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...

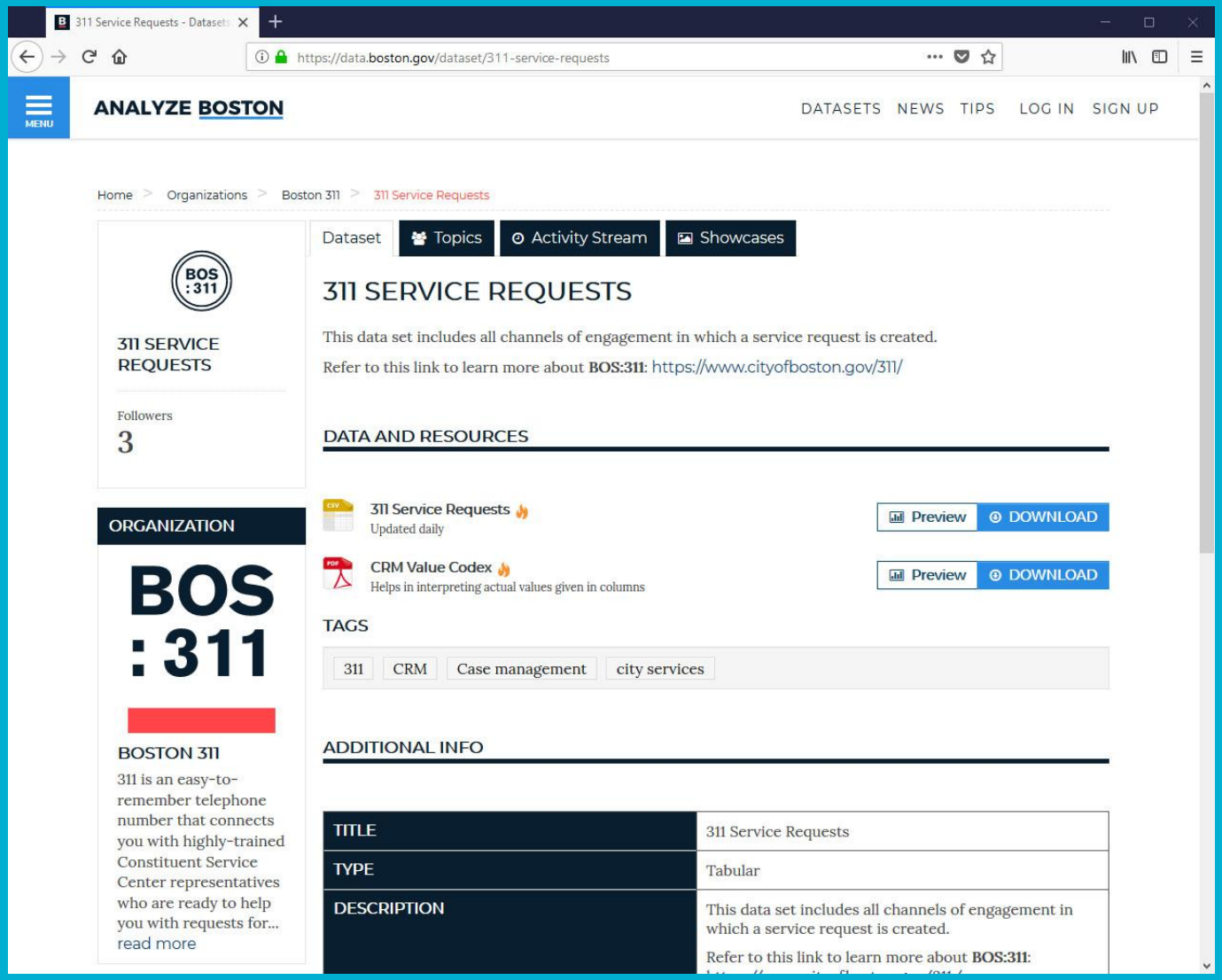
- In any open data initiative, a number of decisions must be made that will influence success. We got some of these right, some wrong.
  - Organization
  - Topic
  - Goals
  - Structure/Plan
  - Participation
- How do these factors impact the success of an open data initiative?

