



Image/Map Credit: Stamen Design

OPEN DATA: WHICH AND HOW

@abhinemani | CAPP 30130 | University of Chicago



CIVIC TECH: KEYS TO SUCCESS

Empathy

- ▶ Build with, not for
- ▶ Talk to users
- ▶ Understand behavior
- ▶ Learn what's out there

Lean

- ▶ It's more than just a technical process
- ▶ Expect failure
- ▶ Only build what's necessary
- ▶ Test, test, test

Data

- ▶ Data "lives" everywhere
- ▶ Be ready to scrape
- ▶ Open data is your friend
- ▶ ETLs are your best friend

Users

- ▶ ~~If you build it they will come~~
- ▶ Listen and learn
- ▶ Define metrics
- ▶ Find partners

Sustainability

- ▶ Consider your options: startup, open source, non-profit, academic, etc
- ▶ Tech -> Policy
- ▶ Expect to train everyone
- ▶ Document everything

CIVIC TECH: KEYS TO SUCCESS

Empathy

- ▶ Build with, not for
- ▶ Talk to users
- ▶ Understand behavior
- ▶ Learn what's out there

Lean

- ▶ It's more than just a technical process
- ▶ Expect failure
- ▶ Only build what's necessary
- ▶ Test, test, test

Data

- ▶ Data "lives" everywhere
- ▶ Be ready to scrape
- ▶ Open data is your friend
- ▶ ETLs are your best friend

Users

- ▶ ~~If you build it they will come~~
- ▶ Listen and learn
- ▶ Define metrics
- ▶ Find partners

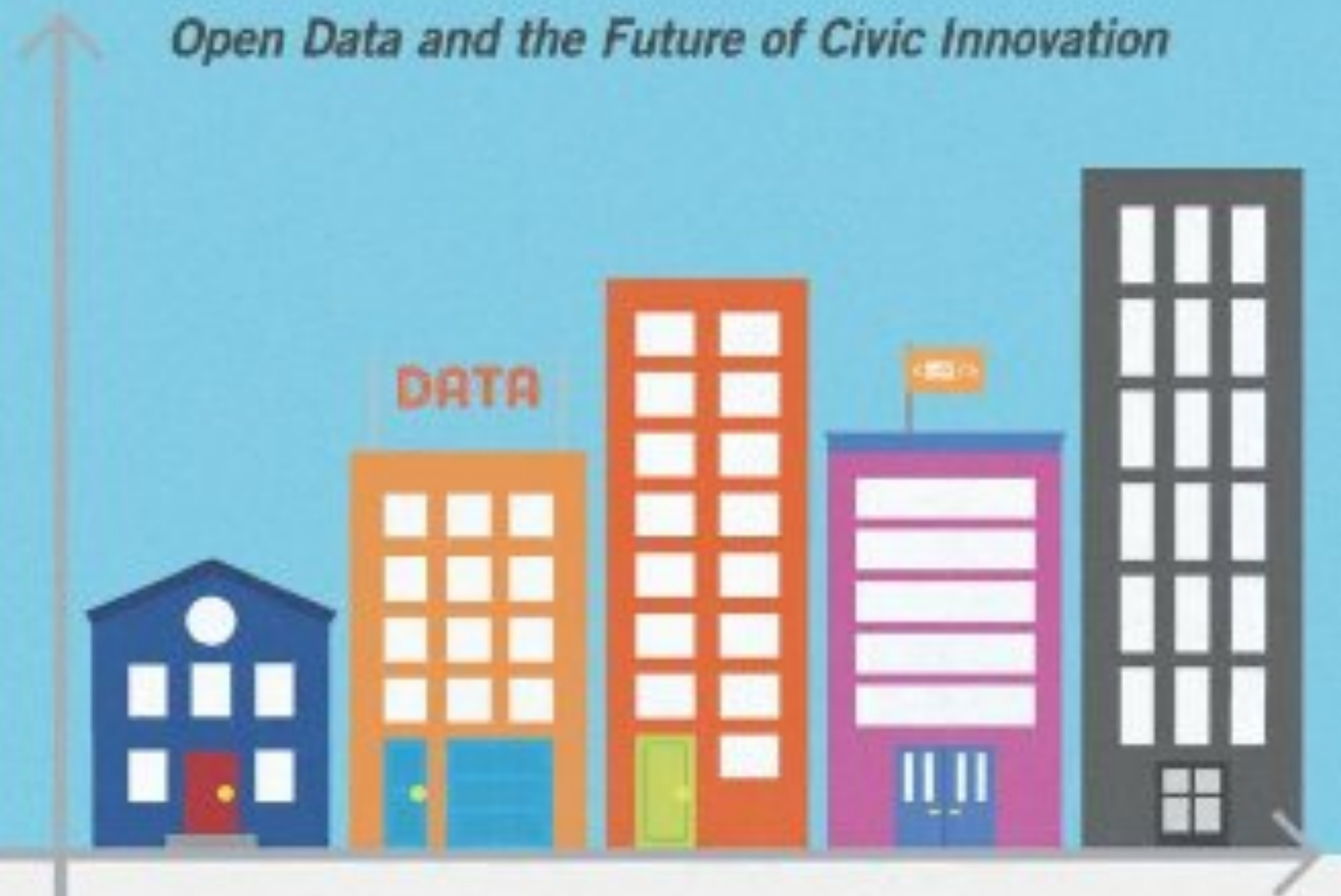
Sustainability

- ▶ Consider your options: startup, open source, non-profit, academic, etc
- ▶ Tech -> Policy
- ▶ Expect to train everyone
- ▶ Document everything

Edited by Brett Goldstein with Lauren Dyson

Beyond Transparency

Open Data and the Future of Civic Innovation



"A must-read for anyone who is passionate about what open data can do to transform city living."
—Boris Johnson, Mayor of London

AGENDA

- Case Study: Philly's Data Portal
 - Public Art
- Which open data?
 - Popular open data sets
 - Data inventories
- How open data?
 - Standard open data publishing workflow
 - Data portal comparisons
 - Challenges



CASE STUDY: OPEN DATA PHILLY

Open data as a group effort

ORIGINAL OPENDATAPHILLY.ORG

The screenshot shows the Open Data Philly website homepage. At the top right, there are links for "Register" and "Login". The main heading is "Connecting people with data." Below this is a sub-heading: "Throughout the Philadelphia region, a multitude of organizations have assembled data. Find it here and start transforming your city." A search bar with the placeholder text "Start your search." is prominently displayed. Below the search bar is a map of Philadelphia with a red dashed circle highlighting a specific area. An orange callout box over the map says "NEED INSPIRATION? VISIT THE IDEA GALLERY". Below the map, there is a tweet from @OPENDATAPHILLY dated Wednesday, April 06, 2011 3:41 PM, mentioning San Diego and refugee populations. To the right of the tweet is a button that says "GOT DATA? MAKE A SUBMISSION" with icons for mobile apps. At the bottom, there is a section for "RECENT ADDITIONS". On the left side of the page, there is a navigation menu under the heading "DATA" with categories like Arts / Culture, Economy, Education, etc.

- *Developed and maintained by community groups in 2011*
 - *Local technology company*
 - *Non-profits*
- *Featured city government and non-government data (NGOs, universities, etc)*
- *Data directory, not data warehouse*
 - *Linked to downloads/APIs*



JOIN THE OPENDATARACE

Nominate the Philly data you would like to see made available to the public!

