

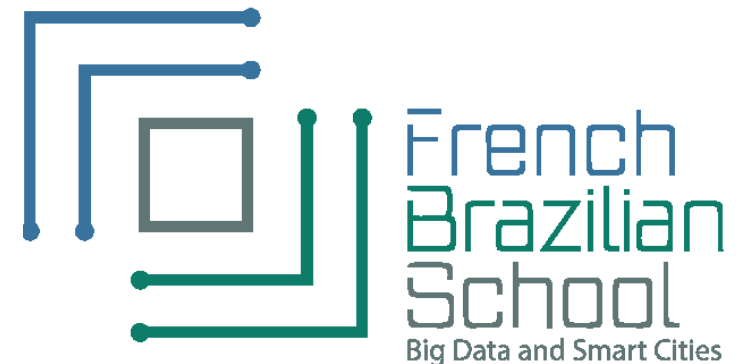
# Using open data for smart cities development

NOVEMBER 8-10, 2017

ANDREIWID SHEFFER CORRÊA



CC BY 4.0

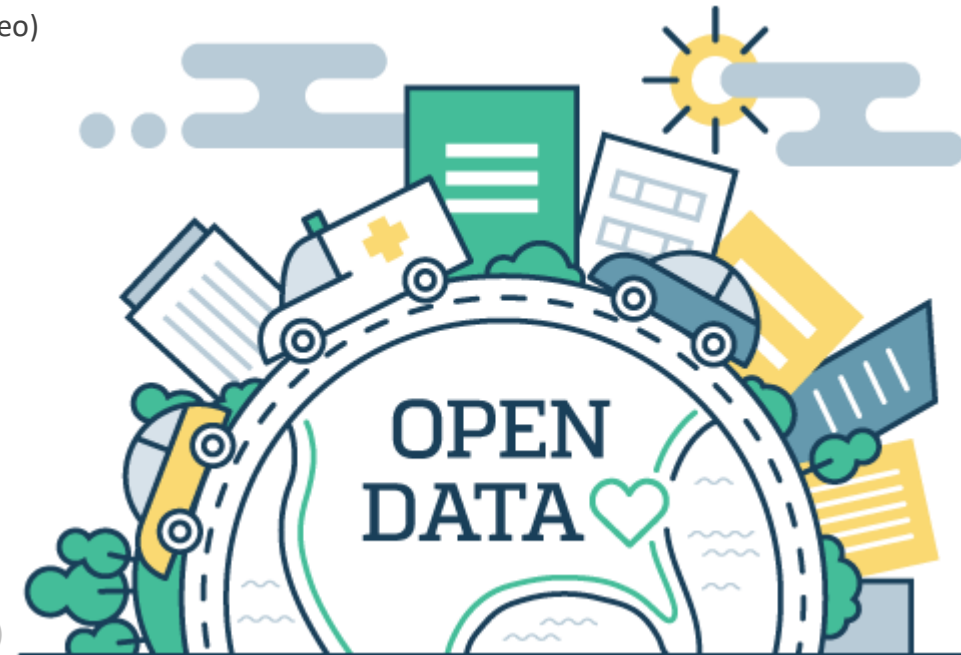


# Open Data and Smart Cities

---

*A smart city is an open city!*

(<https://www.youtube.com/watch?v=BHrsIIHJHeo>)



(images from <https://www.madhattertech.ca/chatter/open-data-and-journey-innovative-governments>)

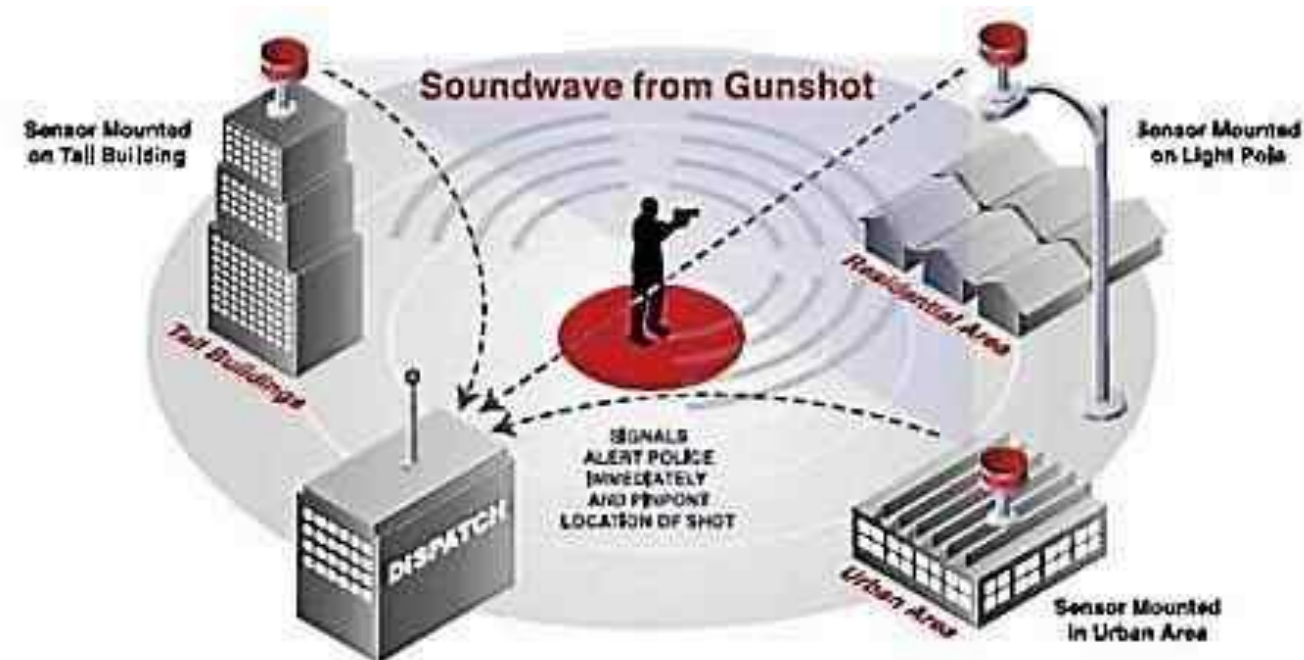
# Data-driven cities

Parking places assisted by sensors



(images from <http://seevider.com/solution.htm>)

Triangulating soundwave from gunshot



(images from <http://www.trentonian.com/article/TT/20140616/NEWS/140619760>)

# Data-driven cities

Intelligent traffic lights with video and sensors



(images from <https://www.smartcityinitiative.com/tag/traffic-light/>)

Sensors measure noise, pollution, dust quantities and garbage levels



(images from [http://www.libelium.com/smart\\_cities/](http://www.libelium.com/smart_cities/))

# What's Open Data

## The Open Data Definition

- “Open data and content can be **freely used, modified, and shared by anyone for any purpose**”

## Open Data ≠ Big Data

- Big Data is defined by its size, complexity, and changeability
  - e.g. Data from sensors, mobile phone GPS signals
- Open Data is defined by its use
  - Must be available to the public
  - May come from Big Data

The screenshot shows the 'Open Definition' website. At the top, there's a navigation bar with links: 'Open Definition', 'The Definition', 'Conformant Licenses', 'Participate', and 'News'. The main heading is 'The Open Definition'. Below it, a paragraph states: 'The **Open Definition** sets out principles that define "openness" in relation to data and content.' Another paragraph explains: 'It makes **precise** the meaning of "open" in the terms "**open data**" and "**open content**" and thereby ensures **quality** and encourages **compatibility** between different pools of open material.' A quote box says: 'It can be summed up in the statement that: "Open means **anyone** can **freely access, use, modify, and share** for **any purpose** (subject, at most, to requirements that preserve provenance and openness)."' Another quote box states: 'Put most succinctly: "Open data and content can be **freely used, modified, and shared by anyone** for **any purpose**"'. A blue button reads 'Read the full Open Definition »'. On the right, a sidebar titled 'THE OPEN DEFINITION IN YOUR LANGUAGE' lists various languages. At the bottom, a footer mentions 'Open Definition is a project of Open Knowledge International - Source Code' and provides links for 'Terms of use', 'Privacy policy', and 'IP Policy'.