But first, some background

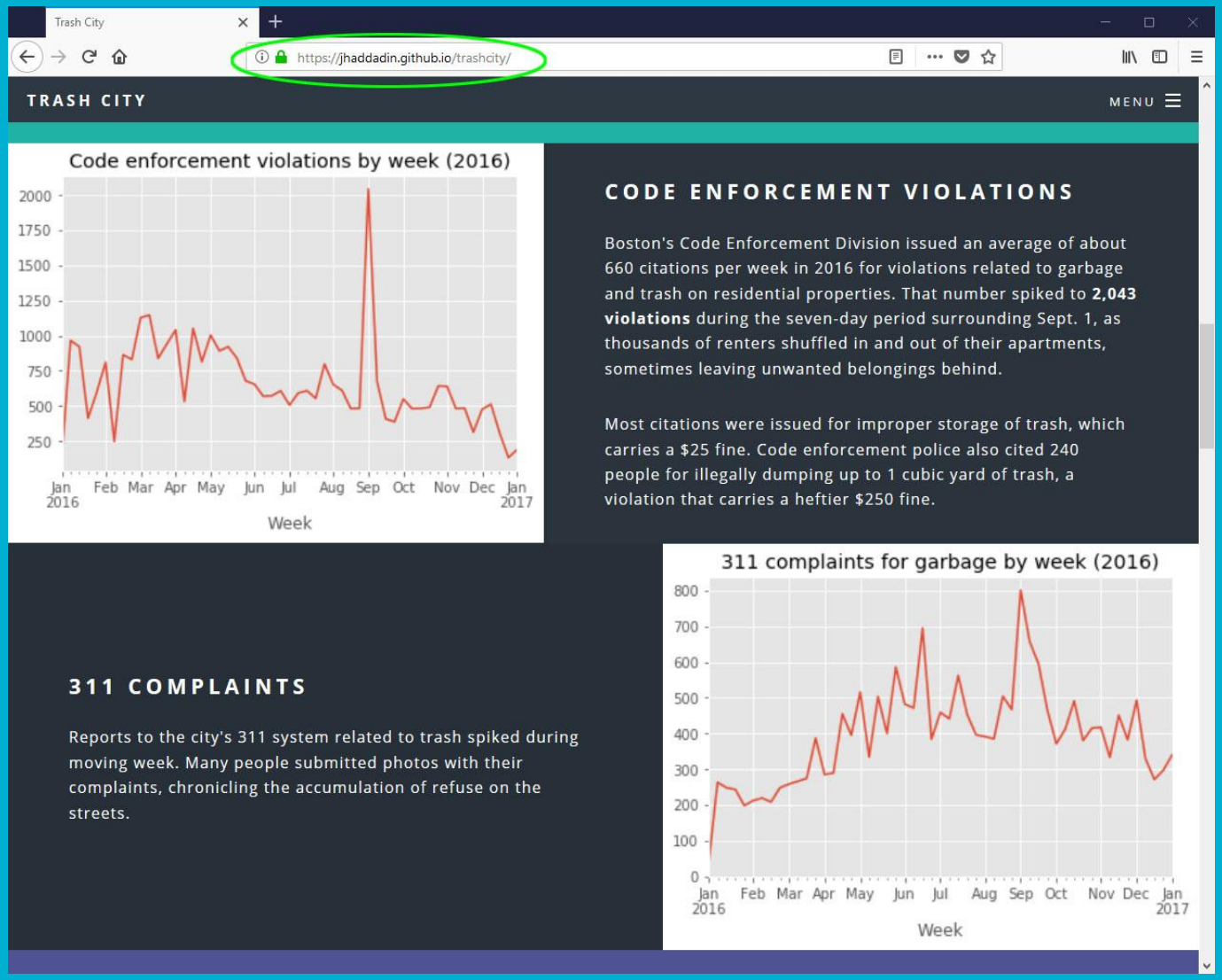
"Open Data Initiatives" come in two flavors:

- 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

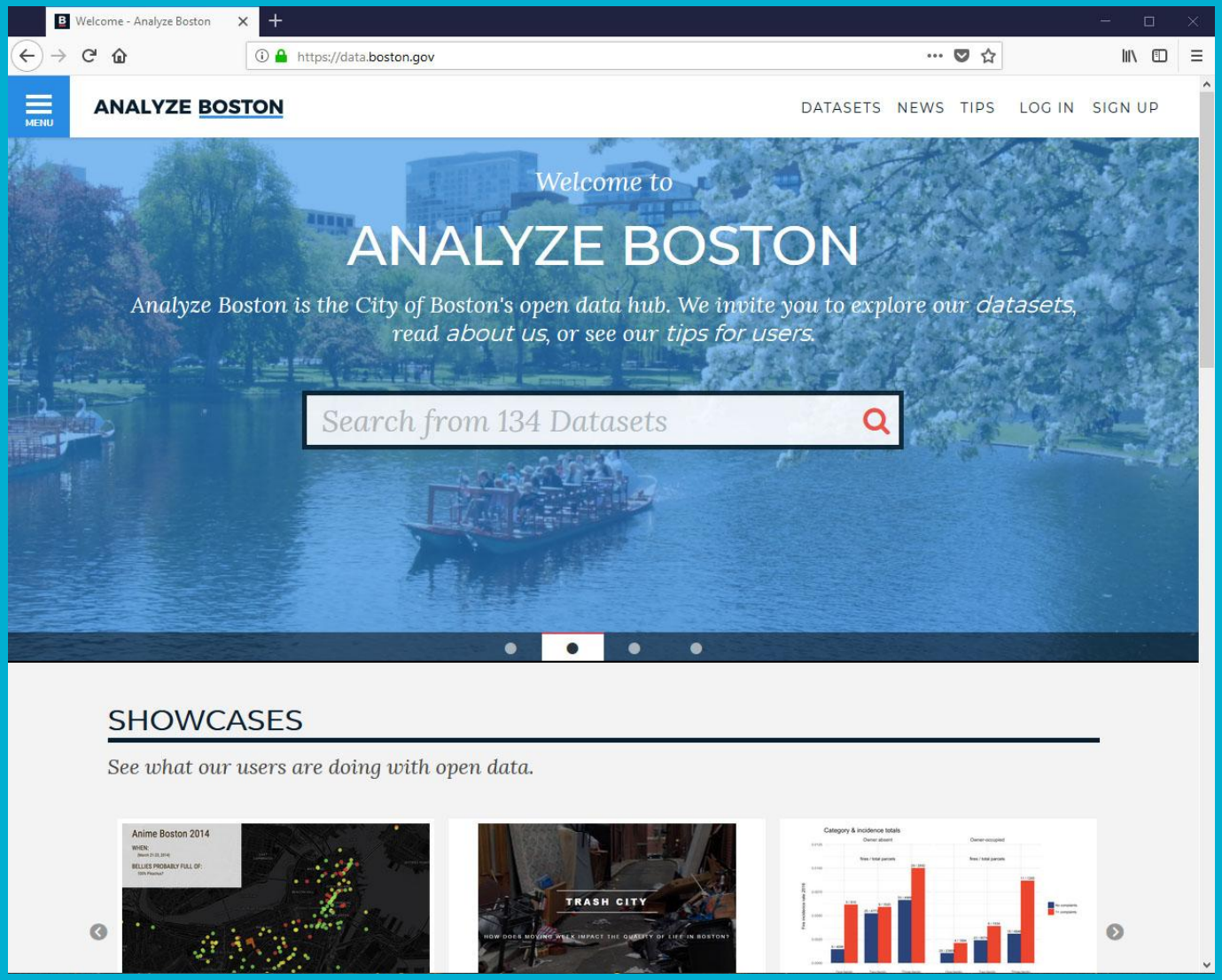
Data Publishing



Data Use


Trash City
https://jhaddadin.github.io/trashcity/
TRASH CITY
MENU
Code enforcement violations by week (2016)
2000
1750
1500
1250
1000
750
500
250
Jan 2016
Feb
Mar
Apr
May
Jun
Jul
Aug
Sep
Oct
Nov
Dec
Jan 2017
Week
CODE ENFORCEMENT VIOLATIONS
Boston's Code Enforcement Division issued an average of about 660 citations per week in 2016 for violations related to garbage and trash on residential properties. That number spiked to 2,043 violations during the seven-day period surrounding Sept. 1, as thousands of renters shuffled in and out of their apartments, sometimes leaving unwanted belongings behind.
Most citations were issued for improper storage of trash, which carries a $25 fine. Code enforcement police also cited 240 people for illegally dumping up to 1 cubic yard of trash, a violation that carries a heftier $250 fine.
311 COMPLAINTS
Reports to the city's 311 system related to trash spiked during moving week. Many people submitted photos with their complaints, chronicling the accumulation of refuse on the streets.
311 complaints for garbage by week (2016)
800
700
600
500
400
300
200
100
0
Jan 2016
Feb
Mar
Apr
May
Jun
Jul
Aug
Sep
Oct
Nov
Dec
Jan 2017
Week

Hybrid? (Not really!)



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.”