**NOMINATE
DATA!**

“The question will be, if the OpenDataRace shows clear community support for certain data releases, will the selected city agency and partnering departments have the capacity, interest and top-level encouragement to follow through.” - [Technical.ly/Philly](https://technical.ly/Philly)

City of Philadelphia
CityOfPhiladelphia

Philadelphia, Pennsylvania
http://phila.gov
Joined on Oct 24, 2012

23 public repos 5 members

- cityofphiladelphia.github.io**
Website to share the City of Philadelphia's open source projects and open data
Last updated 2 days ago
- avi**
You can use this application to estimate your real estate tax under the Actual Value Initiative (AVI).
Last updated 5 days ago
- myphillyrising**
A mobile web app for the City of Philadelphia's PhillyRising Collaborative.
Last updated 15 days ago
- phi-polling-loctions**
City of Philadelphia Polling Locations
Last updated 15 days ago
- flu-shot-spec**
Proposed Data Specification for Flu Shot Locations
Last updated 15 days ago
- phi-airport-data**
A REST and WebSocket API for PHL Flight Information Data
Last updated 19 days ago

README.md

GitHub For Geodata

Vision

This repo will contain information on City of Philadelphia publishing open geodata on GitHub.

Why GitHub?

GitHub is an ideal platform for releasing geodata to the public because it:

- gives GIS units control of when and how their data is updated to the general public
- allows other GitHub users to ask questions and raise issues about the data - creating a dialog around it
- displays a history of changes to the data that gives consumers notifications on updates
- allows for collaboration between staff and outside stakeholders
- Renders [CSV files as tables](#) and [GeoJSON as embeddable web maps](#)

*The City of Philadelphia experimented with various approaches to publishing its own open data. First, they used **GitHub** for hosting its data (left and top right), which was cheap/free and included useful community features (discussions, etc). Then they eventually moving to **Socrata** — which enabled better data visualization and an “easier” UI for non-technical residents.*

Note: This developer portal will be replaced soon - it's gonna be pretty awesome. [Stay tuned...](#)

PHL API

A collection of open data APIs from the City of Philadelphia. These APIs are being made available to participants in civic hackathons and many other events to provide easy access to useful civic data. Build an app or solution to make your city better!

[More Open Data »](#)

The city was able to iteratively open up access to data, building APIs, exposing data, based on citizen requests/input. Above: the now deprecated, [PHLAPI.com](#), which featured (and served) data, documentation, and developer tools.

Airport Parking Availability API

API endpoint

This simple read-only API exposes parking garage availability data (updated every few minutes) for the Philadelphia International Airport.

Data returned for each garage includes:

- Lat/Long coordinates of the garage
- Total number of parking spaces
- Total number of ADA (handicapped) spaces
- Current number of available spaces

API Endpoint

<https://api.phila.gov/airport-parking/v1/>

Sample Response:

```
{
  "garages": [
    {
      "lat": "39.8775100347",
      "lng": "-75.2472191649",
      "total_spaces": "3489",
      "available_spaces": "3489"
    }
  ]
}
```

The city was able to iteratively open up access to data, building APIs, exposing data, based on citizen requests/input. Above: the now deprecated, PHLAPI.com, which featured (and served) data, documentation, and developer tools.

Helper Libraries

Getting started using one of the City of Philadelphia APIs is easy - you can use one of these helper libraries, or even write your own. If you'd like to contribute or enhance any of the libraries listed here, make sure to fork the source code on GitHub and send a pull request to the project sponsor.

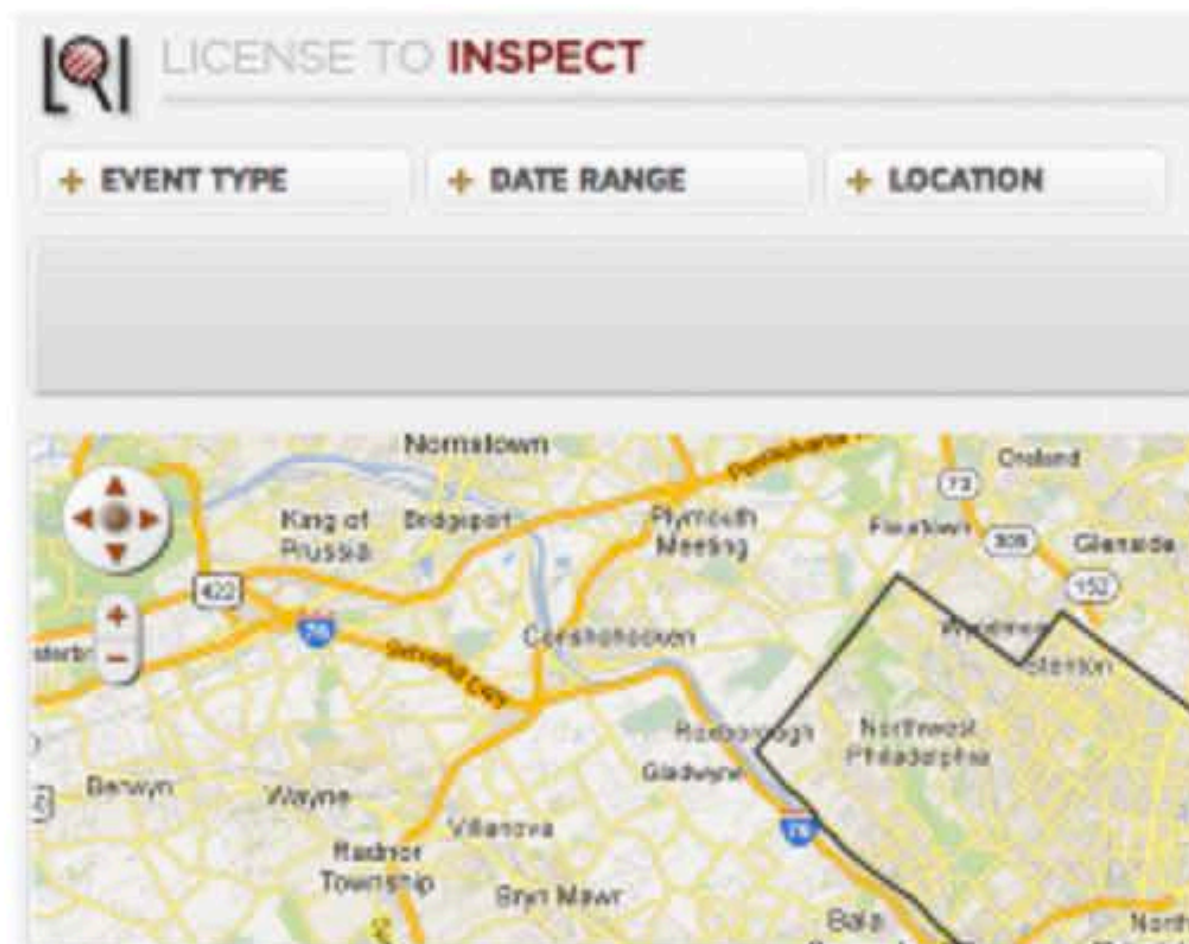
Name:	Language:	Description:
phl-geocode	Node.js	Get latitude and longitude coordinates for a Philadelphia Address.
phl-pac-complaints	Node.js	A module for working with City of Philadelphia's Police Advisory Commission Complaints Data. (Node.js)
phl-poll-finder	Node.js	Find the polling location for a Philadelphia voter. (Node.js)
phli	Node.js	A module for working with City of Philadelphia's Licenses & Inspections Data.
py-li	Python	A thin wrapper for the City of Philadelphia L&I API.
open311	Node.js	A module for interacting with an Open311 API.
open311_php	PHP	Open311 library updated for GeoReport v2.
open311_python	Python	A module for interacting with an Open311 API.

.....

The city was able to iteratively open up access to data, building APIs, exposing data, based on citizen requests/input. Above: the now deprecated, [PHLAPI.com](#), which featured (and served) data, documentation, and developer tools.

Application Showcase

These apps use the [Licenses & Inspection API](#).



L&I Property History

Enter an address to see permits, licenses, violations and appeals from the Department of Licenses & Inspections.