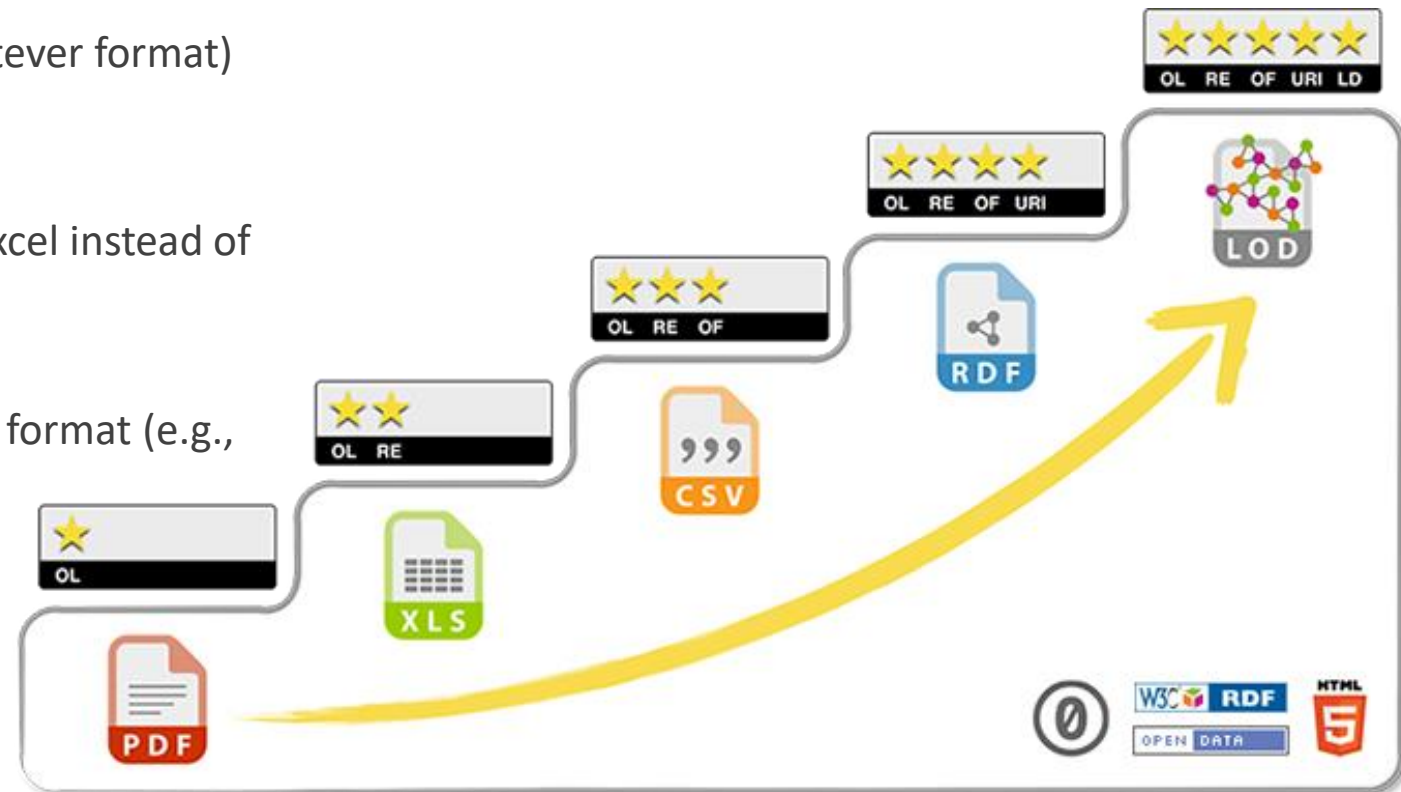
(images from <http://opendefinition.org/>)

# What's Open Data

## 5-star deployment scheme (first appeared 2006)

(images from <http://5stardata.info/en/>)

- OPEN LICENSE (OL)
  - Make your stuff available on the Web (whatever format) under an open license
- STRUCTURED DATA (RE)
  - Make it available as structured data (e.g., Excel instead of image scan of a table)
- NON-PROPRIETARY FORMAT (OF)
  - Make it available in a non-proprietary open format (e.g., CSV instead of Excel)
- USE OF URIs (URI) – identifiers
- LINKED DATA (LD)



# What's Open Data

## The 8 Principles of Open Government Data

- In 2007 a group of 30 advocates wrote a set o principles

Principle	Description	Recommendations
1-Complete	All public data is made available	Made available electronically
2-Primary	As collected at the source (not in aggregate, highest granularity)	Consider multiple views of data
3-Timely	Available as quickly as necessary to preserve the value of the data	Update data! Update data!
4-Accessible	Available to the widest range of users for the widest range of purposes	Provide APIs – automated access
5-Machine processable	Data is reasonably structured to allow automated processing	Avoid PDFs and images. Think tabularly.
6-Non-discrimatory	Available to anyone, with no requirement of registration	Allow anonymous access
7-Non-proprietary	Available in a format over which no entity has exclusive control	Use multiple formats
8-License-free	Not subject to any copyright, patent, trademark or trade secret regulation	Be clear about licensing (mix of public records, personal info, copyrighted work..)



# How Open Data benefits cities?

- Promotes new ideas and businesses (products and services)
- Allows third party apps instead of government apps
- Opportunity for small businesses
- Improves city efficiency



(images from <https://smartcitiesworld.net/news/news/smart-cities-services-worth-225bn-by-2026-1618>)



# Benchmarking

---

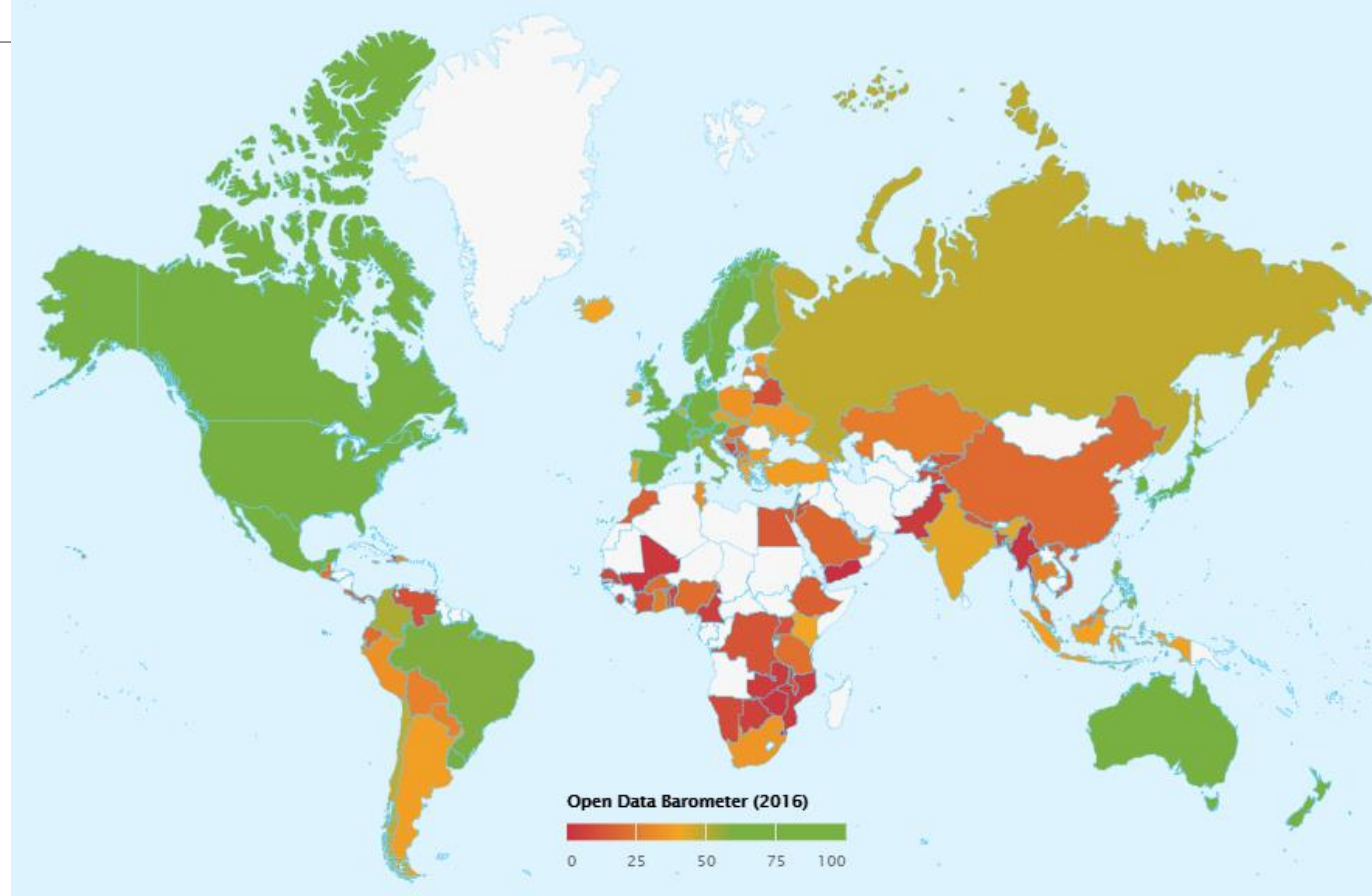
- Means evaluating and ranking countries, organizations and projects
- Improves accountability
- Emphasizes best practices
- Understand use open data for solving problems
- Mainly evaluates country-level initiatives

