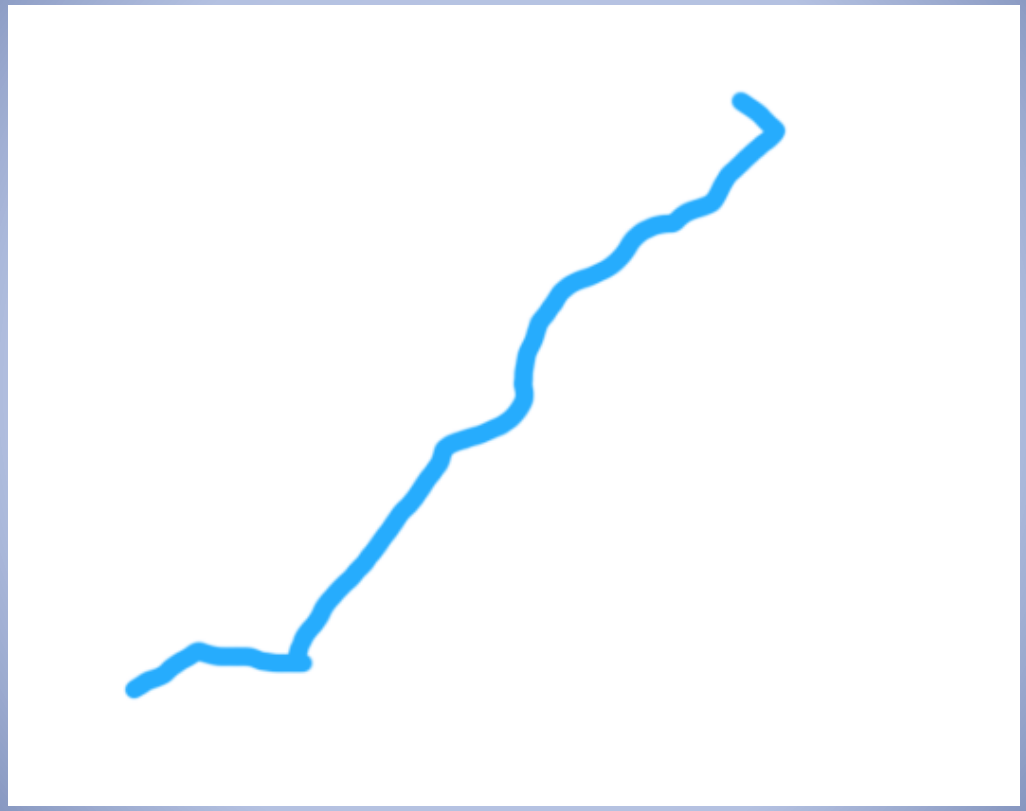
- We often take familiar forms of data delivery for granted. Think about how much information a map contains.
  - Imagine trying to write down all of this information as precisely as a map: “Approximately 2,345 feet from Wilkes-Barre, I-81 curves 4 inches to the northeast. 2,346 feet from Wilkes-Barre, I-81 curves 7.2 inches…” That’s a big book!

1. Data Oriented Governing is Not New

- Technology offers us easier and more efficient ways to use the information we have
- Google Maps, for example, makes using a map easier and *enhances* its information value
- But what happens when we separate "new" data delivery systems from the "old"?



1. Data Oriented Governing is Not New



…You still need the map.

1. Data Oriented Governing is Not New

- In government, the “maps” are documents like the budget, strategic plans, capital plans, and various departmental reports
- Technology can *supplement* these resources and the delivery of services
- But we still depend on documents like the budget to form the *foundation* of good government work
- Other documents, such as the Audited Financial Statement, measure if our map was viable and how we should adjust our efforts going forward

1. Data Oriented Governing is Not New

- Best Practice: “Old” and “New” Data Hybrids
  - Cities such as Jersey City, NJ and Oakland, CA have introduced dynamic digital budgets, which make using vital budget data far easier. These budgets are also widely accessible, available across government departments, and to the public

