

Open Data Initiatives: what works, what doesn't, and what you can expect

—

Lessons from the National Transportation Data Challenge
and other examples

Who am I?

- 20+ years in tech and project management
- No government experience
- Limited academic experience
- Lots of data publishing/analytics experience
- I speak Python in my spare time
- Most recently focused on data
 - interest in geo-spatial
- Came on as volunteer project manager part-way through the National Transportation Data Challenge



National Transportation Data Challenge



“Traffic deaths across the United States increased 14% over the last two years. In 2016, nearly 6,000 pedestrians died – the highest number in more than two decades. Data science collaboration has the potential to reverse this trend – this Challenge aims to do just that.”

“The National Transportation Data Challenge is a series of community problem-solving events, roundtables, hackathons, demonstrations, and tutorials/trainings to build and strengthen collaborative data science projects that advance transportation safety.”

A strong team of backers

- Launched May 2017
- Sponsored by 4 Regional Big Data Innovation Hubs (NSF program)
- Support from federal and state agencies
- Sponsorship from Google, AWS, Satori, Esri and others.



On paper, looks great. But...

- In any open data initiative, a number of decisions must be made that will influence success. We got some of these right, some wrong.
 - Organization
 - Topic
 - Goals
 - Structure/Plan
 - Participation
- How do these factors impact the success of an open data initiative?

But first, some background

“Open Data Initiatives” come in two flavors:

- Data Publishing
 - Collect, aggregate, clean and publish datasets
 - Often done in phases, starting with core data, adding additional layers
 - *Must have an ongoing commitment!*
- Data Use
 - Use/exploit data
 - Often involve outside contributors
 - May be time-bound or open-ended
- These two feed on each other

Data Publishing

311 Service Requests - Dataset

← → ↺ 🏠

🔒 https://data.boston.gov/dataset/311-service-requests

⋮ 📧 ☆


🔍 📄 ☰

☰ MENU

ANALYZE BOSTON

DATASETS NEWS TIPS LOG IN SIGN UP

Home > Organizations > Boston 311 > 311 Service Requests



311 SERVICE REQUESTS

Followers

3

ORGANIZATION

BOS:311

BOSTON 311

311 is an easy-to-remember telephone number that connects you with highly-trained Constituent Service Center representatives who are ready to help you with requests for... read more

Dataset

🔖 Topics

🕒 Activity Stream


🖼 Showcases


311 SERVICE REQUESTS

This data set includes all channels of engagement in which a service request is created.

Refer to this link to learn more about **BOS:311**: <https://www.cityofboston.gov/311/>