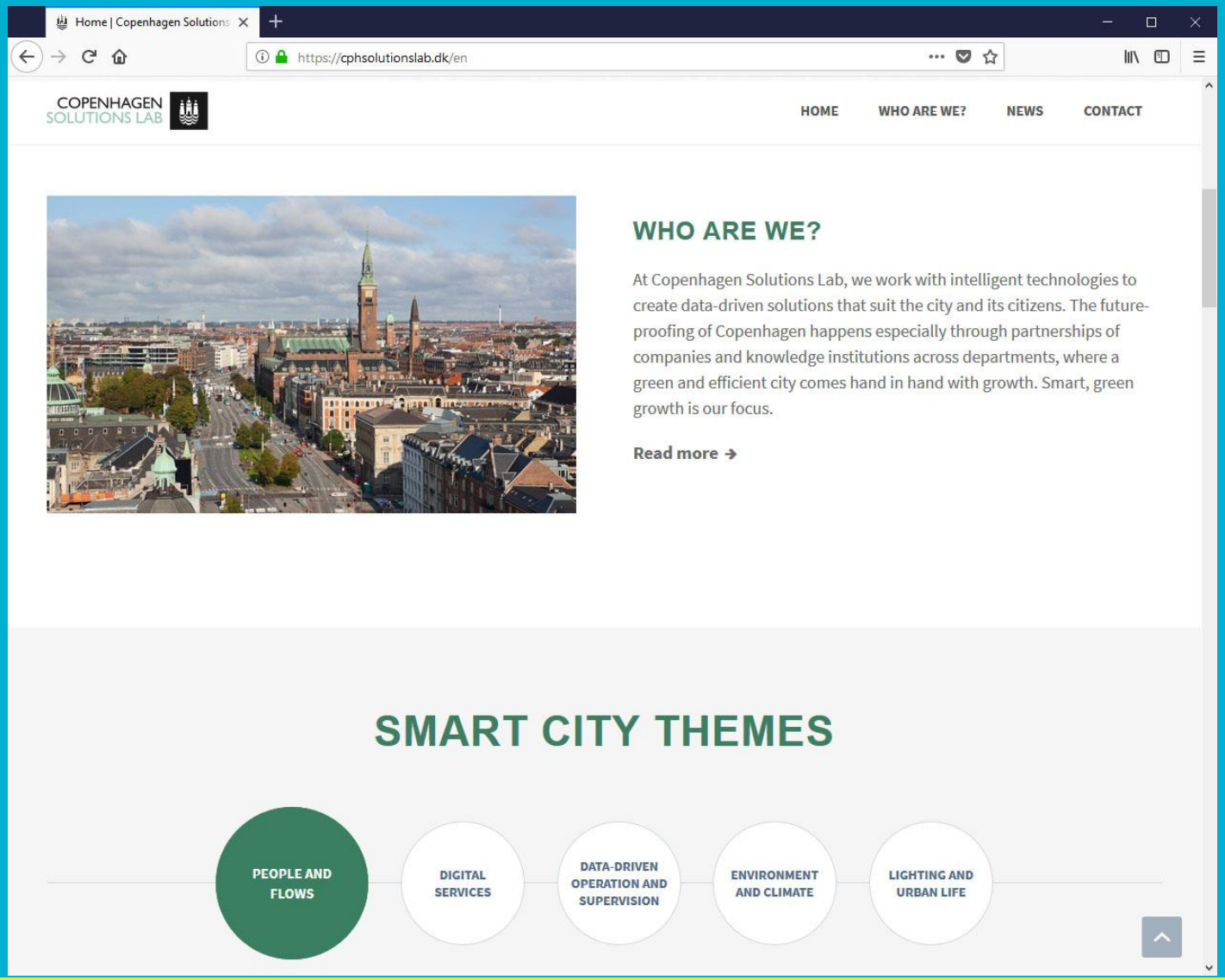
Topic issues

- 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 vauge

- *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