# Benchmarking: Global Open Data Index

Rank	Place	Government Budget	National Statistics	Procurement	National Laws	Administrative Boundaries	Draft Legislation	Air Quality	National Maps	Weather Forecast	Company Register	Election Results	Locations	Water Quality	Government Spending	Land Ownership	Score
1	Taiwan																90%
2	Australia																79%
2	Great Britain																79%
4	France																70%
5	Finland																69%
5	Canada																69%
5	Norway																69%
8	New Zealand																68%
8	Brazil																68%

(images from <https://index.okfn.org/place/>)

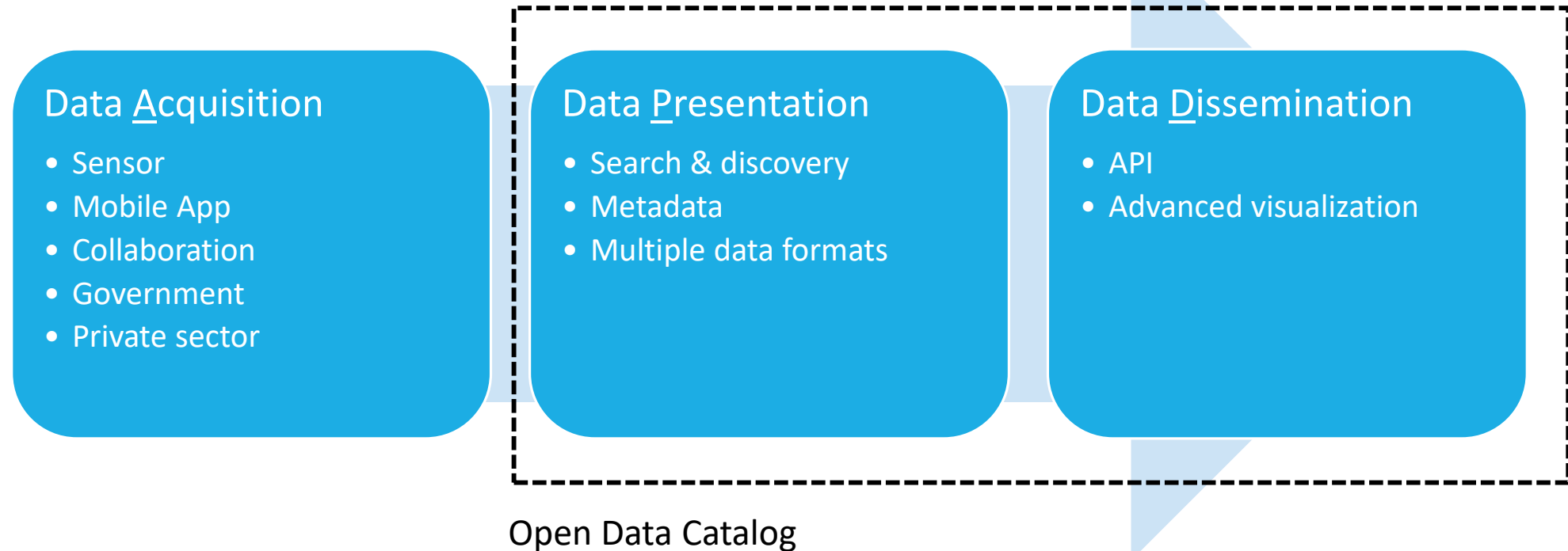
# Benchmarking: Open Data Barometer



(images from [/?\\_year=2016&indicator=ODB](#))

# How Open Data looks like?

## Open Data lifecycle – Acquisition, Presentation and Dissemination



# How Open Data looks like?

---

## Open Data catalogs

- Software to make data available
  - It's possible to implement Open Data Principles
- There are many solution in the market
- The most famous is CKAN - Comprehensive Knowledge Archive Network
- Free & open source software
- Developed by Open Knowledge (<http://okfn.org>)
- Available + languages (including French, Portuguese and English)
- Sandbox at <http://demo.ckan.org>



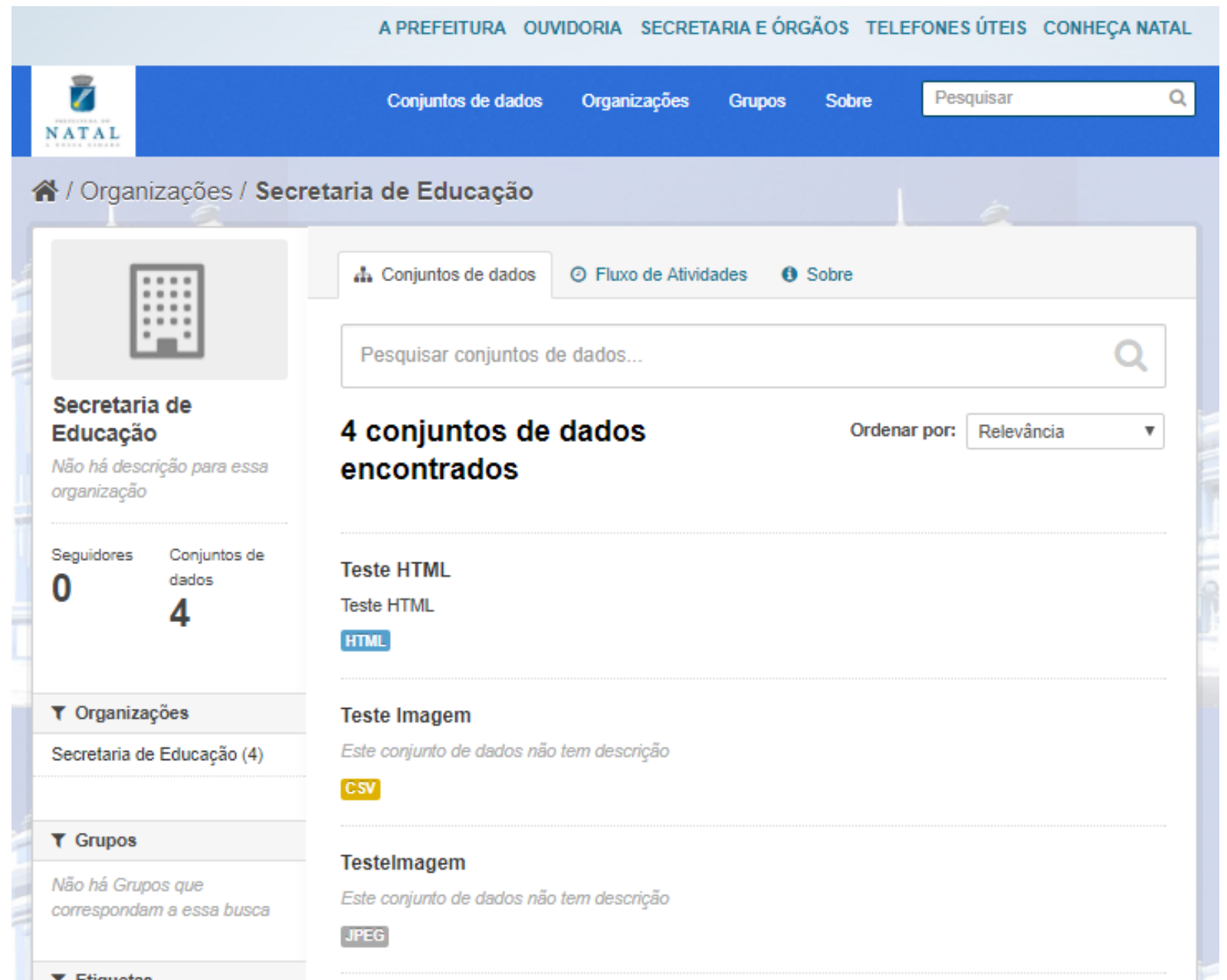
## Natal-RN Data Portal

<http://dados.natal.rn.gov.br>

Seems to be a test instance...

With only four datasets!