PHLI

Download Philadelphia Licenses & Inspections Data

Filter Options

Council Districts

- 1
- 2
- 3
- 4
- 5
- 6
- 7
- 8
- 9
- 10

Zipcodes

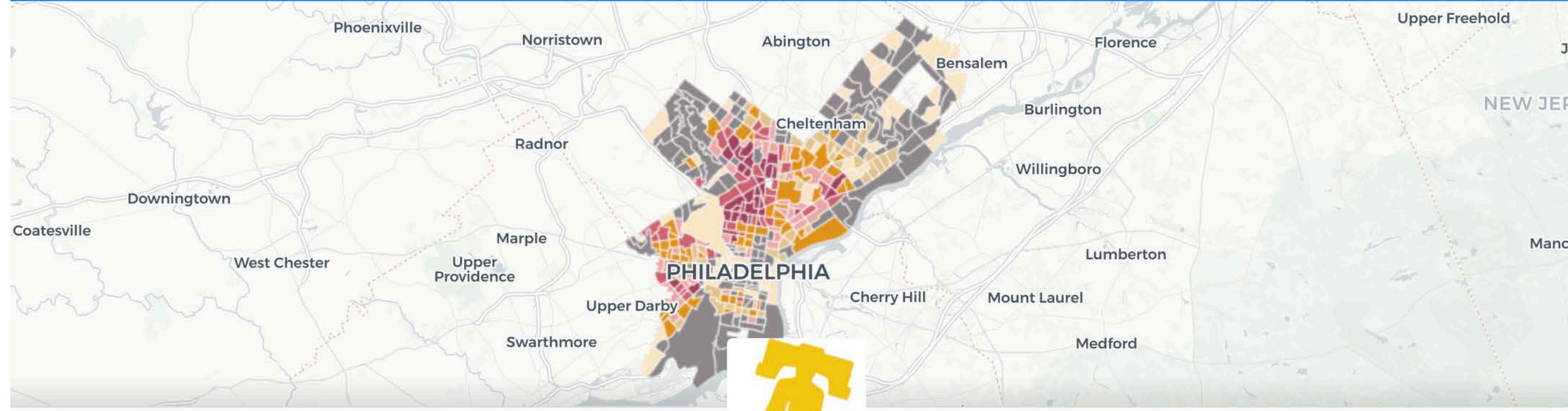
- 19003
- 19004
- 19006
- 19012
- 19020
- 19023
- 19027
- 19033
- 19035
- 19038
- 19046

Wards

- 1
- 2
- 3
- 4
- 5
- 6
- 7
- 8
- 9
- 10
- 11

Permits Licenses Cases

The city was able to iteratively open up access to data, building APIs, exposing data, based on citizen requests/input. Above: the now deprecated, [PHLAPI.com](#), which featured (and served) data, documentation, and developer tools.

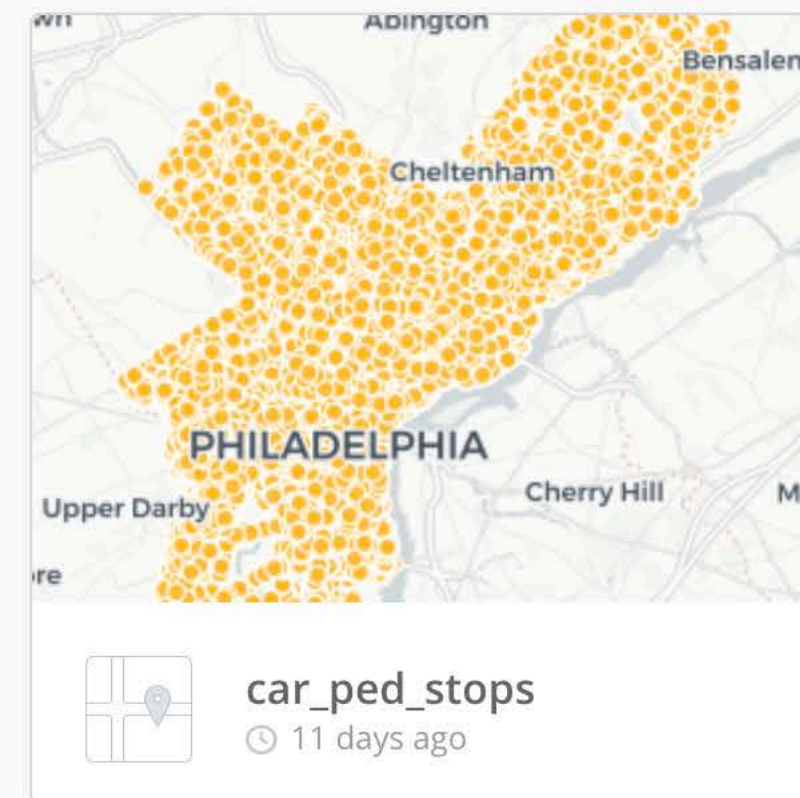
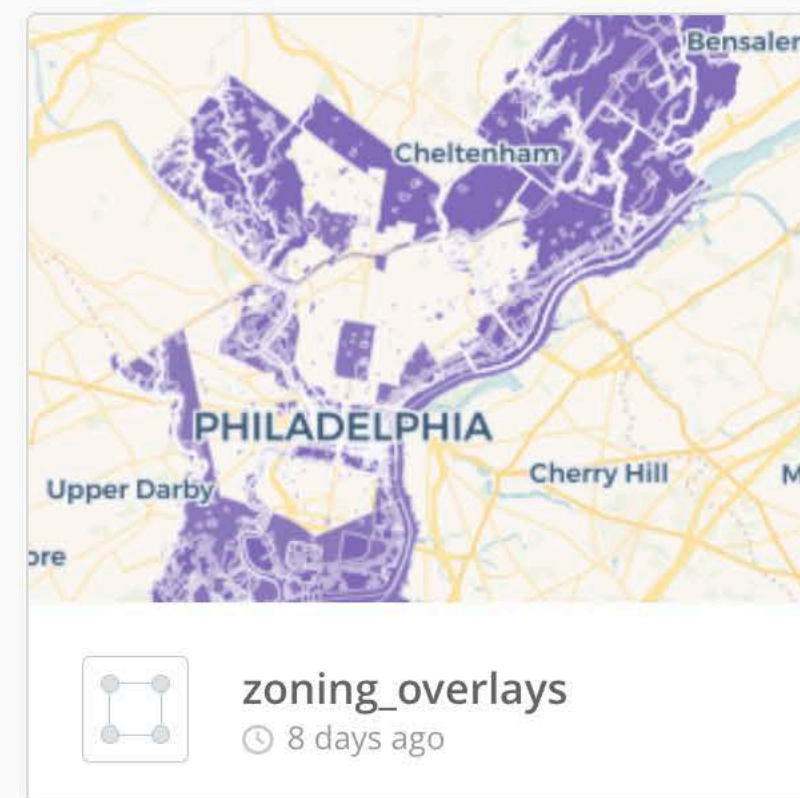
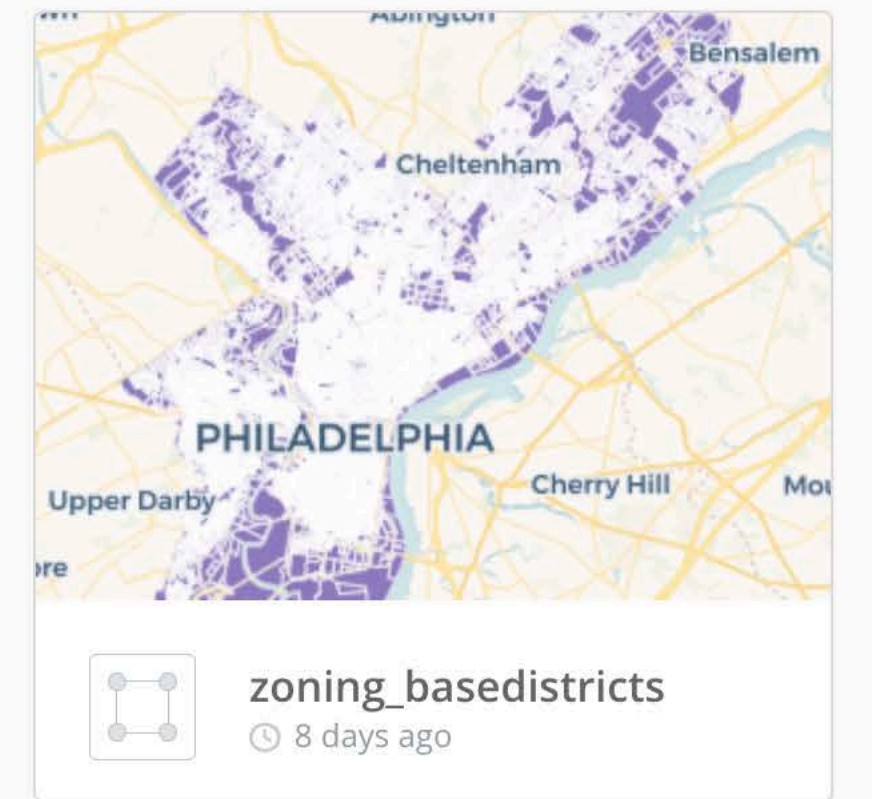
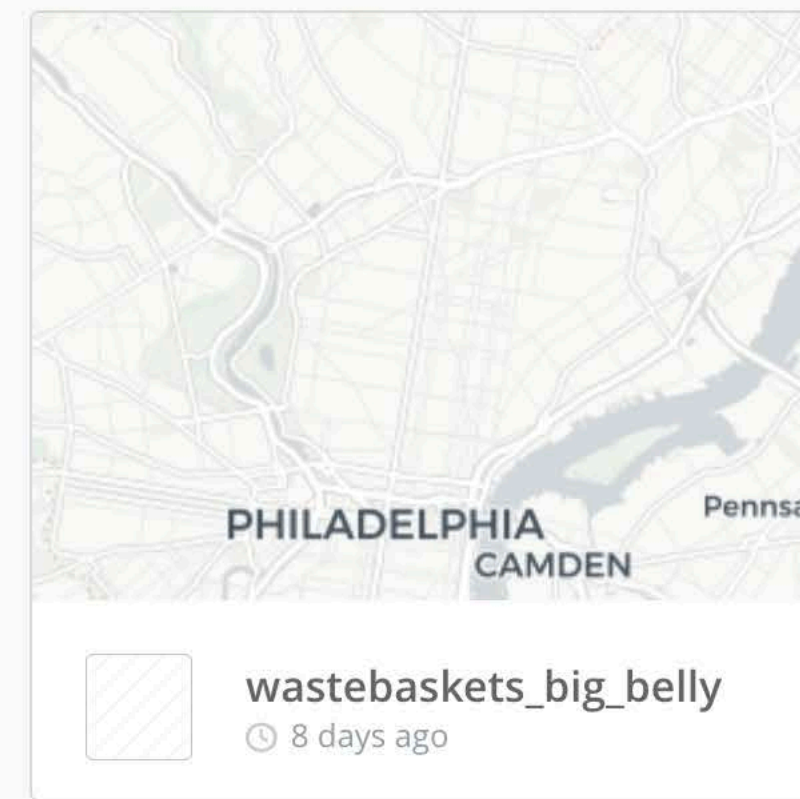