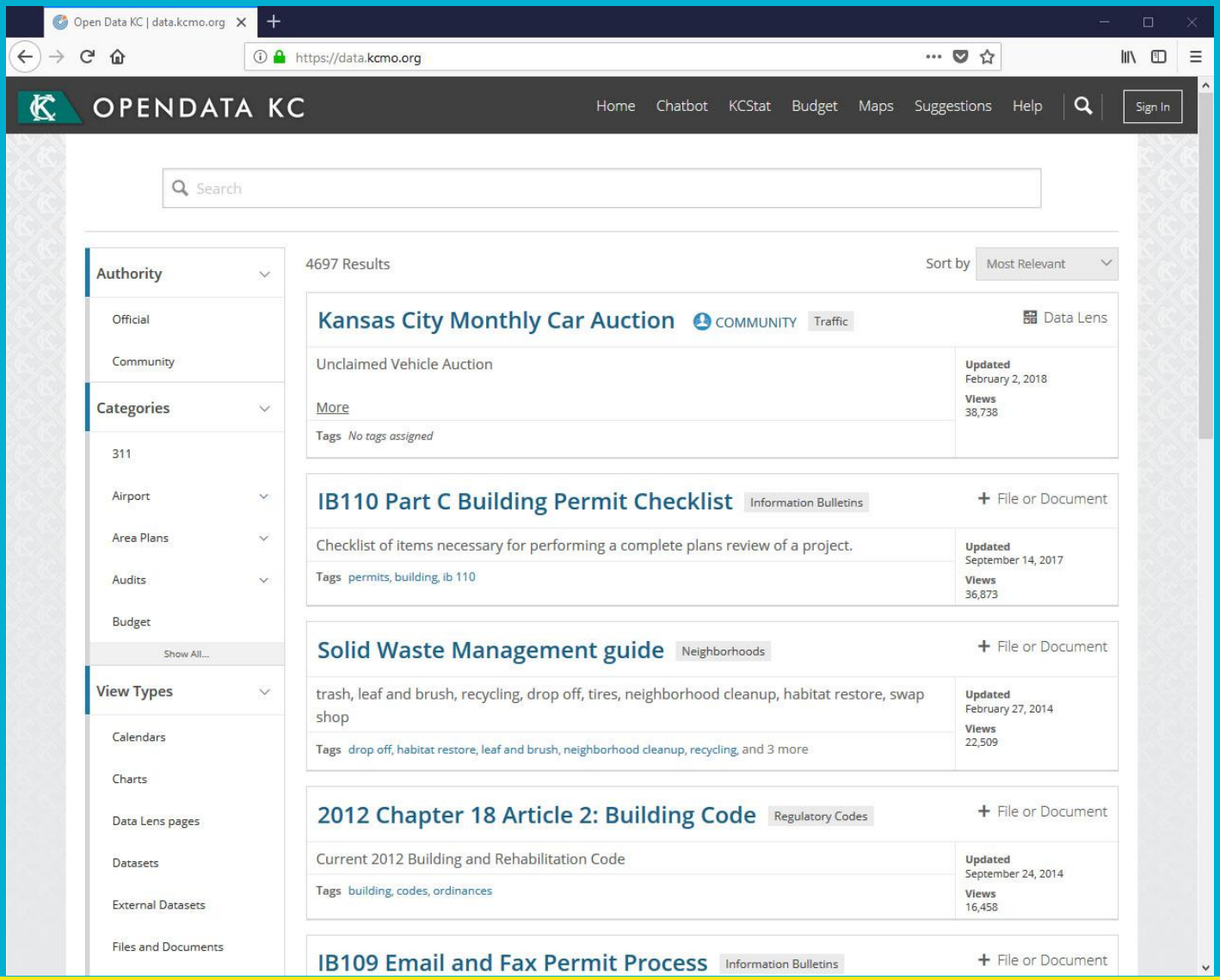
Initiative Structure and Plan

- 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.

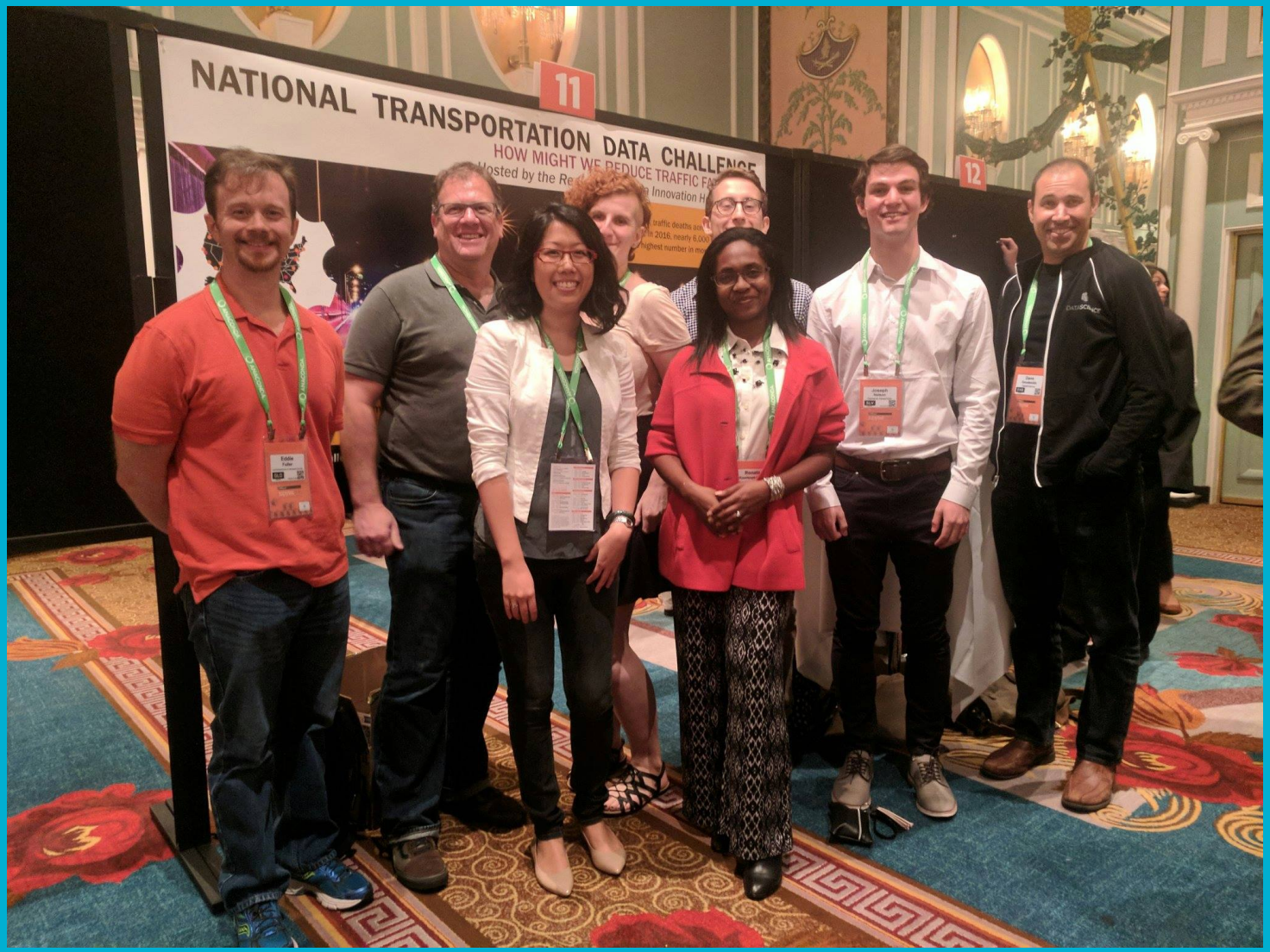
How did Kansas City do it?