(as of Nov, 2017)



The screenshot shows the Natal-RN Data Portal interface. At the top, there is a navigation bar with links: A PREFEITURA, OUVIDORIA, SECRETARIA E ÓRGÃOS, TELEFONES ÚTEIS, and CONHEÇA NATAL. Below this is a blue header with the portal logo and navigation links: Conjuntos de dados, Organizações, Grupos, and Sobre. A search bar labeled 'Pesquisar' is also present. The main content area shows the breadcrumb path: / Organizações / Secretaria de Educação. On the left, there is a sidebar for the 'Secretaria de Educação' organization, showing 0 followers and 4 datasets. The main area displays '4 conjuntos de dados encontrados' (4 datasets found) and a search bar. Below this, there are four datasets listed: 'Teste HTML' (HTML format), 'Teste Imagem' (CSV format), 'Teste Imagem' (JPEG format), and 'Teste Imagem' (JPEG format). Each dataset entry includes a description and a format tag.



# How Open Data looks like?

## Some other CKAN Instances around the world



Data.gov



data.gov.au



Data.gov.uk



datos.gob.mx



Data.Rio



Dados Recife



The Brazilian Senate

And +140 instances  
<https://ckan.org/about/instances/>

# State-of-the-art report and evaluation of Open Data Catalogs (2015)

## Summary of catalog features

FEATURES	CKAN	DKAN	SOCRATA	PUBLISH MY DATA	INFO WKBENCH	ENIGMA	JUNAR	ODS	CALLIM	DATATK	SMWIKI
DATA, METADATA & FILE FORMAT STANDARDS	●	●	●	●	●	●	●	●	●	●	●
SEARCH & INDEXING	●	●	●	●	X	●	●	●	X	●	●
SOCIAL MEDIA, SHARING & COLLABORATION	●	●	●	●	●	X	●	●	●	X	●
PUBLISHING WORKFLOW	●	●	●	●	●	●	●	●	●	●	●
HARVESTING, FEDERATION & CATALOGUE	●	●	●	●	X	X	●	●	●	X	●
DATA ANALYSIS	●	●	●	X	●	●	●	●	X	●	X
VISUALISATION	●	●	●	X	●	X	●	●	X	X	●
PERSONALISATION	●	●	●	●	●	X	●	●	●	●	●
CUSTOMISATION	●	●	●	●	●	NA	●	●	●	●	●
LICENSING FOR DATASET	●	●	●	●	X	X	X	●	X	X	●
ACCESSIBILITY	●	●	●	●	●	NA	●	●	●	●	●
EXTENSIBILITY	●	●	●	●	●	●	●	●	●	●	●
TECHNICAL ENVIRONMENT	Python	PHP, Drupal CMS	Scala	Ruby on rails	Java & Web apps	NA	Java & Python	NA	Java	PHP	PHP
OTHERS	Good manual Simple to use	Easy to use platform	Tracking & Measure of performance	Flexible, cloud-based, easy to use	R stat, support transparency, linked data	Reliable, scalable, large OD Analyses	Track & measures user impact on OD	Remote web services; easy deployment	Guides, videos, tutorial. Linked data	Deal with fraud, aids transparency	None

● denotes full-fledged solution, ● denotes limited solution, x denotes that solution is not provided, NA denotes information not available

Extracted from <http://routetopa.eu/deliverable-2-1-state-of-the-art-report-and-evaluation-of-existing-open-data-platforms-now-available/>

# State-of-the-art report and evaluation of Open Data Catalogs (2015)

## Availability of extensibility

Platforms	Extensible	Open Source	Extension Mechanisms	Guide Available	Customisable Metadata
CKAN	●	●	●	●	●
DKAN	●	●	●	●	●
Socrata	●	x	●	●	●
PublishMyData	●	●	●	●	●
Information Workbench	●	●	●	x	●
Enigma	x	x	●	x	x
Junar	●	x	●	x	x
Open Data Soft	●	x	●	●	x
Callimachus	●	●	●	●	●
DataTank	●	●	●	●	x
Semantic MediaWiki	●	●	●	●	●

● denotes extensive solution, ● denotes limited solution, x denotes that solution is not provided

Extracted from <http://routetopa.eu/deliverable-2-1-state-of-the-art-report-and-evaluation-of-existing-open-data-platforms-now-available/>

# Know by examples: Transportation

---

## Blind Square – Smart Cities For All

- Use data from Foursquare and OpenStreetMap
- App provides audio directions for blind residents



(images from <http://www.perkinselearning.org>)

# Know by examples: Transportation

## Turnaround: Fixing New York City's Buses

- An initiative of a group of New Yorkers riders
- Using open data to raise awareness about bus system conditions
- Report Cards with speed, accuracy, and ridership volume



(images from <http://busturnaround.nyc>)

## Report Card for **BX3**

### **BX3**

#### SPEED

**6.6** mph



#### BUS BUNCHING

**1** of **8** buses bunched (**12.7%**)



#### RIDERSHIP

At **38,411** riders per day, the BX3 ranks **3rd** in ridership out of **43** Bronx routes.  
This represents a **3.9% decrease** in ridership since 2010.

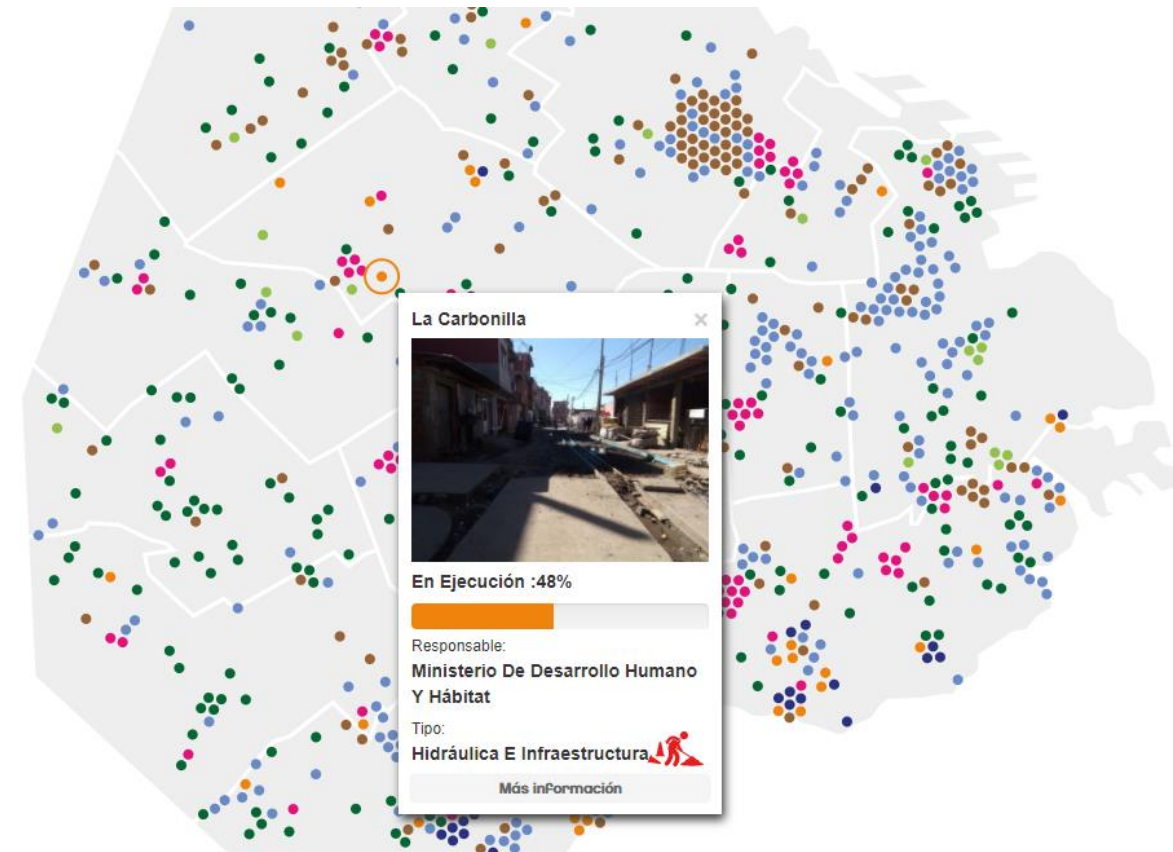
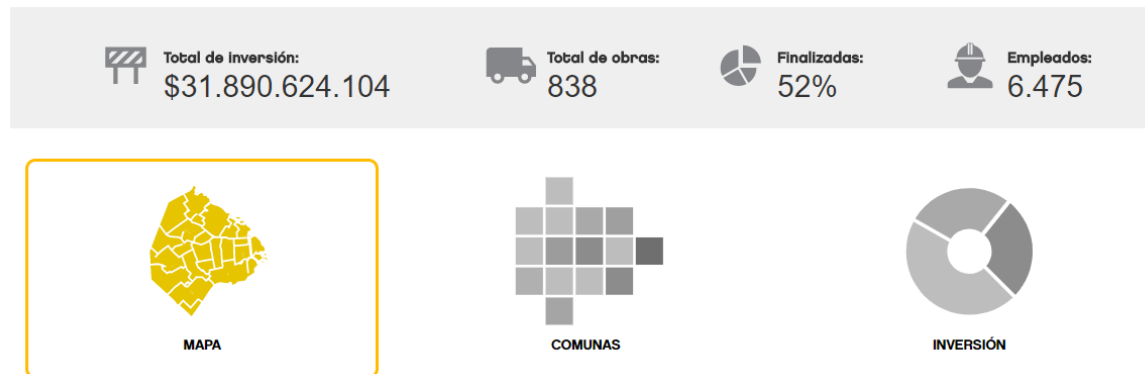