cityofphiladelphia.carto.com

City of Philadelphia / Datasets ▾

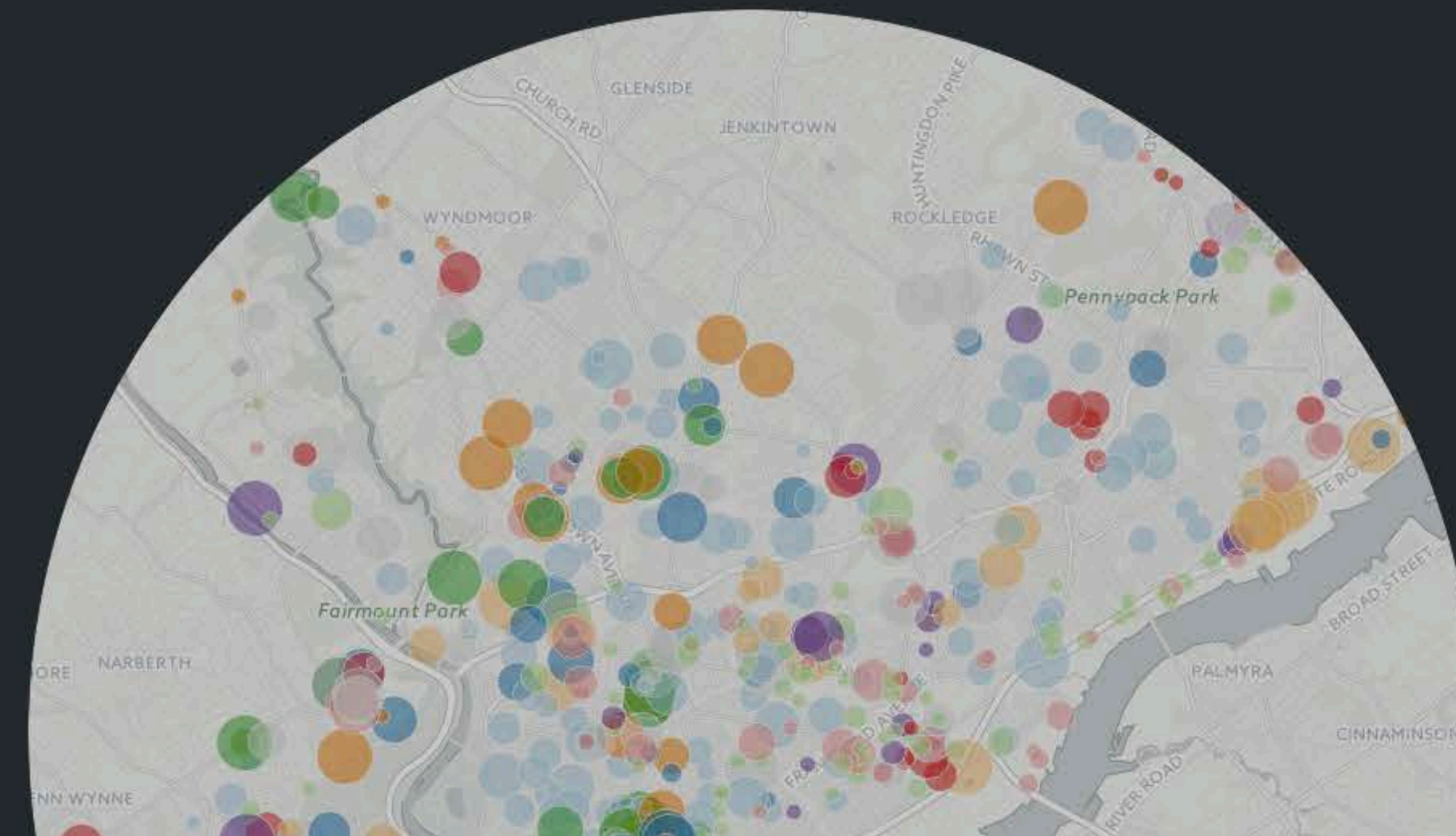
beta.phila.gov [PHLIInnovation](#)

*City of Philadelphia finally decided to host its open data on a popular consumer data platform, **Carto** (formerly CartoDB), which is optimized for **spatial data** (it's PostgreSQL), comes with a **dashboard generating tool**, and provides a **robust API**.*



Your source for open data in the Philadelphia region

OpenDataPhilly is a catalog of open data in the Philadelphia region. In addition to being the official open data repository for the City, it includes data sets from many organizations in the region.