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

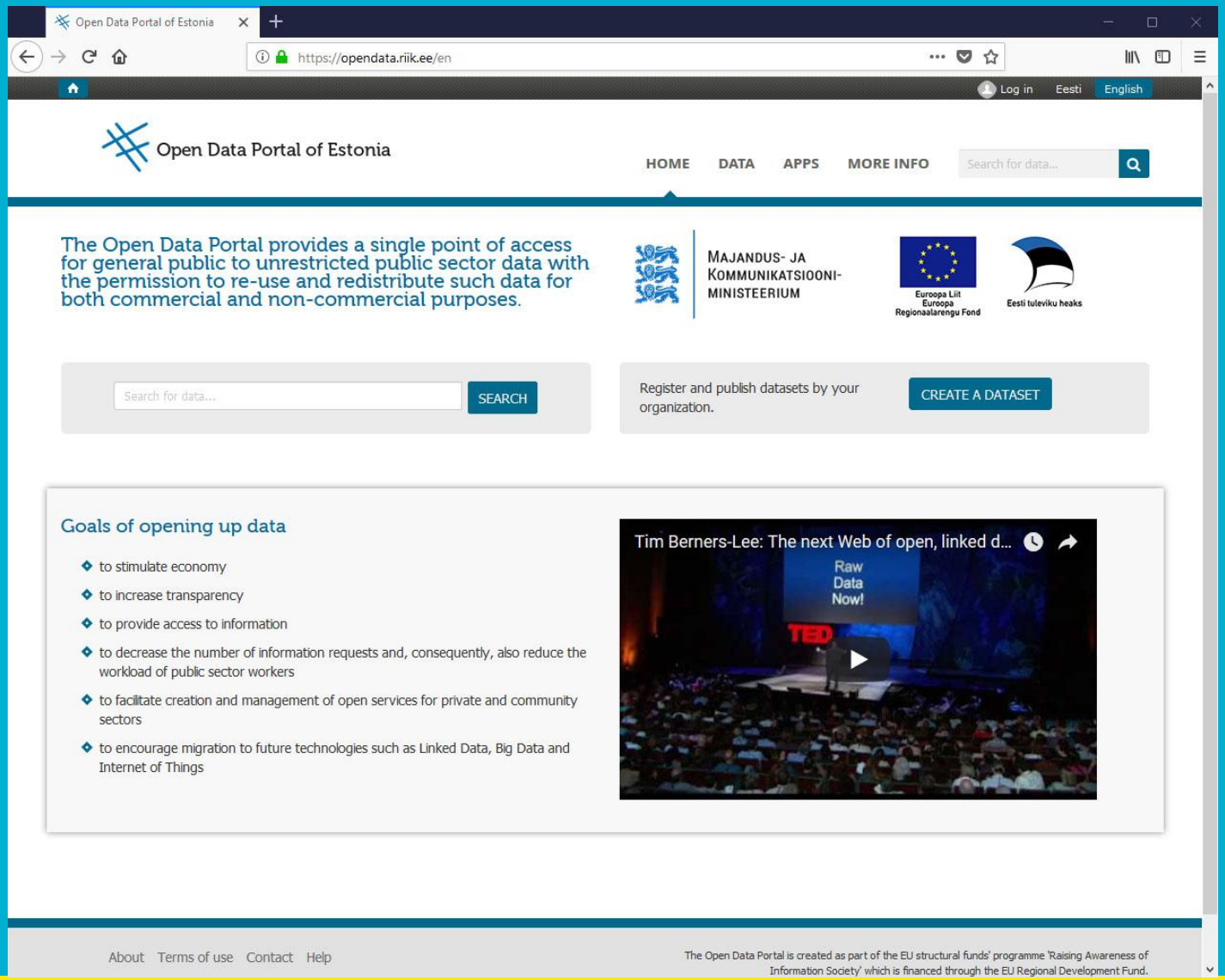
- You can’t launch an initiative without a lot of thought
- You need to have the right organization: Big Data Hubs didn’t fit a national public-private initiative
- There needs to be a lot of visible, public, focused activity
- There needs to be a regular feedback loop, so each event feeds into improvements 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