DATA AND RESOURCES

 311 Service Requests 🔥
Updated daily

 CRM Value Codex 🔥
Helps in interpreting actual values given in columns

📊 Preview

📄 DOWNLOAD

📊 Preview

📄 DOWNLOAD

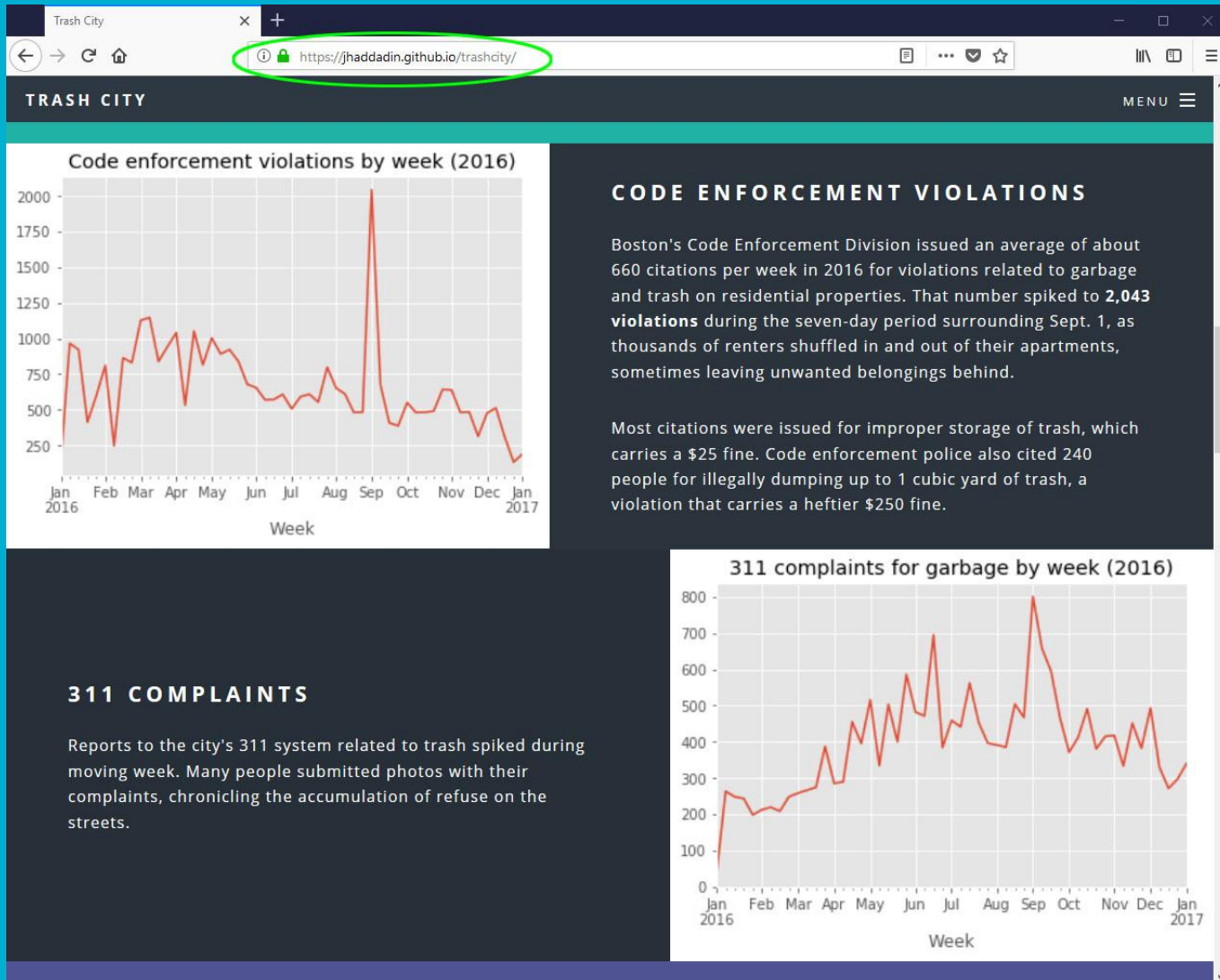
TAGS

311 CRM Case management city services

ADDITIONAL INFO

TITLE	311 Service Requests
TYPE	Tabular
DESCRIPTION	<p>This data set includes all channels of engagement in which a service request is created.</p> <p>Refer to this link to learn more about BOS:311: https://www.cityofboston.gov/311/</p>

Data Use



Hybrid? (Not really!)

The screenshot shows the Analyze Boston website in a web browser. The browser's address bar displays <https://data.boston.gov>. The website's header includes a menu icon, the "ANALYZE BOSTON" logo, and navigation links for "DATASETS", "NEWS", "TIPS", "LOG IN", and "SIGN UP".

The main banner features a scenic image of a boat on a lake in Boston. The text on the banner reads: "Welcome to ANALYZE BOSTON. Analyze Boston is the City of Boston's open data hub. We invite you to explore our *datasets*, read *about us*, or see our *tips for users*." Below this text is a search bar with the placeholder text "Search from 134 Datasets" and a red magnifying glass icon.

Below the banner is a section titled "SHOWCASES" with the subtitle "See what our users are doing with open data." This section displays three data visualizations:

- Anime Boston 2014**: A map showing the locations of anime-related events in Boston, with a text box indicating the event was held on March 21-22, 2014, at the Boston Convention Center.
- TRASH CITY**: A photograph of a cluttered room, with a text box asking "HOW DOES MOVING WASTE IMPACT THE QUALITY OF LIFE IN BOSTON?"
- Category & incidence totals**: A bar chart showing the frequency of various categories of incidents. The chart is divided into two sections: "Owner absent" and "Owner occupied". The "Owner absent" section shows a high frequency of incidents, while the "Owner occupied" section shows a lower frequency. The chart includes a legend for "Frequency" and "Category".

But back to the Transportation Challenge...

- A data-use initiative
 - But that didn't stop some people
- Somewhat hamstrung by organizational constraints
- Sponsor involvement/limitations were also a challenge

“Parent Organization”

“NSF's Directorate for Computer and Information Science and Engineering (CISE) initiated the National Network of Big Data Regional Innovation Hubs (BD Hubs) program in FY 2015. Four BD Hubs – Midwest, Northeast, South, and West – were established to foster multi-sector collaborations among academia, industry, and government, both nationally and internationally. These BD Hubs are serving a convening and coordinating role by bringing together a wide range of Big Data stakeholders in order to connect solution seekers with solution providers.”

Organization

- Extraordinarily complex for an initiative with virtually no dedicated staff and limited budget
 - Four regions
 - Six areas of focus
 - Multiple sponsors
 - Desire for all the above to cross-pollinate
 - Part time contributors
- *But...* Nobody representing interested populations
 - Cyclist groups
 - Pedestrian groups
 - Transit groups
 - Auto manufacturers
 - *This was one of my biggest personal fails!*

Organizational issues

- Not designed to move quickly
 - The real world is moving faster than they are
- Focus areas and geographies sometimes overlapped, but not always
 - A painful matrix
 - It would have worked better if this had been more tightly aligned
- Need to avoid an apparent commercial conflicts of interest was limiting

Topic?