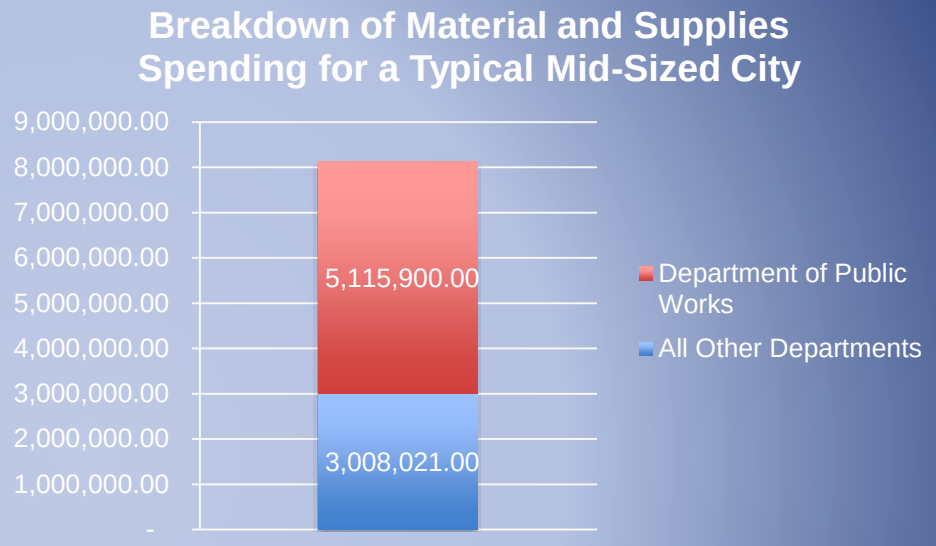
2. Data does not and cannot do anything

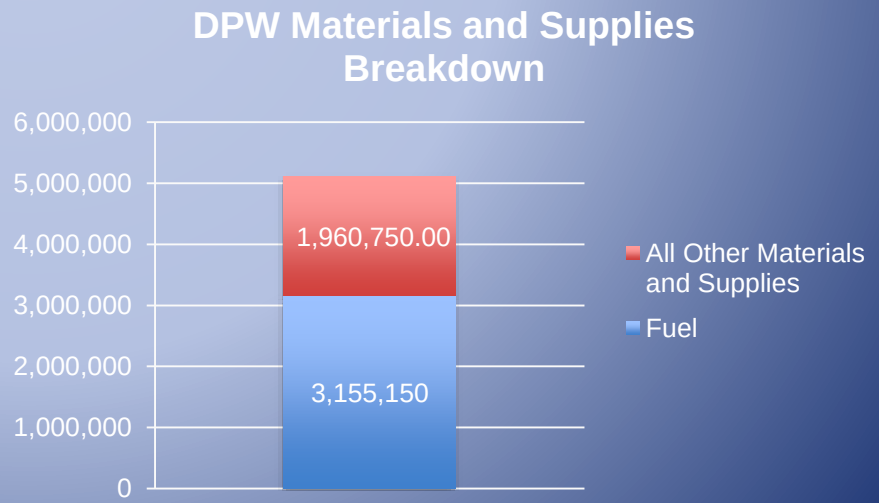
- Most cities typically spend significantly more on personnel than materials and supplies; in this example, nearly 28 times more

- But materials and supplies appears highly concentrated in most cities—usually in the Department of Public Works or equivalent

- That department, in turn, usually spends more than half of its materials and supply budget on fuels.

- Therefore, cities have a very strong incentive to control fuel spend





2. Data does not and cannot do anything

- Enter data!
  - Some options: fuel card storage, GPS tracking system, monitor who owns the car/who it's assigned to, post it on an internal CityStat platform—lots of options
- One city instituted a gas card system to monitor fuel usage by vehicle and prevent abuse (e.g., fueling a personal vehicle)
- When an employee fueled up, he or she would have to enter the odometer reading into the pump for fuel to begin dispensing
- The city was able to keep track of which employees and vehicles seemed to be using too much fuel, promising significant savings

2. Data does not and cannot do anything

- But then one vehicle's odometer broke and a work around needed to be developed for that vehicle
- Then another odometer failed and a work around needed to be developed for *that* vehicle
- Then another
- Then another…
- Eventually, the city had to scrap the odometer aspect of the fuel card system—employees simply entered "00000" to get fuel
- The investment in tracking fuel failed and fuel costs actually *increased*