In 2017, the city, in partnership with the same community groups that launched the original OpenDataPhilly, released an updated version — powered by the open source tool, CKAN — with 300+ datasets, some from the city's Carto warehouse, some from its various APIs (eg PHLApi) and some from community groups / engaged citizens.

Search data

Get started by searching from **360** datasets that cover all dimensions of Philadelphia, from city council districts to park locations.

“

While many open data portals include only municipal government data, OpenDataPhilly also incorporates data from non-profits, universities, and commercial organizations in order to give a more complete picture of the data available on Philadelphia.

-OpenDataPhilly



[Azavea](#) is an award-winning geospatial analysis (GIS) software development firm specializing in the creation of location-based web and mobile solutions, as well as geospatial analysis services to enhance decision-making. Azavea is a certified B Corporation that applies geographic data and technology to promote the emergence of more dynamic, vibrant, and sustainable communities.



The City of Philadelphia [Office of Innovation & Technology](#)'s mission is to increase the effectiveness of the information technology infrastructure, where the services provided are advanced, optimized, and responsive to the needs of the City of Philadelphia's businesses, residents and visitors.



[Temple University School of Media and Communication](#) is a leader in advancing the role of communication in public life. This mission is pursued by educating students for leadership careers in communication and media, through advancing research and creative activity in these fields, and by serving the public need for free and open communication. The school's [Center for Public Interest Journalism](#) has been managing the OpenDataPhilly web site since the re-launch in early 2015.



[Technical.ly Philly](#) is a news site covering the community of people who use technology in Philadelphia. They also operate sister media sites in Baltimore, Brooklyn, Delaware, and Washington DC.



@abhinemani

PUBLIC ART

A case study in community data collection & app development



Philadelphia Public Art API

Followers
0

Organization



Philart.net

This site is a collection of the webmaster's pictures of 935 sculptures, fountains, mosaics and memorials in Philadelphia. <http://philart.net/> [read more](#)

- Dataset**
- Topics
- Activity Stream
- Related

Philadelphia Public Art API

A read-only, RESTful JSON API. The core function of the API is to return descriptions of pieces of public art in Philadelphia Documentation includes a JavaScript tester with visible source code. There are two primary types of data returned by this API: those that exist purely for navigation and those that return descriptions of art. The JSON format for the artworks is consistent in all responses. Some responses describe collections of art. Some collection types include extra information specific to the collection itself. Mostly conforms to a HATEOAS model of navigation and the use of the "links":{"rel":"X","href":"Y"} structure. Also supports a geolocation call which returns a collection that is defined by bb and ll URL arguments

Data and Resources



API Info

Documentation page for the Philadelphia Public Art API.

[Explore](#)



API

Top level data.

[Explore](#)

Arts

Additional Info

Field	Value
-------	-------

0 Comments

Philadelphia Public Art @philart.net

[Titles](#) [Artists](#) [Years](#) [People](#) [Landmarks](#) [Architecture](#) [Size](#) [Content](#) [Tours](#) [Exhibits](#) [All](#) [Nearby](#)
[Home](#) [Resources](#) [Technical](#) [Thanks](#) [Contact](#) [RSS](#) [API](#) Enter Search Terms [Search](#)



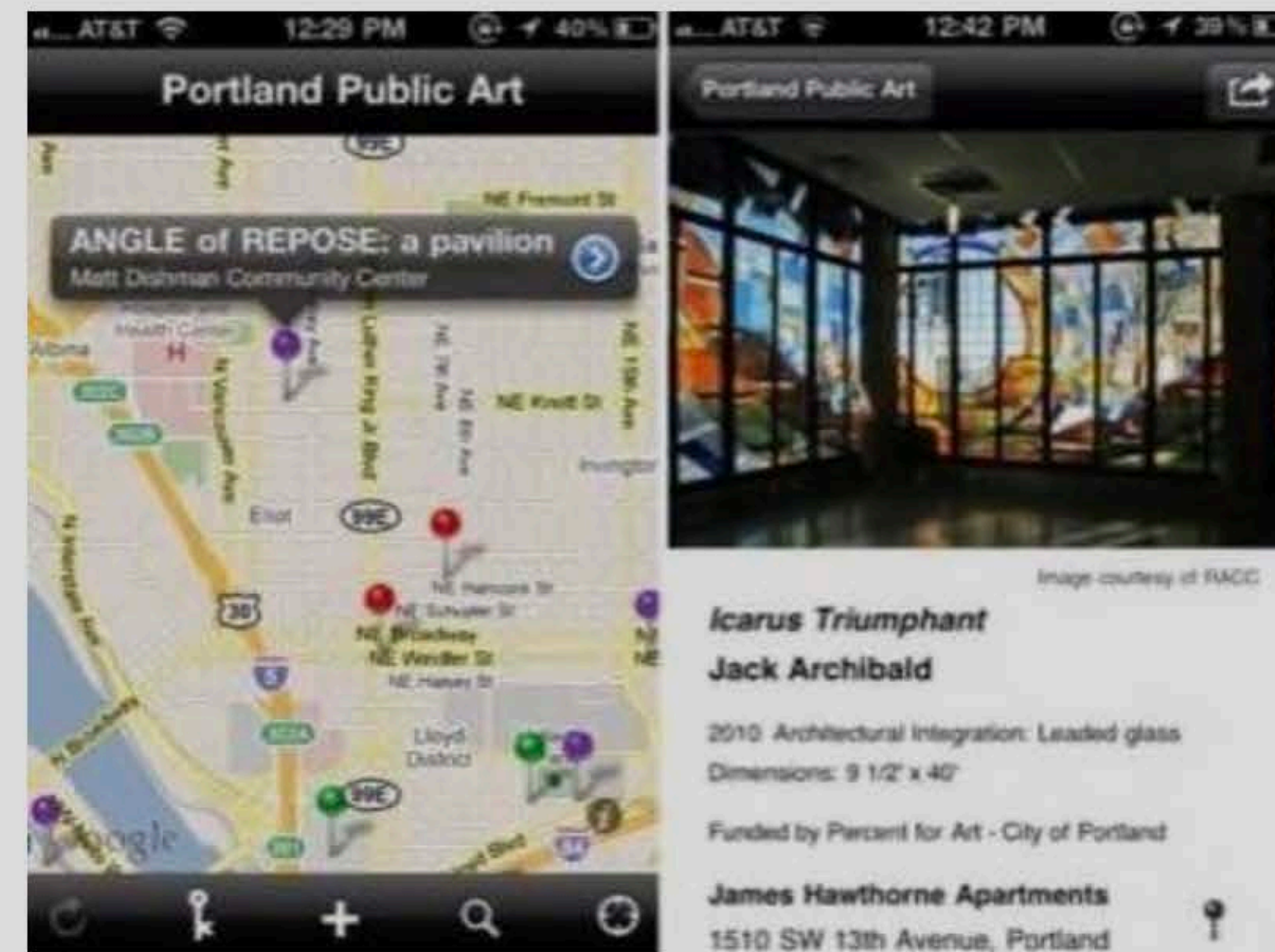
Welcome to Philadelphia Public Art @ philart.net

My personal collection of pictures of **1,075 sculptures, reliefs, mosaics, light displays, fountains and memorials** in Philadelphia, with

- indices for [titles](#), [artists](#), [years](#), [people](#), [landmarks](#), [architecture](#) and [content](#)
- 13 [exhibits](#)
- 35 [tours](#) with both list and map views
- a [nearby](#) feature with geolocation support
- thumbnails that link to the larger pictures
- locations with latitude, longitude and [map] links
- underlined internal links to related and listed entries
- underlined external links (preceded with a +) for artists, artworks, people, and landmarks
- [a read-only JSON API](#) for building your own applications or data mining

Every City Needs This: Mobile Public Art Finder

marshall kirkpatrick / 01 Mar 2011 / Mobile



Portland, Oregon is known for many things but one of them is its availability of public data. When the local government included listings of 429 pieces of public art in the data it made available to independent developers, [Matt Blair](#) took the logical and admirable step of turning that data into an iPhone app. The [Portland Public Art Finder](#) makes it easy to discover and learn about murals, sculptures, installations and other instances of art that you can go and enjoy in public.