“Traffic deaths across the United States increased 14% over the last two years. In 2016, nearly 6,000 pedestrians died – the highest number in more than two decades. Data science collaboration has the potential to reverse this trend – this Challenge aims to do just that.”

“The National Transportation Data Challenge is a series of community problem-solving events, roundtables, hackathons, demonstrations, and tutorials/trainings to build and strengthen collaborative data science projects that advance transportation safety.”

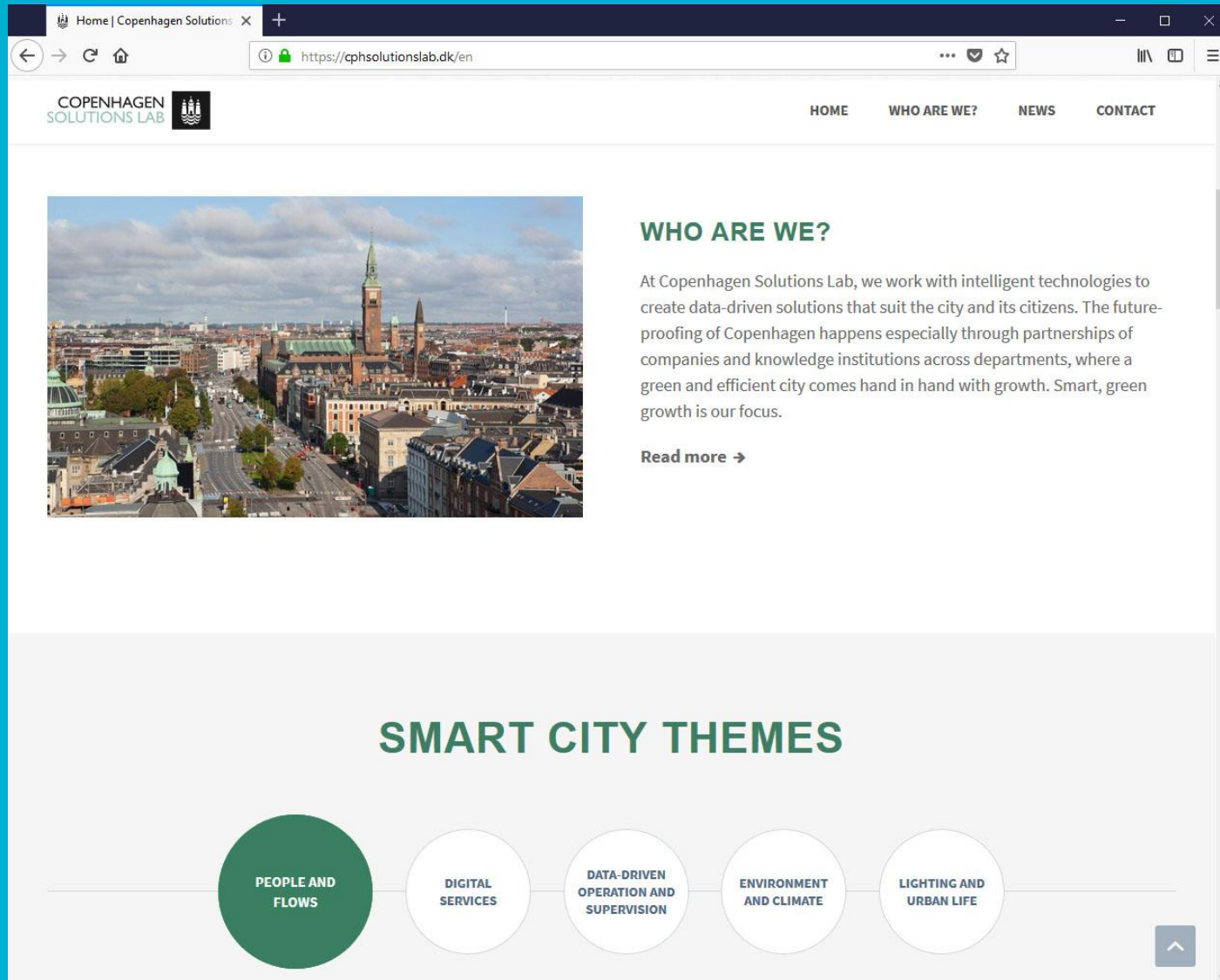
Topic issues

- Initiatives gain traction when the topic is timely, interesting and relevant
- Transportation fits those criteria
- But, *everybody* in the tech world is looking at some aspect of transportation! (Google, Facebook, Amazon, etc.)
- The challenge is to be unique rather than being another “me too” traffic or public transport effort

Topic: Results

- Mixed success. It's a very broad area. Probably too broad.
- Takeaway: For a short initiative, greater simpler focus would have helped
- Regional or particularly timely issues were most successful
 - But how tightly they really related to the topic is debatable.
- Successful example: Copenhagen Solutions Lab Themes

How does Copenhagen do it? Iteratively!