2. Data does not and cannot do anything

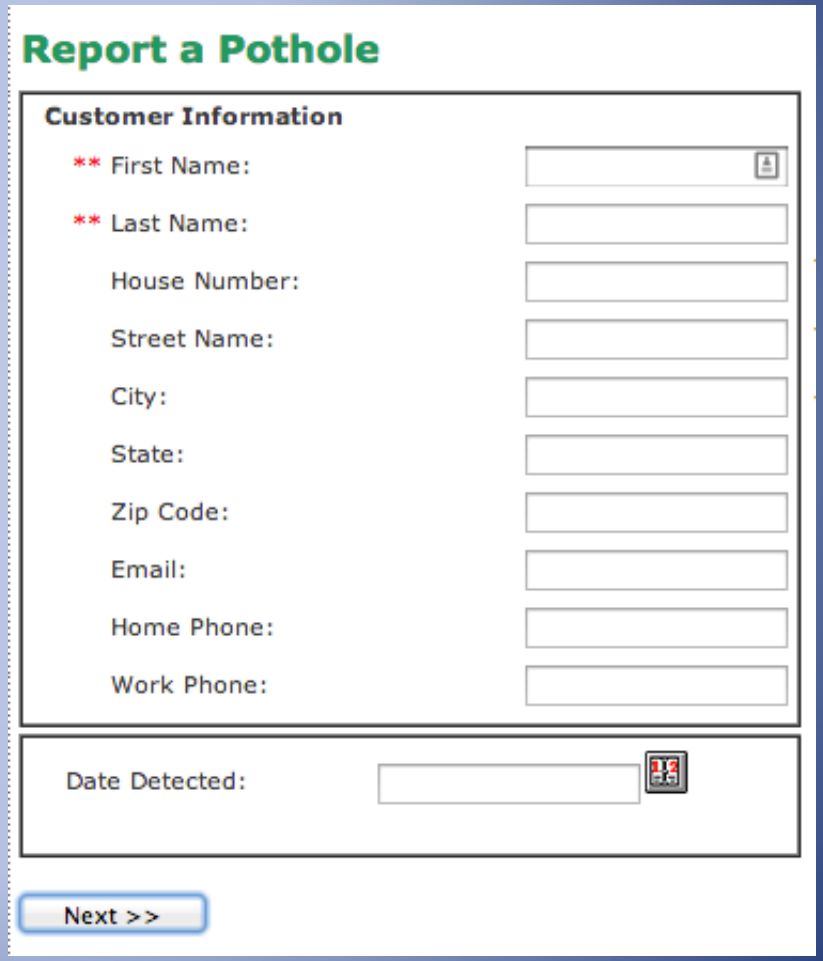
- This—and other data tracking measures—may sometimes be an expensive solution to what is a far more expensive and systemic *management problem*
  - Data might seem like a cheap solution, but it's not. Technology is expensive and usually requires capital funds that already have significant pressures (laying new roads, filling potholes, buying road salt, building a new Fire Department "superstation," etc
- Usually, a *comprehensive* technology solution is necessary

- **You may have a lot of data but you need to find a way to *execute on it!***

2. Data does not and cannot do anything

Best Practice: Data Execution

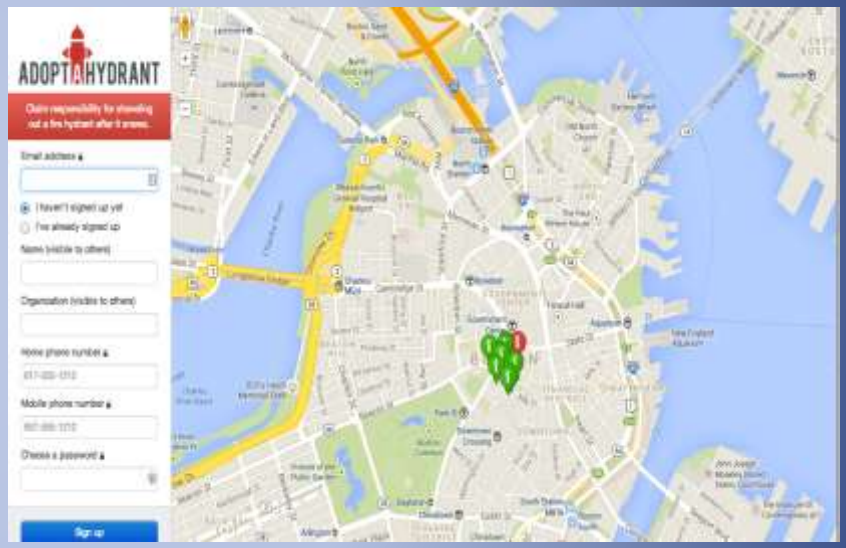
- New York City introduced a data-driven pothole registry and reporting program
- The time between receipt and repair was carefully monitored
- 1.25 *million* potholes were filled between 2002 and 2007
- 99% were filled in 30 days or less (vs. 65% before the institution of the registry and 311 reporting)