Public art became an effective (and fun) example of community engagement, open data, and civic technology. Cities hosted contests, built their own apps, etc.

List

PUBLIC ART FINDER

Find Me



Public art became an effective (and fun) example of community engagement, open data, and civic technology. Cities hosted contests, built their own apps, etc.

@abhinemani

Open Art Tours

Sign in and discover what fabulous public art lies right around the corner!

Sign In

Tours

SELECT STARTING POINT

Race Vine Station - BSL M

Comcast Center

15th St Station - MFL M

Schmidt Dean Gallery

Gross McCleaf Gallery

Carre d'

The Print Center

Kimmel Center

START TOUR

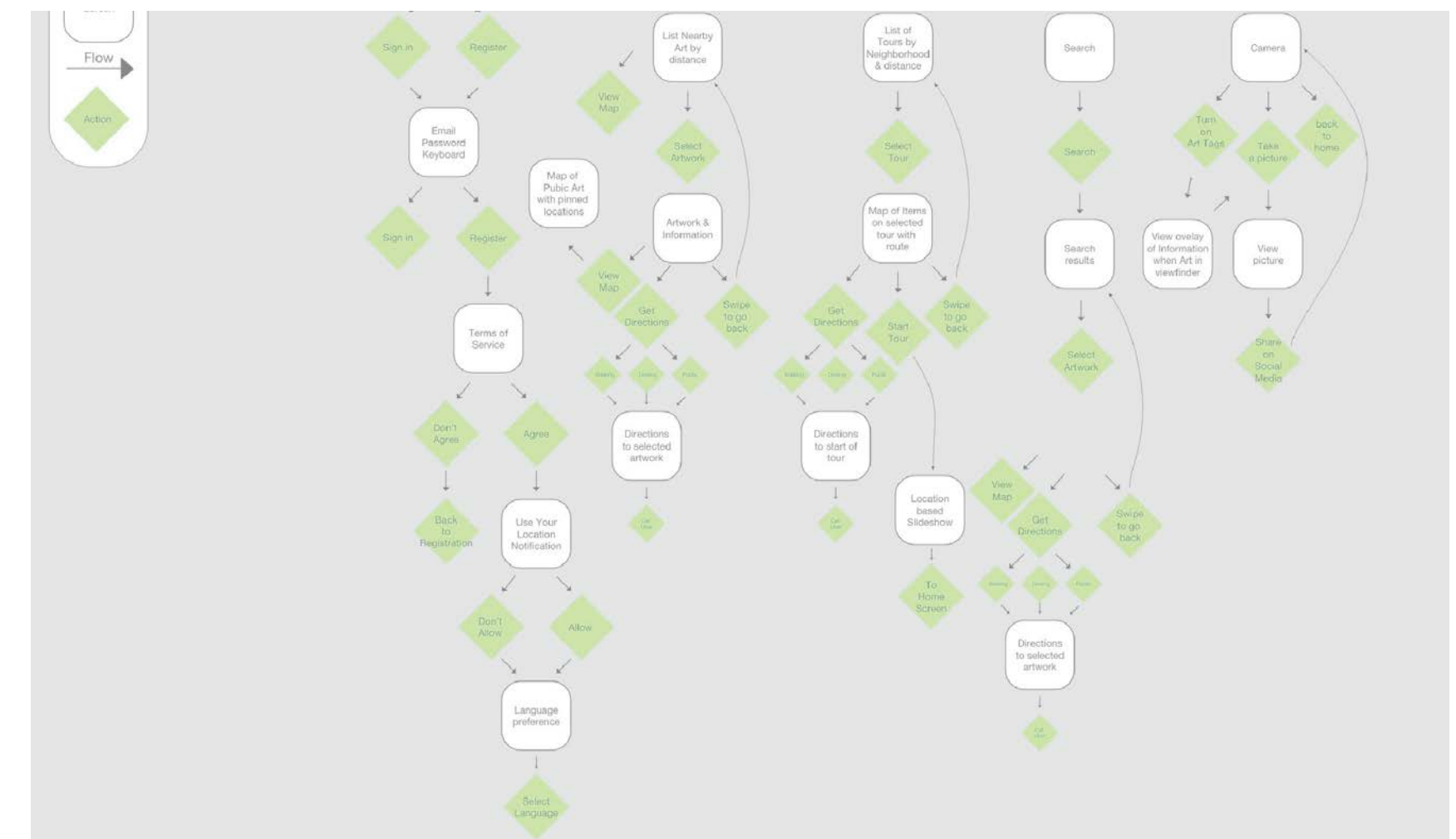
CALL UBER

<Open Data Philly/>

Case Study:

Challenge: Find a way to visualize a data set from OpenDataPhilly.com

Public Art API



@abhinemani

Public art became an effective (and fun) example of community engagement, open data, and civic technology. Cities hosted contests, built their own apps, etc. (Above: Philly Prototype)



@abhinemani

WHICH OPEN DATA?

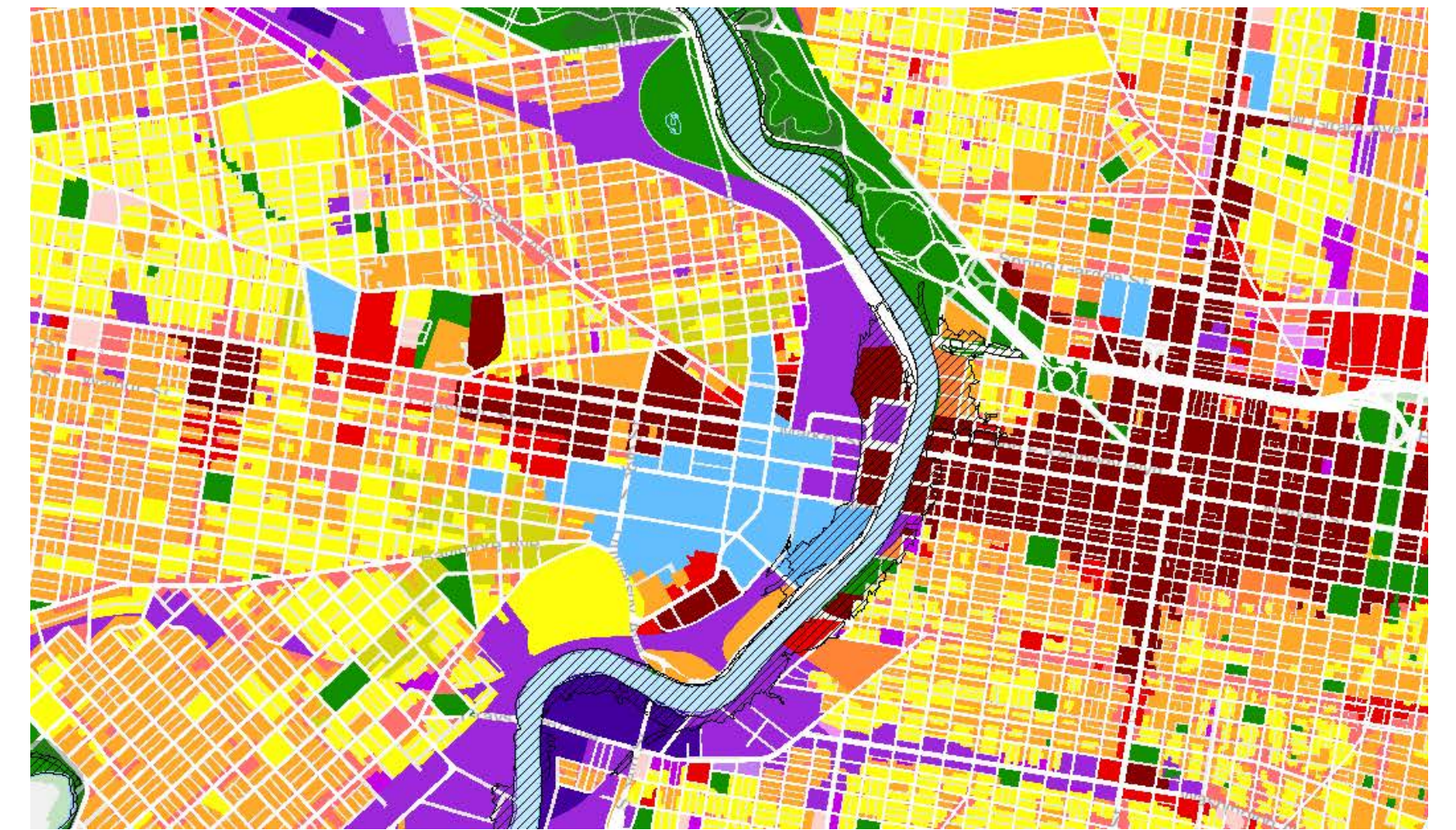
Prioritizing which open data to publish (first)