Goals: Too many and too vague

- *DISTRACTED DRIVING: How might we determine if there is a distracted driving problem? How might we mitigate the issue?*
- *WEATHER / EMERGENCY RESPONSE: How might we improve transportation safety in extreme weather? In cases of crises?*
- *CURRICULUM DEVELOPMENT: How might we develop a ready-to-go high school or undergraduate curriculum introducing transportation safety data to build awareness and spark new contributions/solutions?*
- *BIKE / PEDESTRIAN SAFETY: How might we improve bike/pedestrian safety in our communities?*
- *AUTONOMOUS OR CONNECTED VEHICLES: How might we leverage autonomous or enhanced vehicles to improve our constituents' safety? How might we navigate technological, policy, and social considerations involved?*
- *MULTIPLE DATA STREAMS: How might we utilize multiple data streams (particularly from different sectors, sources) to capture actionable insights to make a local transportation corridor more safe? Are there lessons and best practices that can be shared more broadly?*

Issues with our “goals”

- Too “academic” an approach
- “Challenge” was never defined
 - Who is being challenged?
 - To do what?
 - Judged by whom?
 - With what prizes/results/recognition?
- A topic is not a goal. A question is not a goal. Discussing an approach is not a goal.
- Every goal should have a “business owner,” those don’t really exist at a national level.
- We had too many possible goals -> Lack of focus

Initiative Structure and Plan

- We stated what we would do... sort of
 - Scheduled a number of events, discussions, phone calls etc.
- Had a plan
- Defined key elements
- Developed metrics

Planning: what worked, what didn't

- Lots of good events, but they were isolated from each other
- Too many parallel events that we could not learn from
 - In large part because of the parallel nature of the six focus areas
- Nobody owned cross-pollination of ideas
- Nobody owned key elements for the national challenge
- Key takeaway: One thing at a time, towards a shared goal
- Successful example: Open Data Kansas City, successfully started with just a single dataset (census) upon which much more content could be built.

How did Kansas City do it?

The screenshot displays the Open Data KC website interface. The browser address bar shows the URL <https://data.kcmo.org>. The website header includes the "OPENDATA KC" logo and navigation links: Home, Chatbot, KCStat, Budget, Maps, Suggestions, Help, and a Sign In button. A search bar is located below the header.

The main content area shows search results for 4697 items, sorted by "Most Relevant". The left sidebar contains filters for Authority (Official, Community), Categories (311, Airport, Area Plans, Audits, Budget), and View Types (Calendars, Charts, Data Lens pages, Datasets, External Datasets, Files and Documents).

The search results list the following datasets:

- Kansas City Monthly Car Auction** (COMMUNITY, Traffic)
 - Unclaimed Vehicle Auction
 - Updated: February 2, 2018
 - Views: 38,738
 - Tags: No tags assigned
- IB110 Part C Building Permit Checklist** (Information Bulletins)
 - Checklist of items necessary for performing a complete plans review of a project.
 - Updated: September 14, 2017
 - Views: 36,873
 - Tags: permits, building, ib 110
- Solid Waste Management guide** (Neighborhoods)
 - trash, leaf and brush, recycling, drop off, tires, neighborhood cleanup, habitat restore, swap shop
 - Updated: February 27, 2014
 - Views: 22,509
 - Tags: drop off, habitat restore, leaf and brush, neighborhood cleanup, recycling, and 3 more
- 2012 Chapter 18 Article 2: Building Code** (Regulatory Codes)
 - Current 2012 Building and Rehabilitation Code
 - Updated: September 24, 2014
 - Views: 16,458
 - Tags: building, codes, ordinances
- IB109 Email and Fax Permit Process** (Information Bulletins)

Initiative Participation

- If there's no outside participation, all you've done is publish a website, which will not last long
- Need participation at many levels
 - Short term contributors
 - Longer term contributors/volunteers/part time help
 - Sponsors/vendors/suppliers/event hosts
- Getting participation requires significant ongoing outreach and publicity

Initiative outreach



Participation: what we did

- Great sponsors
- Good-great longer term participants
- Struggled in to attract shorter-term contributors
 - In part because of the unclear nature of what the “challenge” was
- Key takeaway: Understand the constituencies and design around their needs and interests.
 - They will probably not build what you thought they would build
 - Example: NYC crime/police data used to demonstrate racial bias by police, not to help improve policing!
- Successful approach: NASA annual hackathons.