2. Data does not and cannot do anything

Best Practice: Execution Outsourcing

- The City of Boston has over 13,000 fire hydrants
- Snowy winters and older infrastructure make keeping fire hydrants accessible a priority
- Through a digital mapping tool, City residents can choose to “Adopt-A-Hydrant,” maintaining access to the hydrant and reporting when more serious issues arise
- This saves the city money and, more importantly, allows the Fire Department to spend more time on fighting fires and fire prevention
- Thorny political issues, such as private ownership or outsourcing the management of public assets, can also be avoided, thus making execution easier



3. All forms of data are always situated in a particular context

- Titusville, PA Population:
  - 1860: 438
  - 1870: 8,639 (up 1,872.4%)
- What changed?
  - Discovery of oil
- Detroit, MI: Population
  - 1950: 1.85M
  - 2010: 714,000 (down 61%)
- What changed
  - Decline of automobile manufacturing base

3. All forms of data are always situated in a particular context

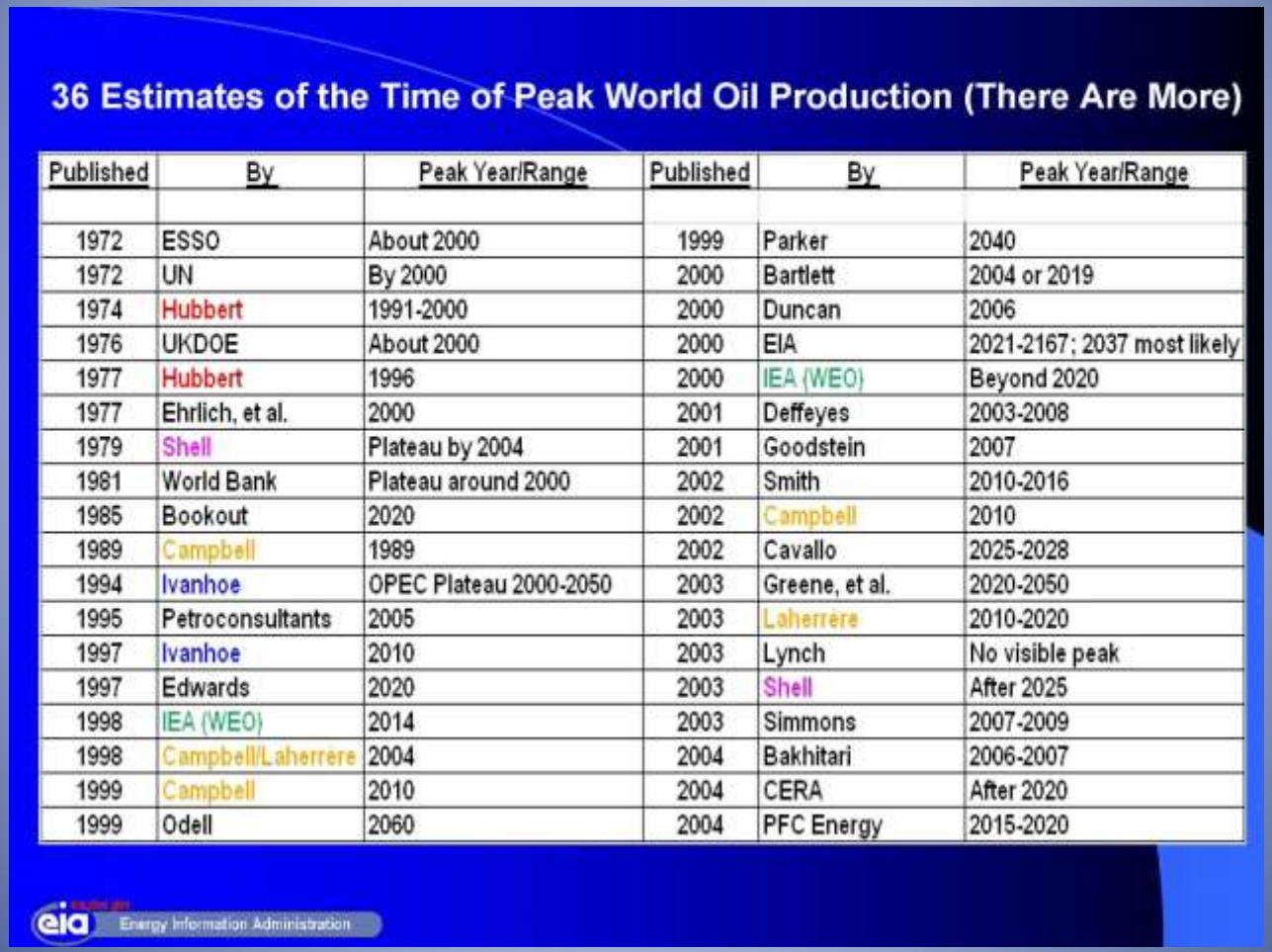
- But change came again: “In 1875, Henry E. Wrigley, the head of the Pennsylvania Geological Survey, issued a doomsday warning that the state—and hence the world—production of oil had peaked and would soon experience a precipitous decline, aggravating fears that had overshadowed the oil industry since its inception. ”