# Know by examples: City Planning

## BA Obras Buenos Aires

- Website for monitoring public works
- Based on official data
- Rich interface with maps and graphics
- Suggestions for data visualization

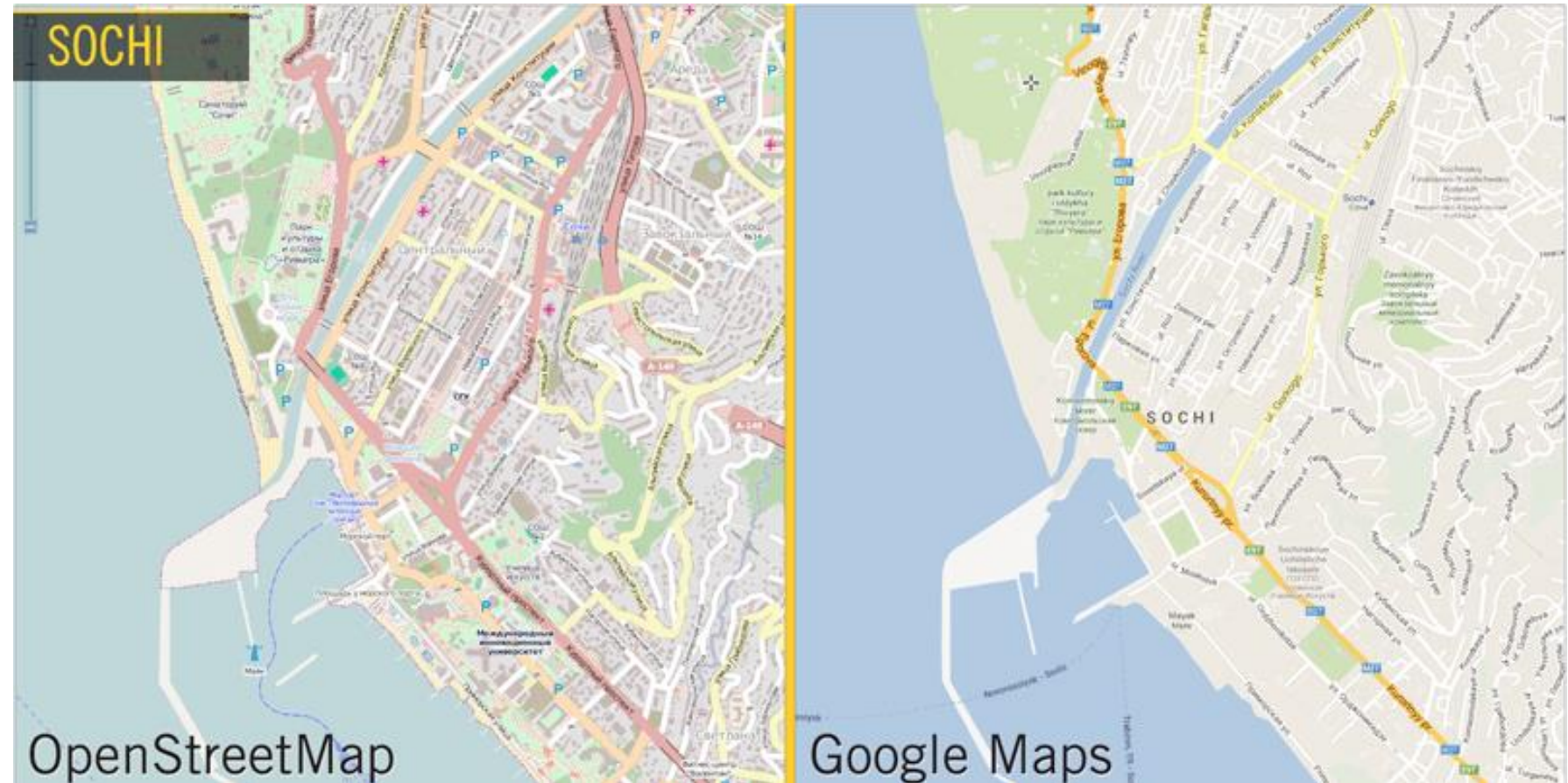




# Know by examples: City Planning

**OpenStreetMap:**  
“the Wikipedia of maps”

Current OSM map  
vs. Google Map of  
Sochi, Russia where  
the 2014 Olympic  
Games begin on  
Feb. 7



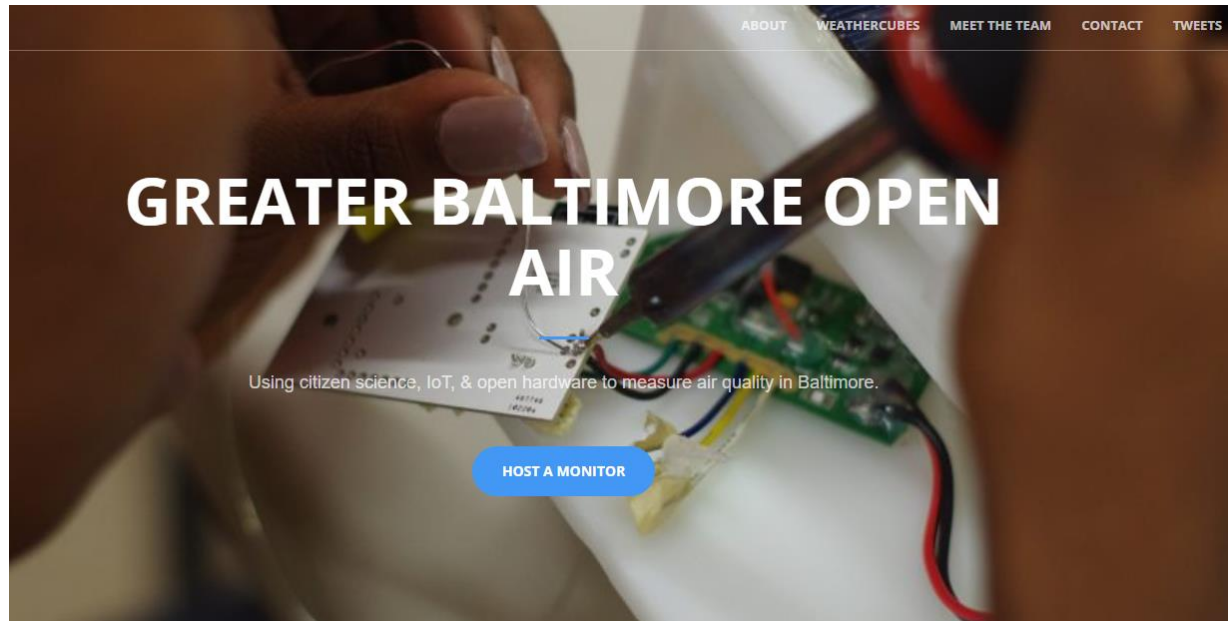
(images from <https://stevecoast.com/2014/01/30/its-time-to-make-openstreetmap-your-only-street-map/>)