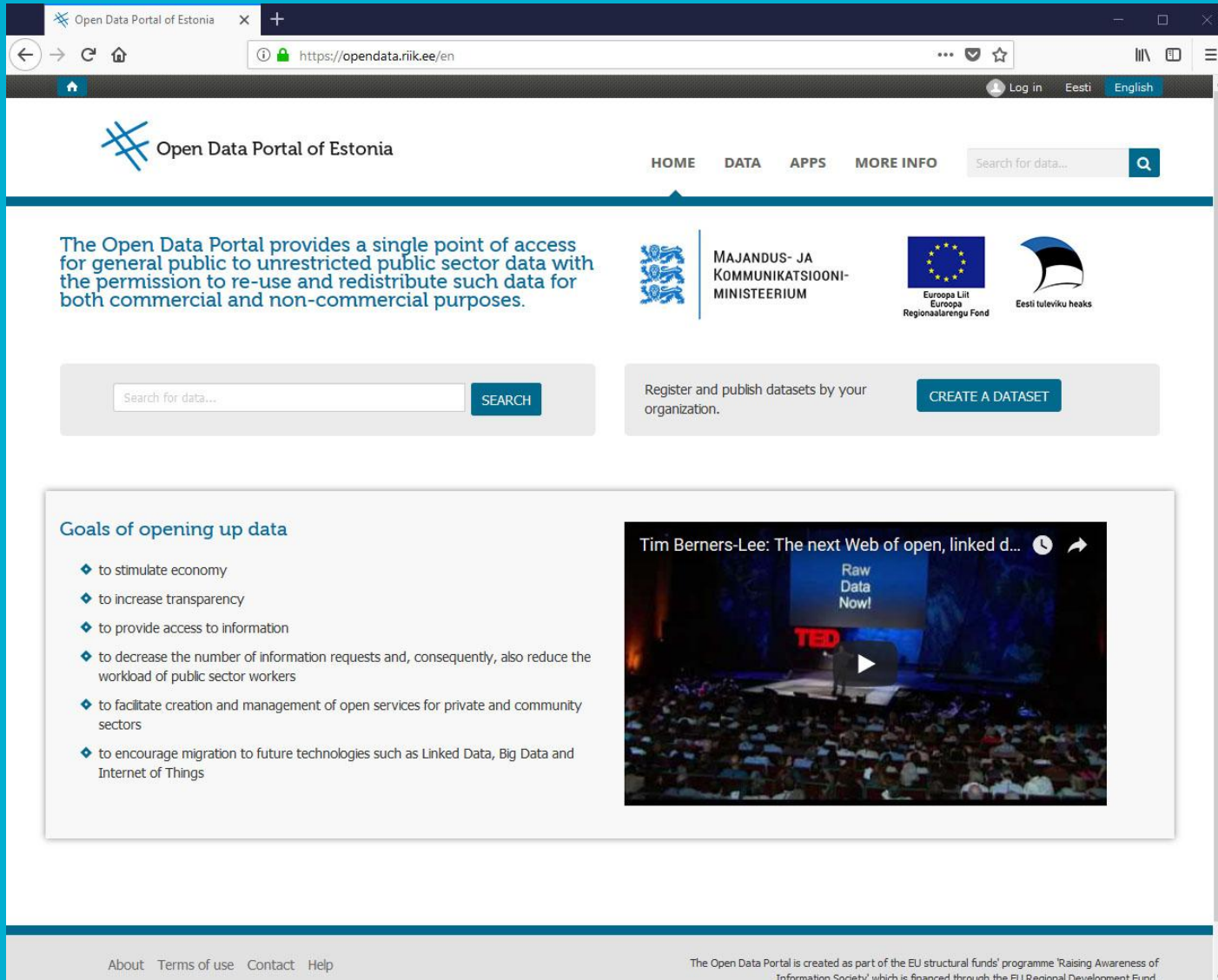
Final Thoughts

- You can't launch an initiative without a lot of thought
- You need to have the right organization: Big Data Hubs didn't fit a national public-private initiative
- There needs to be a lot of visible, public, focused activity
- There needs to be a regular feedback loop, so each event feeds into improvements to the next one.
- There need to be metrics to evaluate all this, especially during the initiative. The transportation challenge only measured success at the end.

Final thoughts: A great all around example

- Open Data Estonia
 - <https://opendata.riik.ee/en>
- Open Data New Zealand
 - <https://www.data.govt.nz/>
- Include:
 - Datasets
 - Toolkit
 - Case studies
 - Community events
 - Outreach
 - Social media presence
- Both long-term initiatives and short-term projects
 - Many good examples of both types

Open Data ES



The screenshot displays the Open Data Portal of Estonia website. The browser's address bar shows the URL <https://opendata.riik.ee/en>. The website's header includes the logo, navigation links (HOME, DATA, APPS, MORE INFO), and a search bar. The main content area features a descriptive paragraph about the portal's purpose, logos for the Estonian government and the European Union, and a section titled 'Goals of opening up data' with a list of objectives. A video player shows a TED talk by Tim Berners-Lee. The footer contains links for 'About', 'Terms of use', 'Contact', and 'Help', along with a note about the portal's funding.