- Today, this sounds absurd

- According to the data available at the time—which necessarily limited itself to known oil fields (Northwestern Pennsylvania), known extraction methods, and uses (kerosene for light)—this prediction was correct.

- Excerpt From: Ron Chernow. “Titan.” 1998. Vintage Books

3. All forms of data are always situated in a particular context



36 Estimates of the Time of Peak World Oil Production (There Are More)

Published	By	Peak Year/Range	Published	By	Peak Year/Range
1972	ESSO	About 2000	1999	Parker	2040
1972	UN	By 2000	2000	Bartlett	2004 or 2019
1974	Hubbert	1991-2000	2000	Duncan	2006
1976	UKDOE	About 2000	2000	EIA	2021-2167; 2037 most likely
1977	Hubbert	1996	2000	IEA (WEO)	Beyond 2020
1977	Ehrlich, et al.	2000	2001	Deffeyes	2003-2008
1979	Shell	Plateau by 2004	2001	Goodstein	2007
1981	World Bank	Plateau around 2000	2002	Smith	2010-2016
1985	Bookout	2020	2002	Campbell	2010
1989	Campbell	1989	2002	Cavallo	2025-2028
1994	Ivanhoe	OPEC Plateau 2000-2050	2003	Greene, et al.	2020-2050
1995	Petroconsultants	2005	2003	Laherrère	2010-2020
1997	Ivanhoe	2010	2003	Lynch	No visible peak
1997	Edwards	2020	2003	Shell	After 2025
1998	IEA (WEO)	2014	2003	Simmons	2007-2009
1998	Campbell/Laherrère	2004	2004	Bakhitari	2006-2007
1999	Campbell	2010	2004	CERA	After 2020
1999	Odell	2060	2004	PFC Energy	2015-2020

eia Energy Information Administration

3. All forms of data are always situated in a particular context

- When times were good, the City of Detroit tried to mirror the grandeur of the auto industry, granting substantial pensions and benefits, as well as embarking on large infrastructure projects
- However, as the auto industry—and thus the City's financial base—declined, the City's liabilities remained in place
- This represents an essential failure to take key indicators into account and read "the data on the wall"
- It also reveals some of the general limitations governments have with respect to execution