# Know by examples: Resilient City

---

## Baltimore Open Air Project

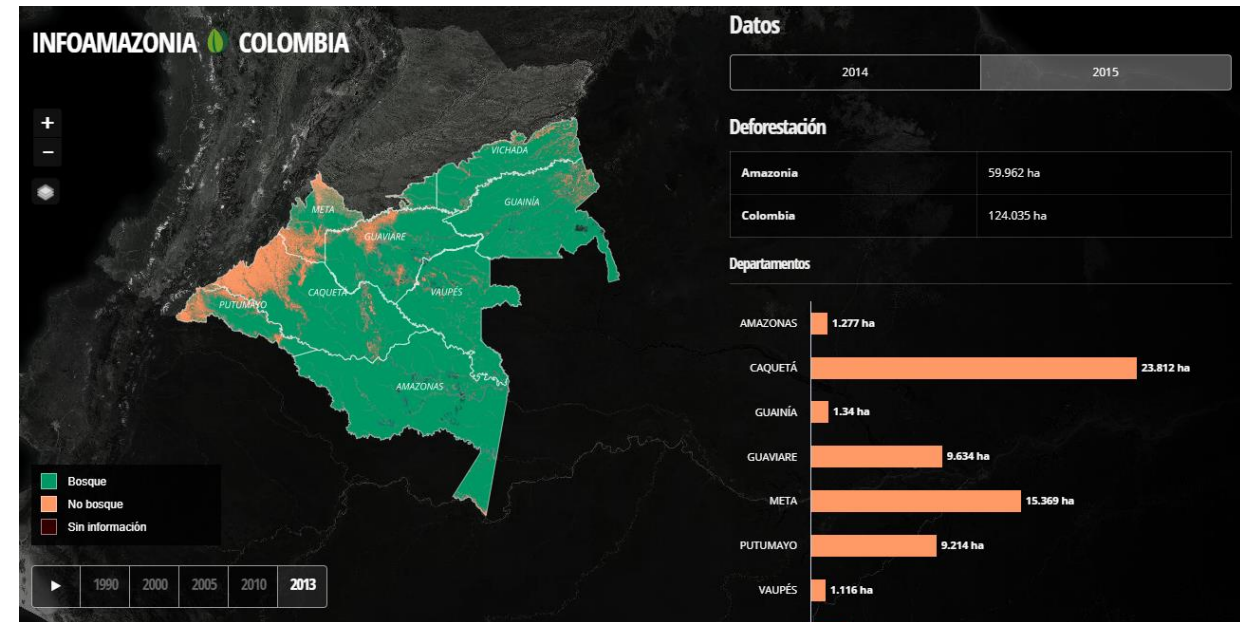
- Project to distribute stations to monitor air quality
- They call those stations WeatherCubes
- Monitor temperature, humidity, ozone, and nitrogen dioxide
- Everyone can apply for hosting a WeatherCube and be part of the network



# Know by examples: Resilient City

## InfoAmazonia Colombia

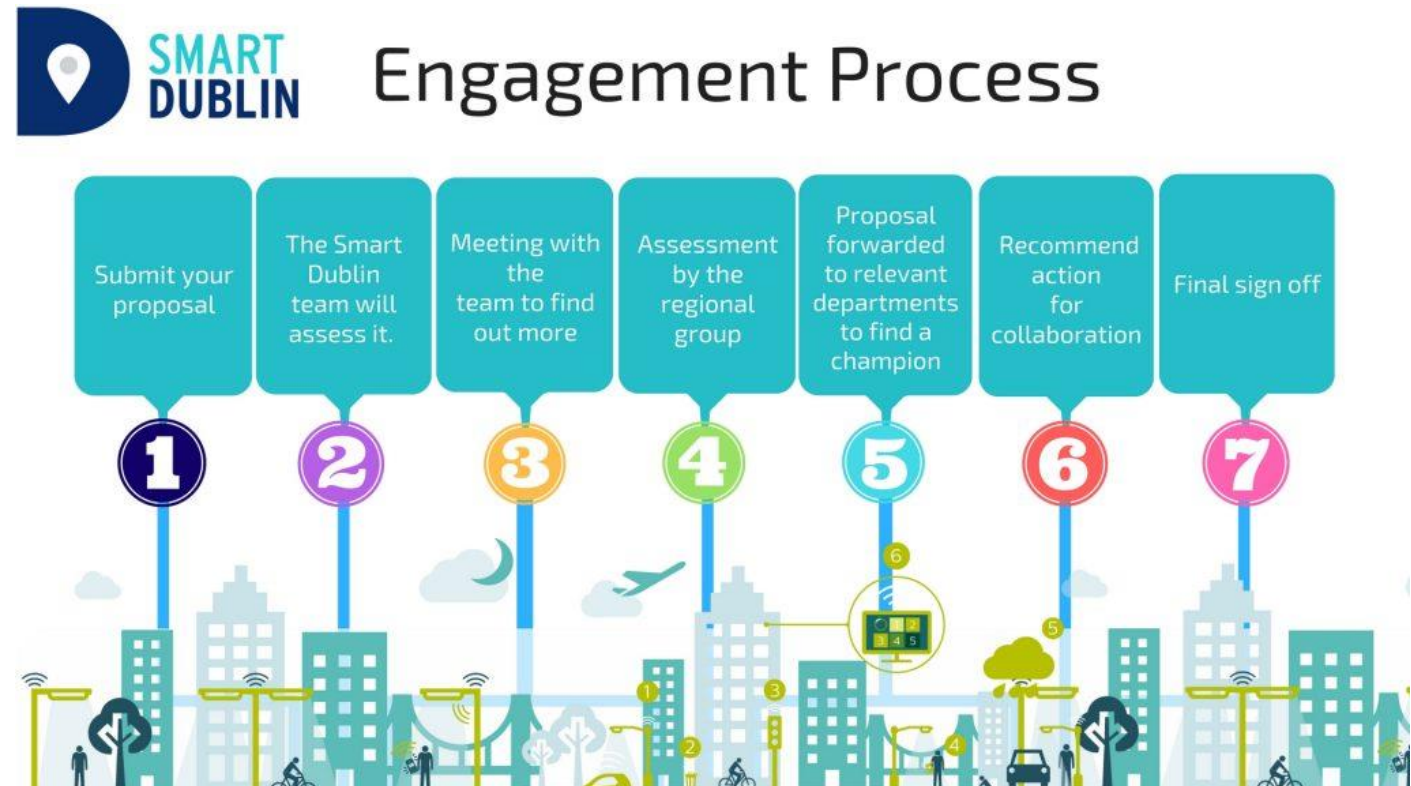
- A project to reduce deforestation
- It uses data from satellites
- Also data crowdsourced from data journalists
- Provide alerts to citizens when identify new construction projects



# Know by examples: Engagement

## Smart Dublin

- A project to engage with smart technology providers
- Providers have benefits from local authorities
- Data is shared to solve challenges and improve city life
- There is an engagement process

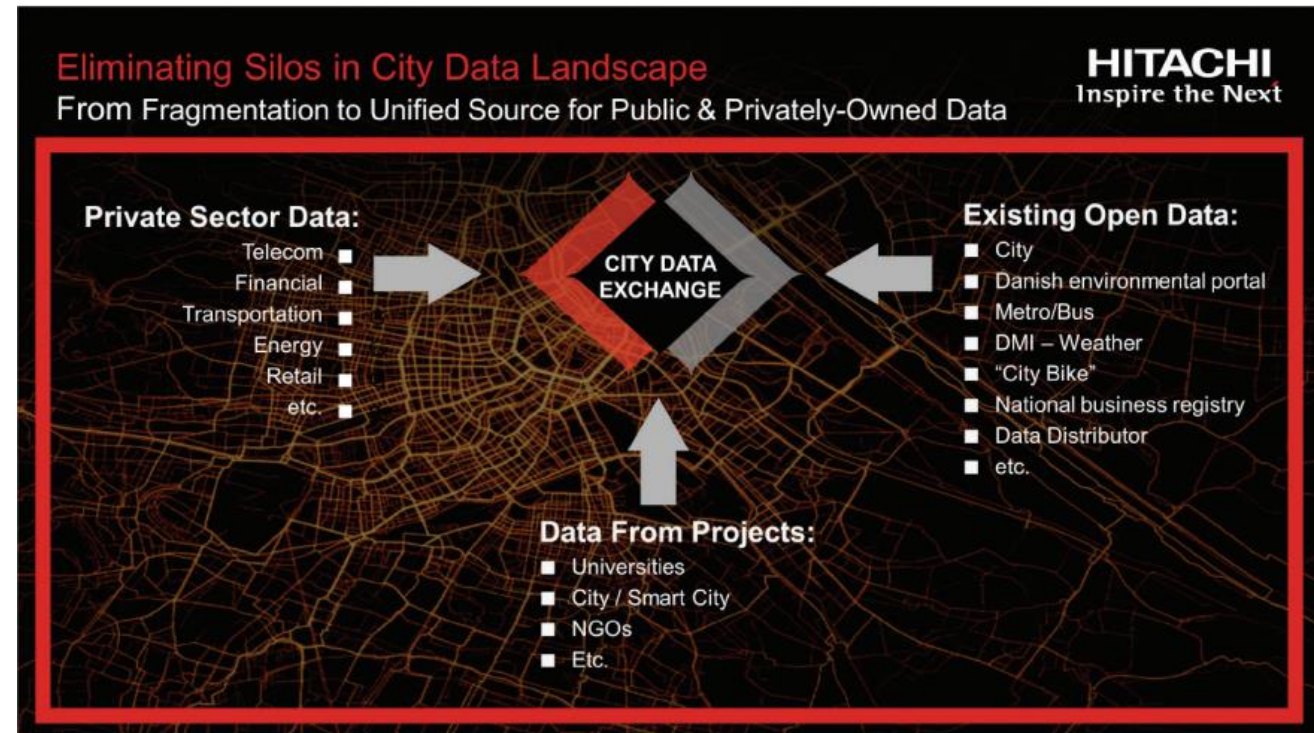




# Know by examples: Public and Private

## The City Data Exchange (Copenhagen)

- Integrates data from public and private data owners
- Tries to avoid silos of data
- Based on a proprietary platform from Hitachi
- Owners can charge to access data