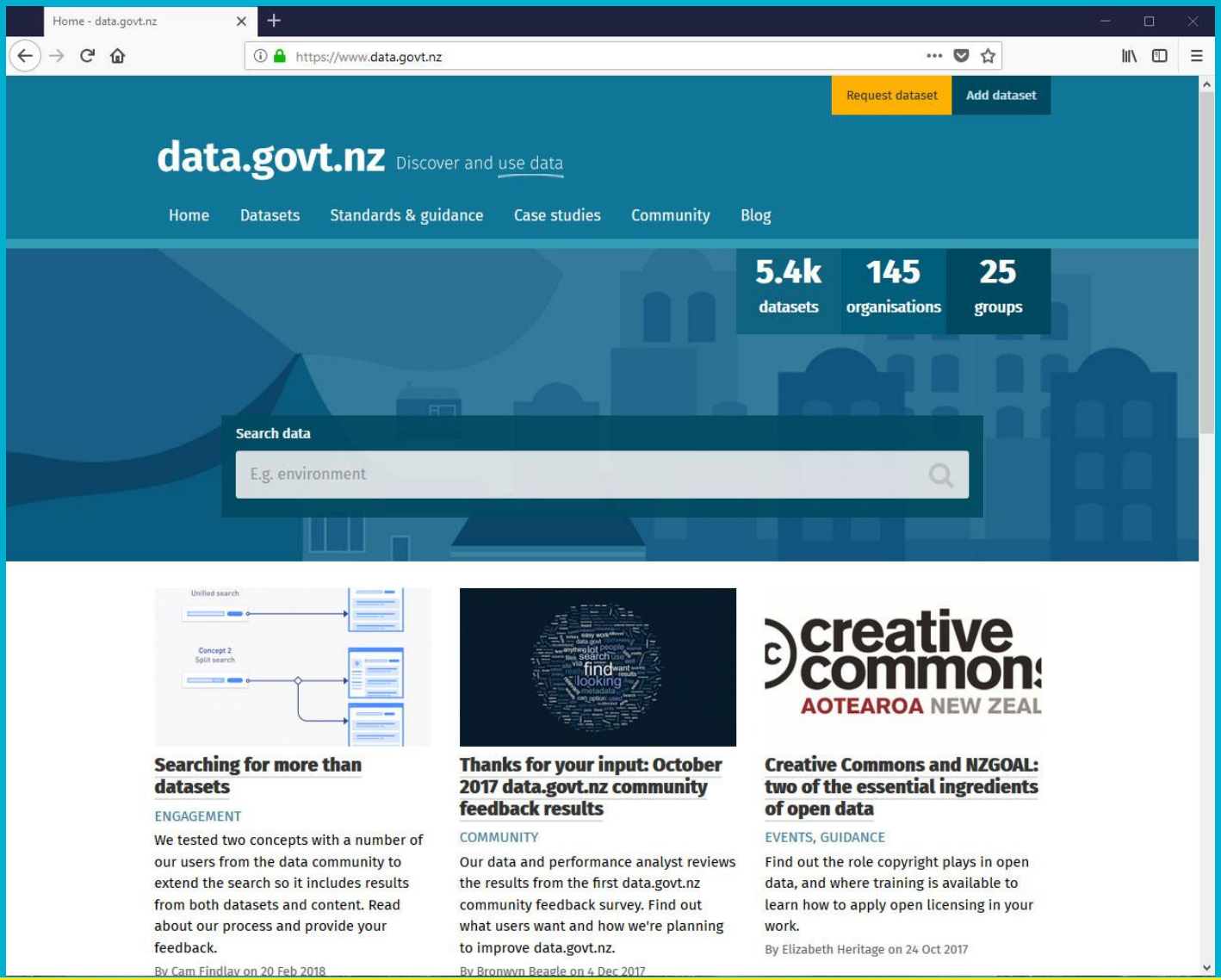
—

- 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

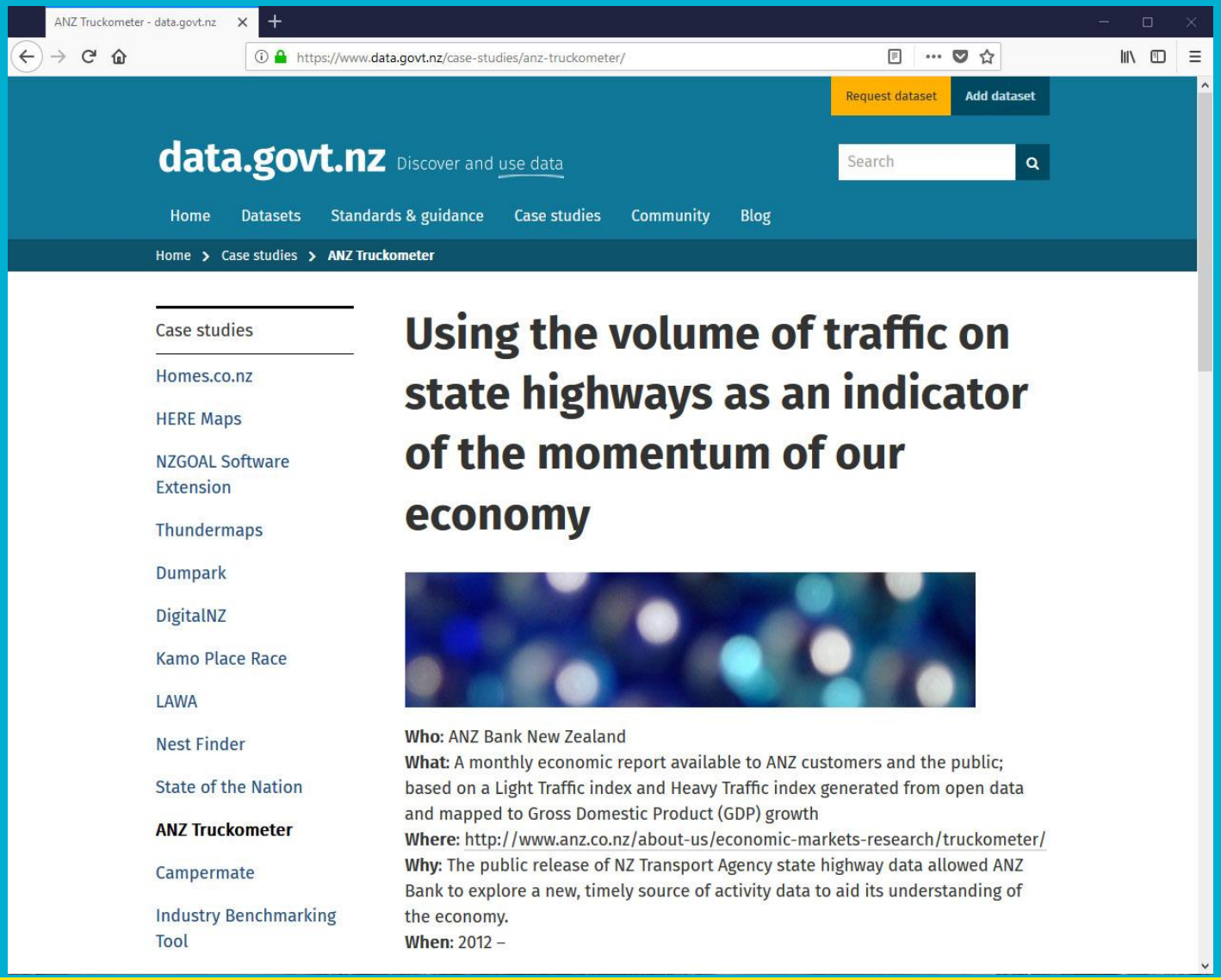
Open Data ES


Open Data Portal of Estonia
https://opendata.riik.ee/en
Log in
Eesti
English
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!
TED
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


Home - data.govt.nz