3. All forms of data are always situated in a particular context

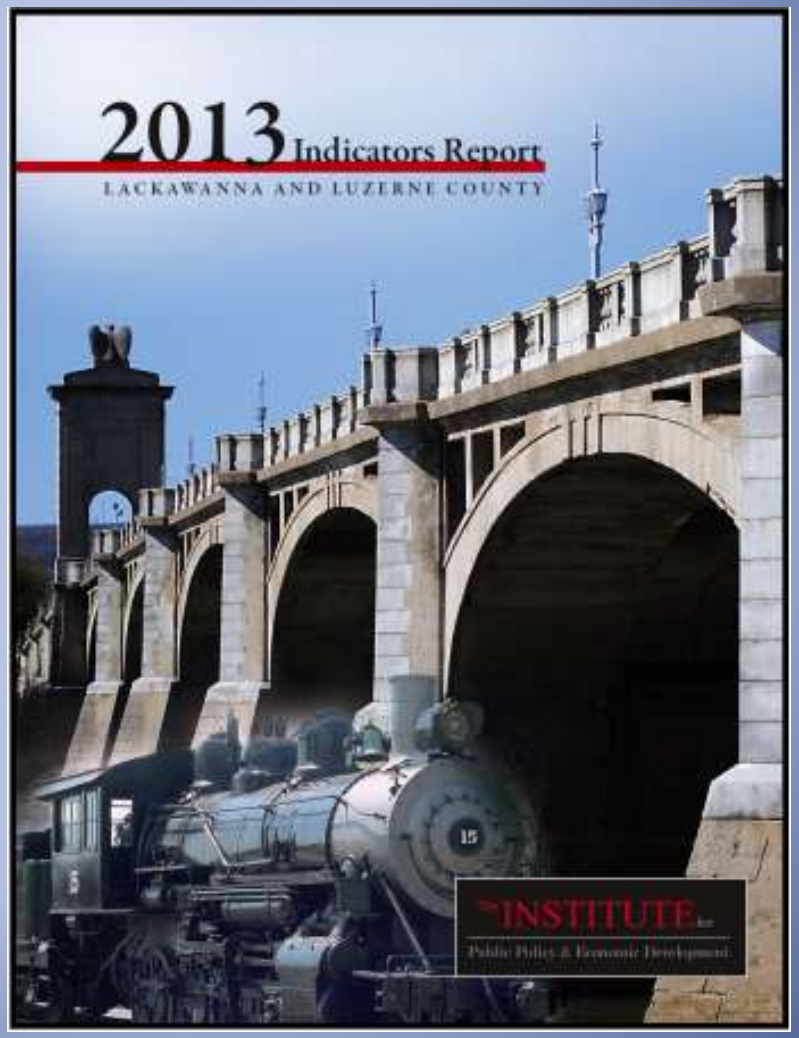
- Think about how much has changed since 1875
- Think about how much has changed since *2005*
- There is no way to predict the future, but there are ways to *model* the future using "scenarios"

3. All forms of data are always situated in a particular context

- Best Practice: Scenarios
- Royal Dutch Shell brought in experts to help construct possible scenarios for the oil industry—e.g., new technology (deepwater drilling, hydraulic fracturing) and political issues (carbon taxes, cap and trade systems, disruptions in the supply line, nationalization, cartel formation)
- We need more detailed "scenarios" in government planning
  - Austin, TX, has an entire Department devoted to both planning and *strategic* planning initiatives
  - These plans offer a great chance for public participation, as well as academics and other experts to weigh in
  - True inclusion is key for these efforts to be successful

3. All forms of data are always situated in a particular context

- Best Practice:



4. Data does not exist unless we know data exists

This is what raw data looks like:

Image Credit: Whe*re's Waldo*

Without organization and signposting, data is frustrating or, worse, useless

Henry Amoroso, Seton Hall/HJA Strategies

4. Data does not exist unless we know data exists

Cities must take a more proactive approach with data

Image Credit: Where's *Waldo*

This is who we're looking for

Henry Amoroso, Seton Hall/HJA Strategies

4. Data does not exist unless we know data exists

- Too often, one department does not know what the other department has
  - The Finance Director does not have what the Business Administrator has does not have what the Office of Management and Budget Director has
- Procurement data offers an excellent example
- In short, data *needs* to be easily accessible *to all*.
- It should not take months for a government employee to find data in their own government!