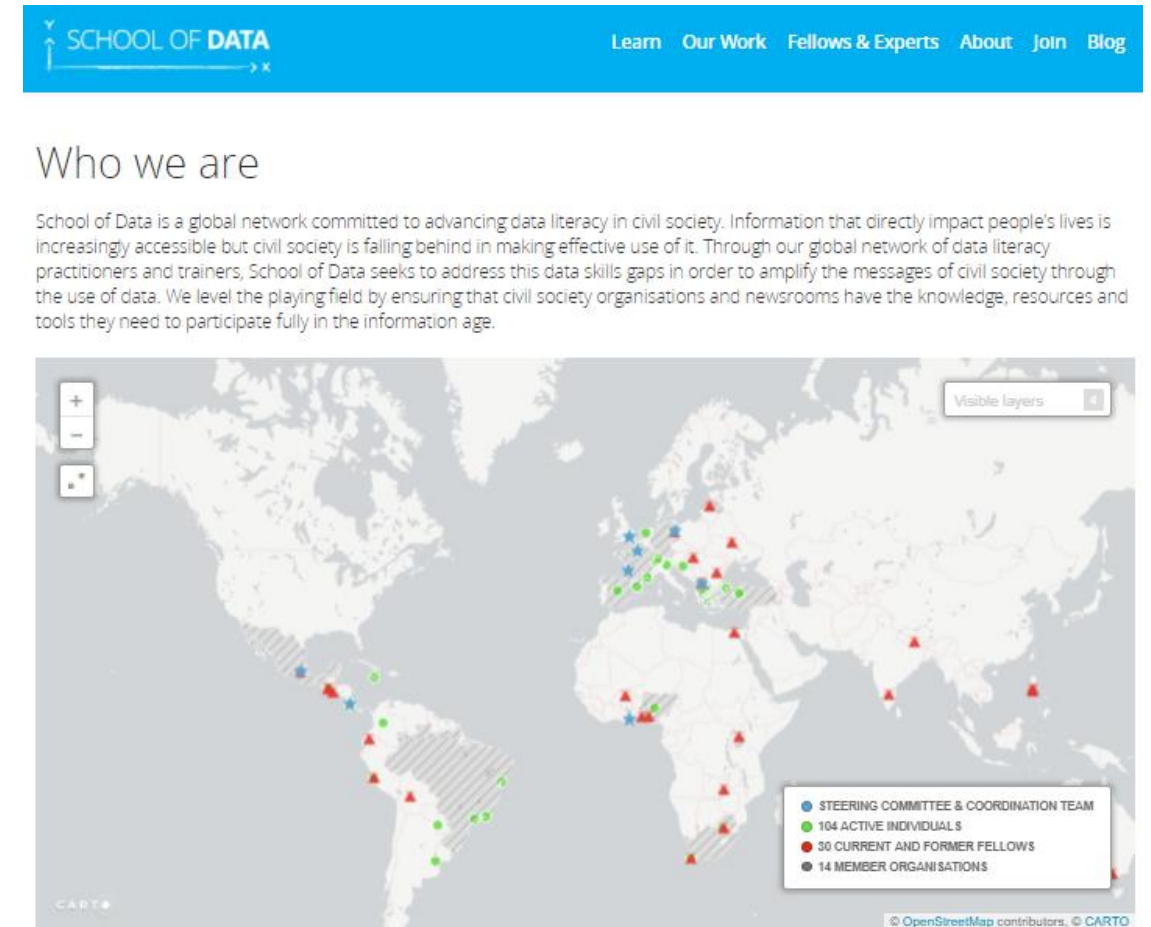


(images from <https://www.hitachivantara.com/en-us/pdf/solution-profile/hitachi-solution-profile-city-data-exchange.pdf>)

# Know by examples: Spreading the word!

## School of data (<https://schoolofdata.org>)

- Global network committed to advancing data literacy
- Aims at empowering civil society, journalists and citizens
- Promote training, workshops and community events
- 13 organizations and 101 active individuals (as of Nov 2017)
- Presence in developing countries





# Some points to think about (carefully!)

---

- Keep the mantra: “OPEN BY DEFAULT”
- Opening data is not only for governments
- Legislation is not enough!
- Disclosing data is not enough: we need social benefits (open data impact)
- Clear understanding of licensing issues
- The spread of documents instead of data (specially at municipal level)
- Go beyond hackathons to a continuous development of smart cities

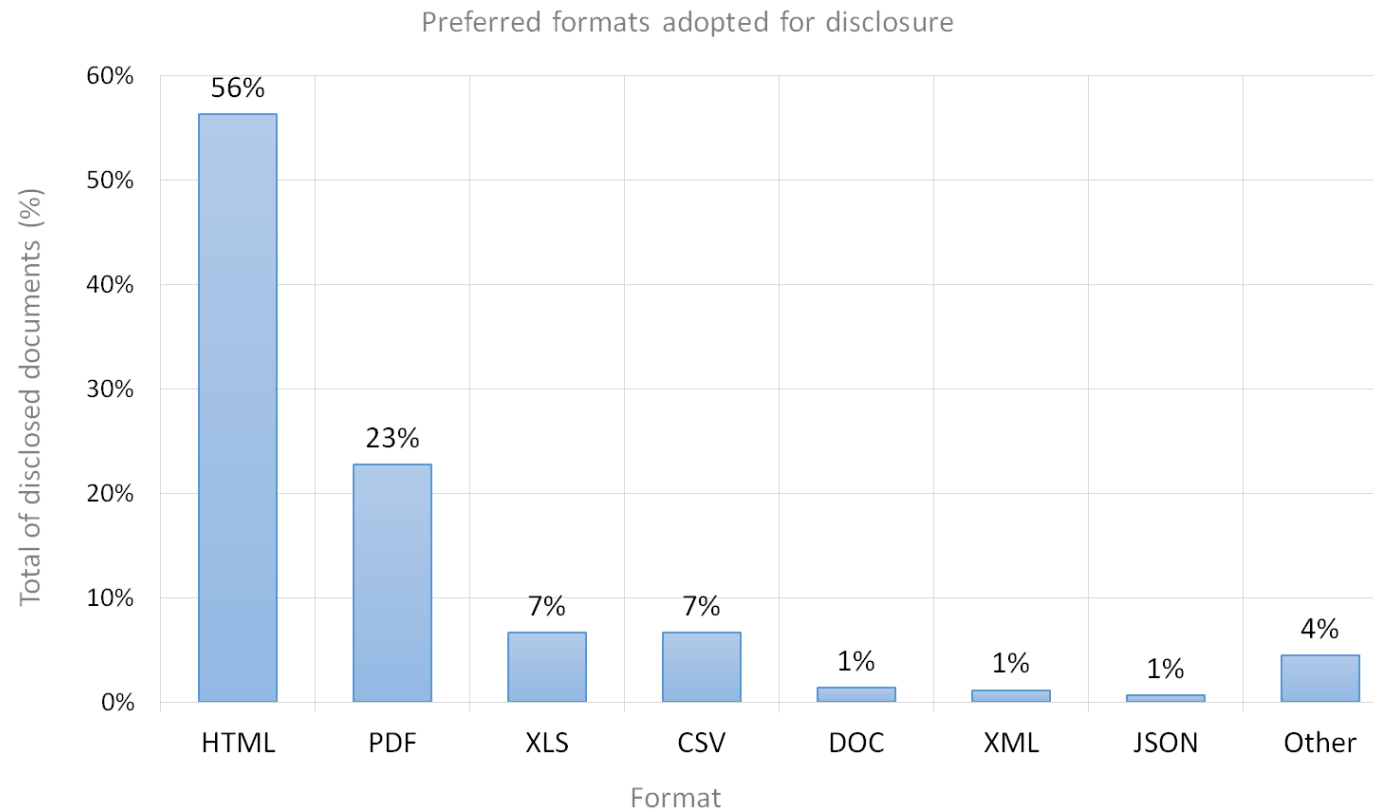
# Data openness in Brazilian local governments