https://www.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 ZEAL
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



Open Data NZ case study: Beer!

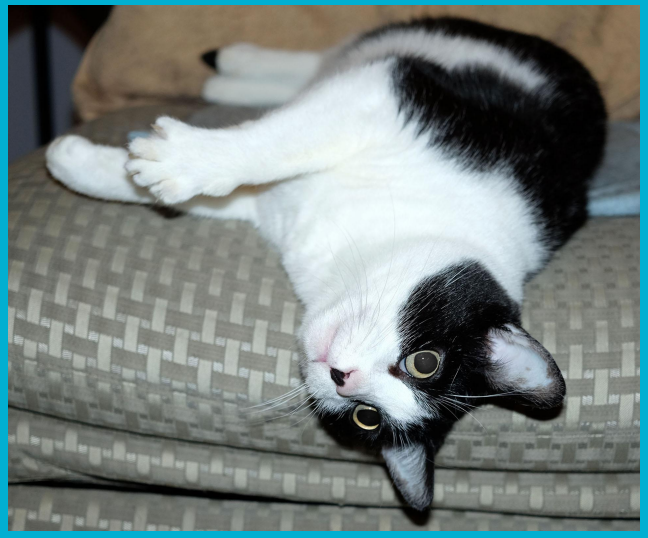
More final thoughts: Open data in the US

- 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 entites 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