4. Data does not exist unless we know data exists

- Best Practice: Sharing data
  - One primary use of data is *responsive*—e.g., there's a pothole in front of Ms. Doe's house on Maple St.
  - Another primary use is *predictive*—e.g., Maple St. is more likely to develop potholes than Spruce St.
    - Boston, Chicago, Los Angeles, New York City, Philadelphia, San Francisco and Seattle (the G7 Initiative) have joined in an effort to share the data they collect and increase their predictive capabilities
    - But small and mid-sized cities that have excellent data programs—e.g., Dubuque, IA—have small sample sizes and limited financial capacity
    - To help combat this problem, the City of Boston has made its "Citizen Connect" app available to 54 other communities around Massachusetts, which allows citizens to report issues via their Smartphone

4. Data does not exist unless we know data exists

- Best Practice: *Making* people share data the old fashioned way
  - Mayor Rahm Emmanuel of Chicago issued an Executive Order "mandating that city agencies publish public data sets under their control as well as update them on a regular basis."
  - Since then, Nashville, TN, Louisville, KY, the State of Connecticut, and Boston, MA have followed suit with similar executive orders

5. Data must be taught to speak

- In 2014, every day 2.5 *quintillion* (25,000,000,000,000,000,000) bytes of data are created, with 90% of the world's data, at that time, created since 2012 alone
- Government related operations are not an insignificant portion of that amount
  - The British government's tax collection arm alone holds 80 times more data than the British Library, the largest library by catalog in the world (and the UK's population is about a fifth of America's)
- But data is useless unless we can translate these huge data sets into actionable and executable items
- We need to make more maps!

5. Data must be taught to speak

- Best Practice: Chicago's "SmartData Platform" and "Data Dictionary"
  - The aggregate total of Chicago's data collection efforts is 7 million rows *per day*
  - But this data exists in dozens of separate databases across units of government
  - The SmartData platform allows users to quickly query the database for vital information
  - It also uses collected data to predict needs
    - E.g., this system has allowed the City to predict the location of rodent activity spikes 7 days in advance
  - The Data Dictionary provides an invaluable tool for the non-computer scientist or trained data analyst to understand and make use of complex data indicators by defining key terms. It further explains what data the City has, where it may be found, and what formats it is available in.

5. Data must be taught to speak

- Best Practice: Listening to "Old-Fashioned" Indicators
  - Tampa, FL's downtown used to primarily be a 9-5 community (there were only 600 residents in 2004)
  - Mayor Bob Buckhorn has made transforming downtown into a 24/7 community a priority, with five new residential towers under construction between 2014-2016
  - How does he measure the community's *true* growth (and not just real estate *sales*)?
  - He drives around and sees how many people are walking their dogs downtown!

Canaries

- The most important use of data, in itself, is to provoke reactions that lead to actions
  - Because of the structure of government and the common distribution of data, insiders are often the only ones given a chance to react
  - Even when the ‘data on the wall’ is clear to see, decision-makers can push action into the future
- That’s where we see the role of “canaries” come in.
  - A canary’s job is to have a keen grasp of the vital indicators *in their context* and be able to relay those indicators to *all* stakeholders.
  - They must interpret data and be able to explain their significance both inside and outside of the government apparatus
  - A canary *must* be equipped to inspire decision makers to move beyond reaction and urge them to take genuine action.

Listening to the Canaries

- The popularity of downtown living has grown impressively in the last decade, with Generation X'ers and Millenials accounting for the majority of demand
  - 71 percent of the market for new dwelling units in downtown Wichita, KS is comprised by young people
  - Even downtown Detroit's population of college educated under-35s grew by 59% in the last census period—in contrast to a 25% decline city-wide
- Harnessing this migration is absolutely key and that means capitalizing on the existing infrastructure of older cities—e.g., walkability, historic architecture, anchor institutions, and transportation capacity

Conclusion

- Don’t forget about “old data” (the budget!)
- Plan true execution strategies
- Pay attention to the indicators and create scenarios!
- Be transparent
- Be clear and accessible

Thank You