Open Data Portal of Estonia

HOME DATA APPS MORE INFO Search for data...

The Open Data Portal provides a single point of access for general public to unrestricted public sector data with the permission to re-use and redistribute such data for both commercial and non-commercial purposes.

MAJANDUS- JA KOMMUNIKATSIOONI-MINISTEERIUM

Euroopa Liit Euroopa Regionaalarengu Fond

Eesti tuleviku heaks

Search for data... SEARCH

Register and publish datasets by your organization. CREATE A DATASET

Goals of opening up data

- to stimulate economy
- to increase transparency
- to provide access to information
- to decrease the number of information requests and, consequently, also reduce the workload of public sector workers
- to facilitate creation and management of open services for private and community sectors
- to encourage migration to future technologies such as Linked Data, Big Data and Internet of Things

Tim Berners-Lee: The next Web of open, linked d... Raw Data Now!

About Terms of use Contact Help

The Open Data Portal is created as part of the EU structural funds' programme 'Raising Awareness of Information Society' which is financed through the EU Regional Development Fund.

Open Data NZ

The screenshot shows the homepage of data.govt.nz. At the top, there's a navigation bar with links for Home, Datasets, Standards & guidance, Case studies, Community, and Blog. A search bar is prominently displayed in the center, with the placeholder text "E.g. environment". To the right of the search bar, there are three statistics: 5.4k datasets, 145 organisations, and 25 groups. Below the search bar, there are three featured articles. The first article is titled "Searching for more than datasets" and discusses a unified search concept. The second article is titled "Thanks for your input: October 2017 data.govt.nz community feedback results" and mentions a performance analyst review. The third article is titled "Creative Commons and NZGOAL: two of the essential ingredients of open data" and discusses the role of copyright in open data. The Creative Commons logo is also visible next to the third article.

Home - data.govt.nz

Request dataset Add dataset

data.govt.nz Discover and use data

Home Datasets Standards & guidance Case studies Community Blog

5.4k datasets 145 organisations 25 groups

Search data

E.g. environment

Searching for more than datasets

ENGAGEMENT

We tested two concepts with a number of our users from the data community to extend the search so it includes results from both datasets and content. Read about our process and provide your feedback.

By Cam Findlay on 20 Feb 2018

Thanks for your input: October 2017 data.govt.nz community feedback results

COMMUNITY

Our data and performance analyst reviews the results from the first data.govt.nz community feedback survey. Find out what users want and how we're planning to improve data.govt.nz.

By Bronwyn Beagle on 4 Dec 2017

creative commons
AOTEAROA NEW ZEALAND

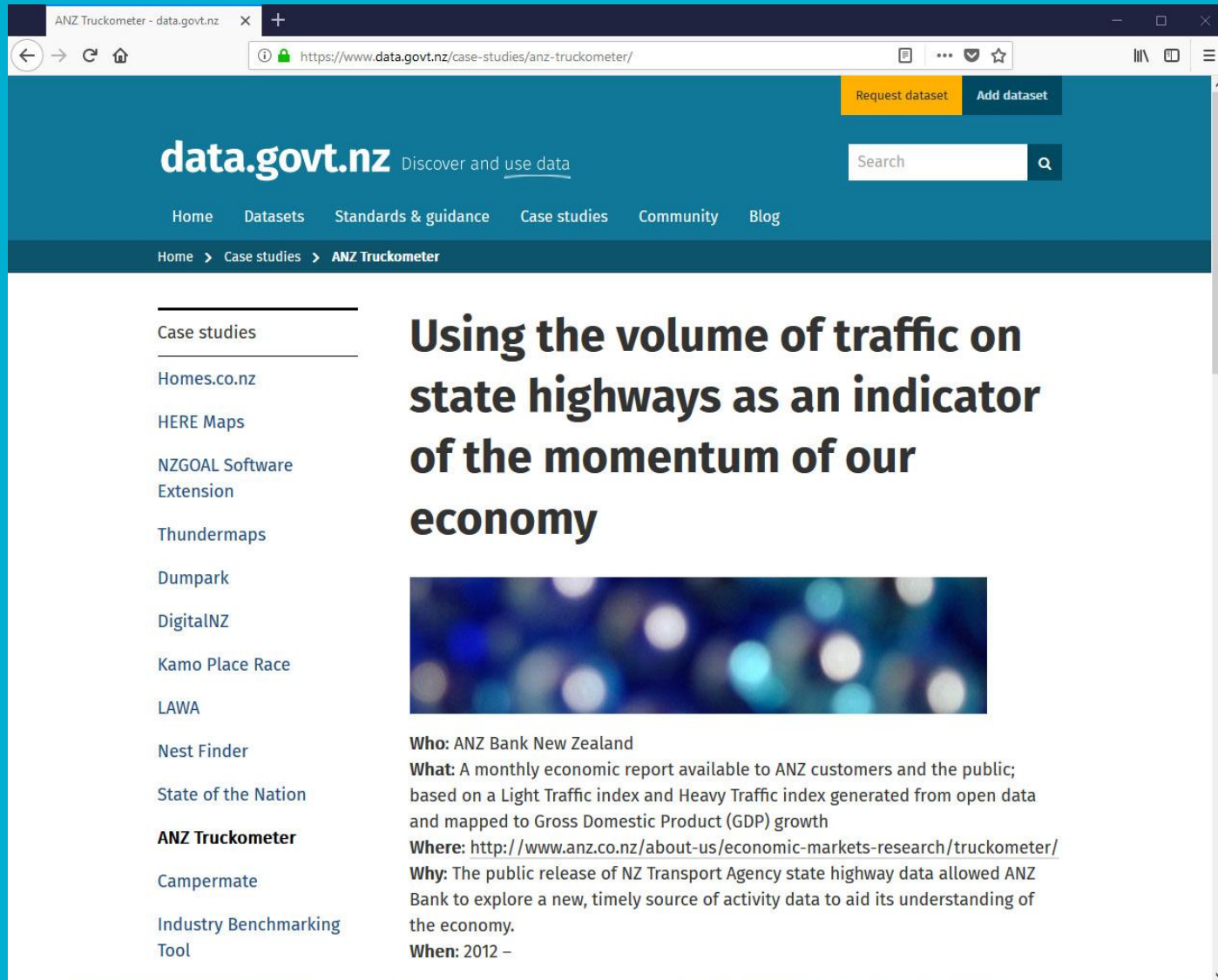
Creative Commons and NZGOAL: two of the essential ingredients of open data