“3-B’s of Open Data: Buses, Budgets, and Bullets”

- Mark Headd



*Buildings, Bills, and Boundaries
(An additional 3 B's)*



“Gamifying open data”: The Open Data Census gives cities guidance on what datasets to prioritize and motivation to publish them well (and fast) by ranking cities for compliance.

LA RANKED #1 IN US CITY OPEN DATA CENSUS

MAY 21, 2015



Just one year after Mayor Garcetti launched LA's first-ever Open Data Portal. LA given ISO 37210 Platinum Certification for Data by World Council on City

The power of rankings: "Mayor Garcetti today announced the Los Angeles has been ranked #1 by the US City Open Data Census for its open data. The Census... ranks cities' data on the availability, quality, and accessibility of its open data. Just one year after Mayor Garcetti launched the City's first-ever open data portal, LA tops the list, beating New York, San Francisco, Boston, and Philadelphia."

DataSF In Progress

Plan Progress

Our open data legislation requires 1) an inventory of City datasets and 2) publishing plans from each department. March 31, 2015 was the target date for all departments to complete the dataset inventory. We accept department submissions on a rolling monthly basis. Publishing plans were due contingent on inventory completion and will be updated every six months.

52 Departments

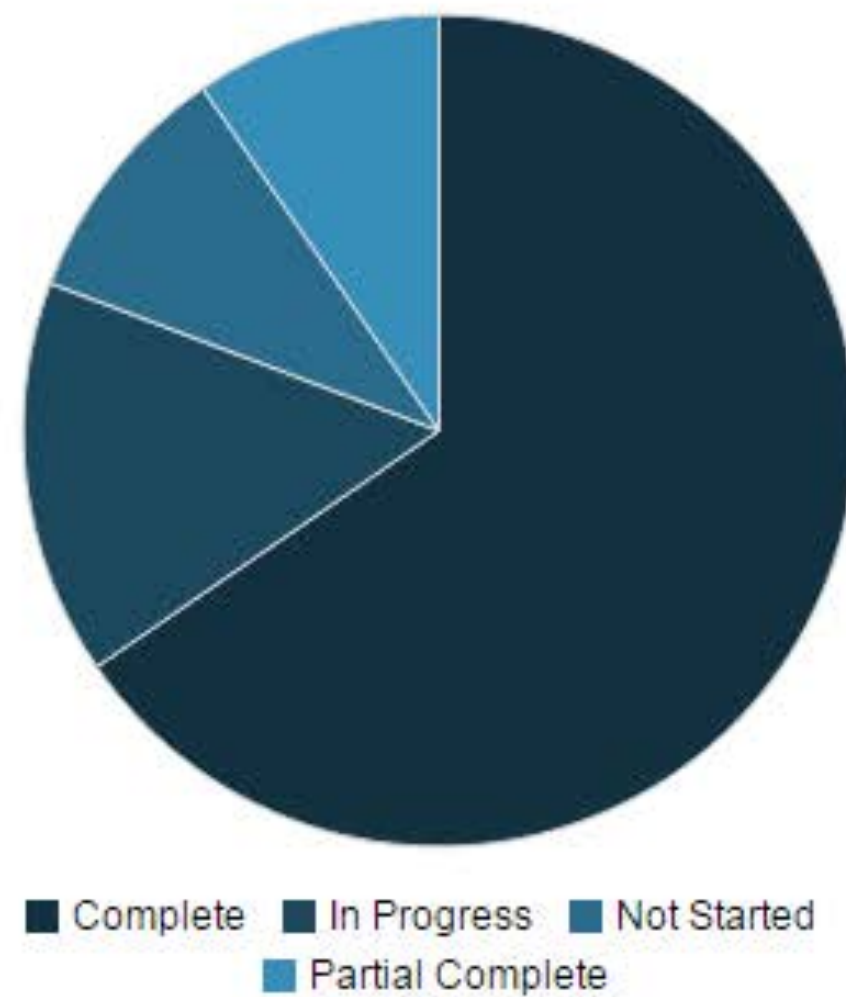
34 Inventories Complete

25 Plans Complete

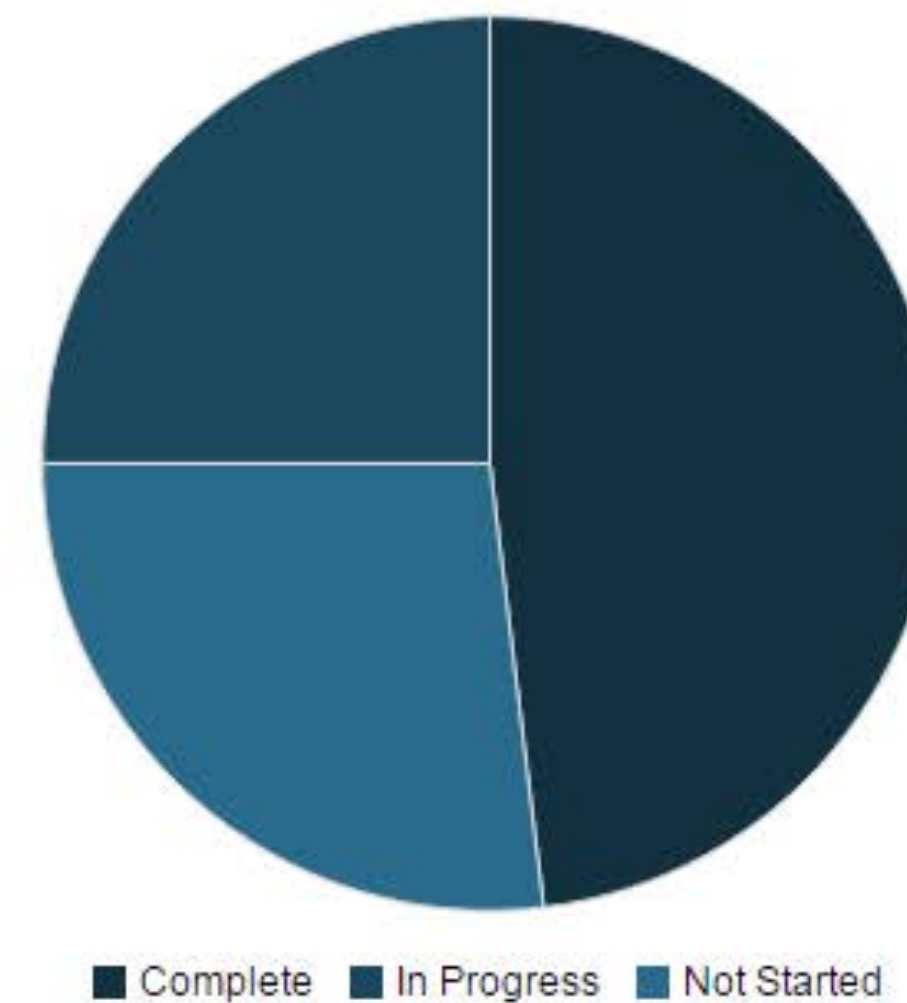
[Read the Publishing Plans](#)