---

## Research paper highlights

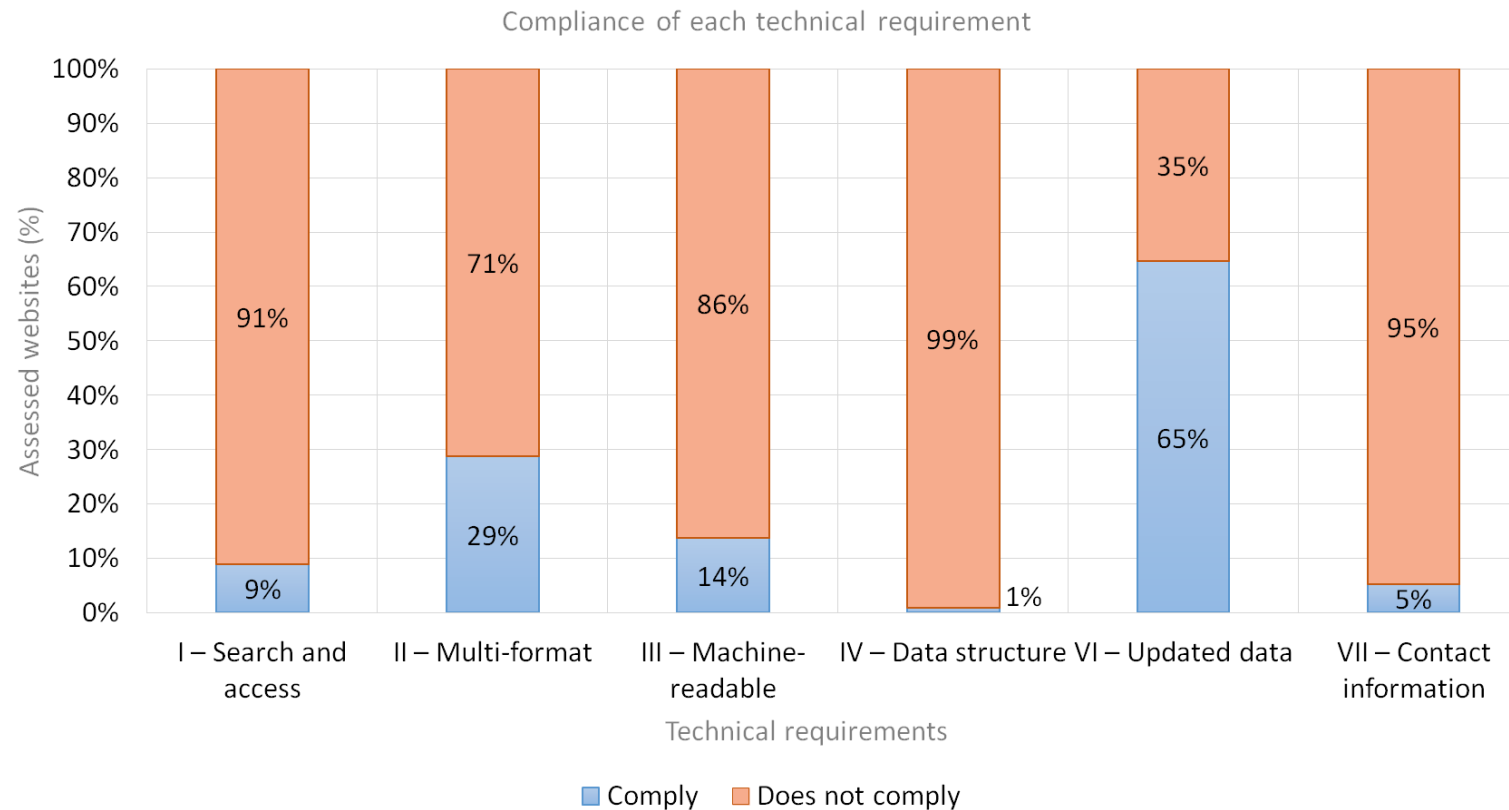
- Understand how current data portals comply with open government data (OGD)
- Focus on local governments (city halls)
- Assessed 561 municipalities
- Findings showed the vast majority of assessed data portals did not comply with OGD

# Data openness in Brazilian local governments



Andreiwid Sh. Corrêa, Evandro Couto de Paula, Pedro Luiz Pizzigatti Corrêa, Flávio Soares Corrêa da Silva, (2017) "Transparency and open government data: A wide national assessment of data openness in Brazilian local governments", Transforming Government: People, Process and Policy, Vol. 11 Issue: 1, pp.58-78, doi: 10.1108/TG-12-2015-0052

# Data openness in Brazilian local governments

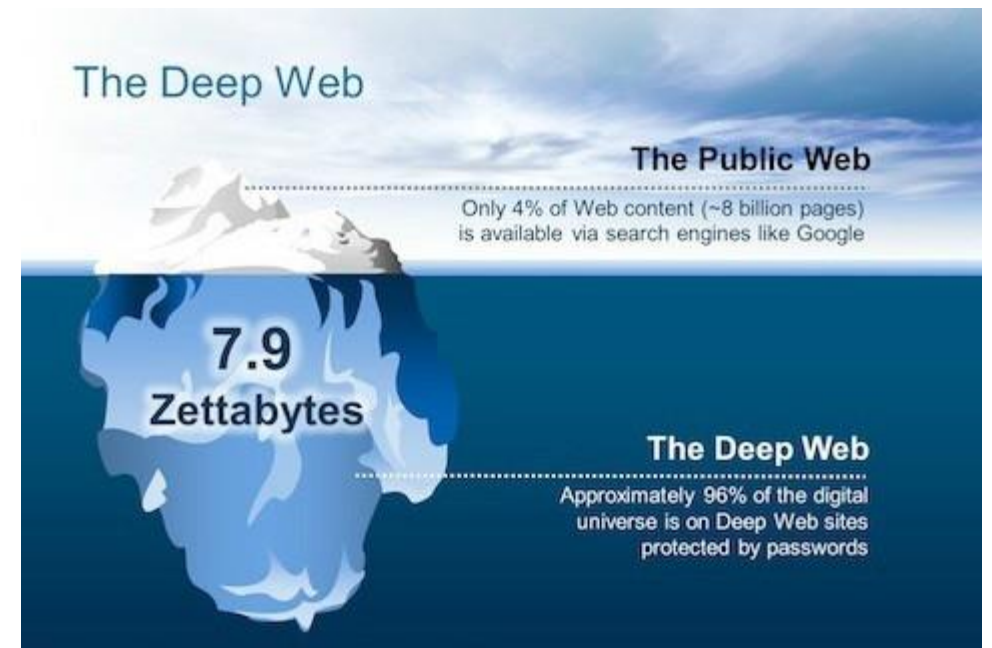


Andreiwid Sh. Corrêa, Evandro Couto de Paula, Pedro Luiz Pizzigatti Corrêa, Flávio Soares Corrêa da Silva, (2017) "Transparency and open government data: A wide national assessment of data openness in Brazilian local governments", Transforming Government: People, Process and Policy, Vol. 11 Issue: 1, pp.58-78, doi: 10.1108/TG-12-2015-0052

# Current research

---

- Aims at assessing government data portals according to OGD
- Look into the Deep Governmental Web
- Dynamic-generated data (HTML)
  - Data behind HTML forms and parameters
- Preliminary tests in the 27 Brazilian capitals
  - Crawled ~20 millions of links
  - Currently processing potential deep web pages (with certain kinds of HTML forms)



(images from <https://www.deepweb-sites.com/>)

# Interesting sources of information

---

- Open Data Charter (2015)
  - <https://opendatacharter.net>
- Open Data Institute (2012)
  - <https://theodi.org>
- Open Data for Development (2010)
  - <http://od4d.net>



# Thank you!

---

[andreiwid@ifsp.edu.br](mailto:andreiwid@ifsp.edu.br)

<http://andreiwid.info>

# Bibliography

---

<https://carto.com/blog/forty-brilliant-open-data-projects-preparing-smart-cities-2018/>

<http://busturnaround.nyc/>

<http://www.blindsquare.com/>

<https://smartnet.niua.org/sites/default/files/resources/Open-Data-Guide-8-24-2015.pdf>

<https://opengovdata.org/>

[https://www.pwc.ru/en/government-and-public-sector/assets/ddc\\_eng.pdf](https://www.pwc.ru/en/government-and-public-sector/assets/ddc_eng.pdf)

[http://cdn.cs50.net/2016/fall/seminars/smart\\_cities/smart\\_cities.pdf](http://cdn.cs50.net/2016/fall/seminars/smart_cities/smart_cities.pdf)

<https://www.hitachivantara.com/en-us/pdf/solution-profile/hitachi-solution-profile-city-data-exchange.pdf>

<http://routetopa.eu/deliverable-2-1-state-of-the-art-report-and-evaluation-of-existing-open-data-platforms-now-available/>