EVENTS, GUIDANCE

Find out the role copyright plays in open data, and where training is available to learn how to apply open licensing in your work.

By Elizabeth Heritage on 24 Oct 2017

Open Data NZ case study: Trucking



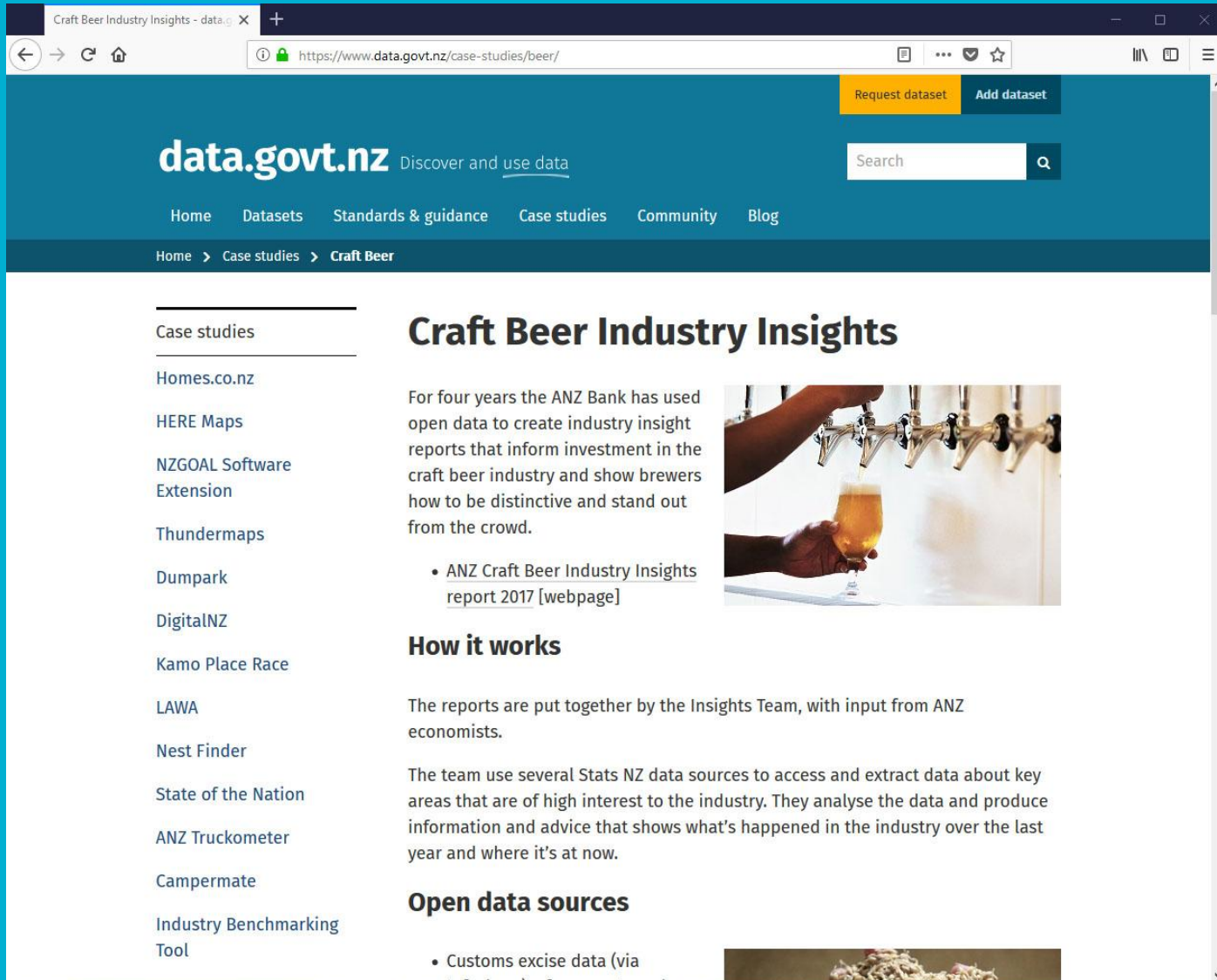
The screenshot shows a web browser window with the URL <https://www.data.govt.nz/case-studies/anz-truckometer/>. The page features a dark blue header with the **data.govt.nz** logo and the tagline "Discover and use data". Navigation links include Home, Datasets, Standards & guidance, Case studies, Community, and Blog. A search bar is also present. Below the header, a breadcrumb trail reads "Home > Case studies > ANZ Truckometer".

On the left side, there is a sidebar with a list of case studies under the heading "Case studies". The list includes: Homes.co.nz, HERE Maps, NZGOAL Software Extension, Thundermaps, Dumpark, DigitalNZ, Kamo Place Race, LAWA, Nest Finder, State of the Nation, **ANZ Truckometer** (highlighted), Campermate, and Industry Benchmarking Tool.

The main content area features the title "Using the volume of traffic on state highways as an indicator of the momentum of our economy" in large, bold, black text. Below the title is a blurred image of blue and white bokeh lights. To the right of the image, there is a list of details about the case study:

- Who:** ANZ Bank New Zealand
- What:** A monthly economic report available to ANZ customers and the public; based on a Light Traffic index and Heavy Traffic index generated from open data and mapped to Gross Domestic Product (GDP) growth
- Where:** <http://www.anz.co.nz/about-us/economic-markets-research/truckometer/>
- Why:** The public release of NZ Transport Agency state highway data allowed ANZ Bank to explore a new, timely source of activity data to aid its understanding of the economy.
- When:** 2012 –

Open Data NZ case study: Beer!



The screenshot shows a web browser window with the URL <https://www.data.govt.nz/case-studies/beer/>. The page features a dark blue header with the **data.govt.nz** logo and the tagline "Discover and use data". Navigation links include Home, Datasets, Standards & guidance, Case studies, Community, and Blog. A search bar is also present. Below the header, a breadcrumb trail reads "Home > Case studies > Craft Beer".

The main content area is titled "Craft Beer Industry Insights". It includes a paragraph: "For four years the ANZ Bank has used open data to create industry insight reports that inform investment in the craft beer industry and show brewers how to be distinctive and stand out from the crowd." Below this is a list item: "• ANZ Craft Beer Industry Insights report 2017 [webpage]". To the right of the text is an image of a hand pouring beer from a tap.

A sidebar on the left lists various case studies under the heading "Case studies": Homes.co.nz, HERE Maps, NZGOAL Software Extension, Thundermaps, Dumpark, DigitalNZ, Kamo Place Race, LAWA, Nest Finder, State of the Nation, ANZ Truckometer, Campermate, and Industry Benchmarking Tool.

Below the paragraph is a section titled "How it works" with the text: "The reports are put together by the Insights Team, with input from ANZ economists." and "The team use several Stats NZ data sources to access and extract data about key areas that are of high interest to the industry. They analyse the data and produce information and advice that shows what's happened in the industry over the last year and where it's at now."

At the bottom is a section titled "Open data sources" with a list item: "• Customs excise data (via". To the right of this list item is a small image of a pile of food.

More final thoughts: Open data in the US

- The US (Federal) government is not the optimal place to try to do open data
 - Nobody has a mandate
 - Opendata.gov run by GSA! Departments publish as they wish
 - Weaponization of data by virtually all political entities makes open, unfiltered data a difficult thing to do
 - No long term commitments
 - Difficult to involve outside and especially commercial parties
 - A professional civil service does help with things like this.
- States, Cities and private enterprise are where it's at
- Commercial interests are moving faster than most government entities can.

Final, final thoughts

- Enthusiasm was great
- Everybody really believed the topic mattered
- Everybody happily put in their own time
- We generated useful ideas and approaches that I believe will be exploited beyond the challenge
- We got good political exposure despite an administration that is somewhat hostile

Thanks, questions

SCaLE 16x

Dave Goodsmith / Datascience.com

Meredith Lee / West Big Data Hub

The Regional Big Data Innovation
Hubs

Flash the Cat

Michael Gat

@michaelgat

<http://www.michaelgat.com>