[View Progress by Department](#)

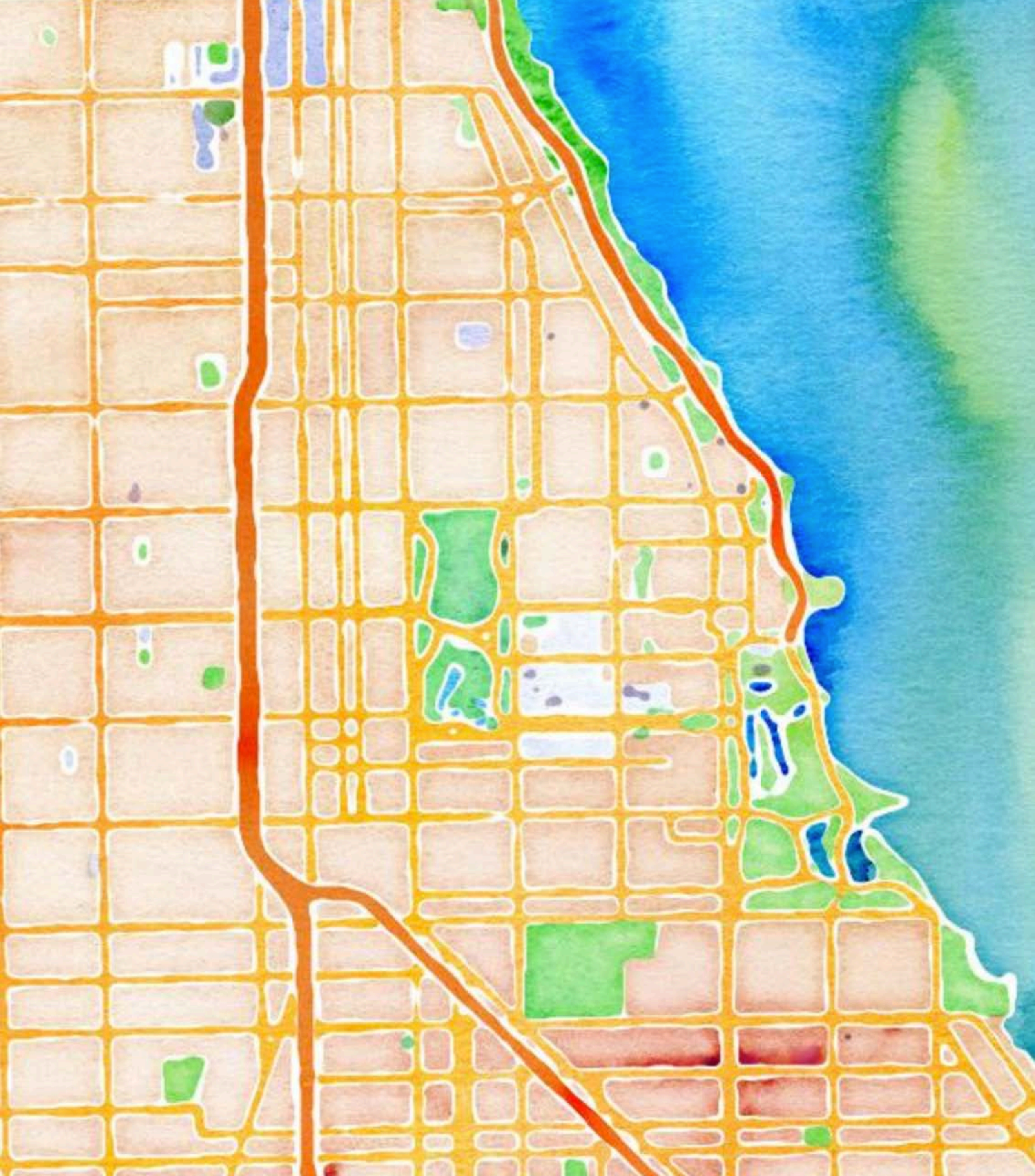
Inventory status



Publishing plan status



Data Inventories: Governments will often take a systematic approach to publishing their data, going department-to-department inventorying all the applications in use, and then the data coming into and the data going out of them. Then, based on priorities, access, and utility, they will publish data — ideally tracking their own performance as a data organization, like DataSF (above).



DATA DISCOVERY

- What information systems does your department use?
- What databases does your department use?
- What applications capture information or are used in your business processes?
- Are some data resources kept in spreadsheets (on shared or individual drives)?
- What information are we already publishing and where did that information come from?

PRIORITIZING DATA PUBLISHING: QUESTIONS TO ASK

➤ Internal value?:

- Is the dataset aligned with the department's core function?
- Is the dataset a direct or indirect source of a department's Headline Performance Measure(s) and/or Supporting Measure(s)?
- Will posting the data result in administrative time saved for a department(s)?
 - For instance, as departments often spend substantial time and resource complying with FOIA?
- Is the data already collected and readily accessible so as not to require new or additional business processes?
- Is the data used across departments and does its publication facilitate collaboration?
- Will publishing this dataset enable the retirement of a legacy system or database?

➤ External value?:

- Is the data typically used when deciding policies that impact residents' quality of life?
- Will publishing this dataset allow a business or resident to perform a process more efficiently and/or effectively?
- Does the data contribute to civic engagement?
- Does the data create economic opportunity?
- Will publishing this dataset make the department more accountable and/or responsive?



@abhinemani

OPEN DATA WORKFLOW

How it's supposed to work. And how it does.





Source Data(base)

TRADITIONAL OPEN DATA WORKFLOW



Source Data(base)

- *Excel files on desktops*
- *Google Spreadsheets*
- *On-Premise Databases*
- *On-Premise Applications*
- *Web Databases*
- *Web Applications*

*Data
Centralization*



Source Data(base)

TRADITIONAL OPEN DATA WORKFLOW

Data Centralization



Source Data(base)



- *File Upload*
- *On-prem to web scripts*
- *Web-to-web scripts*

*CSV Upload
ETL Script*



Datastore



Source Data(base)

TRADITIONAL OPEN DATA WORKFLOW

*Data
Centralization*



Datastore

- *Hosted Database(s)*
- *(Public) Web interface*
- *Administration/Access control*
- *API access*



Source Data(base)



DIFFERENT DATA PORTALS / SYSTEMS

- Open v Closed
 - Proprietary
 - Socrata
 - Esri
 - Opensource
 - CKAN
 - DKAN
 - JKAN
- Note: Most, if not all, support **data.json**
 - Standard data format/file listing all available data

*Data
Centralization*



Datastore

*Data
Consumption*



Source Data(base)

TRADITIONAL OPEN DATA WORKFLOW

*Data
Centralization*



Datastore

*Data
Consumption*



- *CSV Download*
- *JSON Feed*
- *GeoJSON Feed*



Source Data(base)

*Data
Centralization*



Datastore

*Data
Consumption*

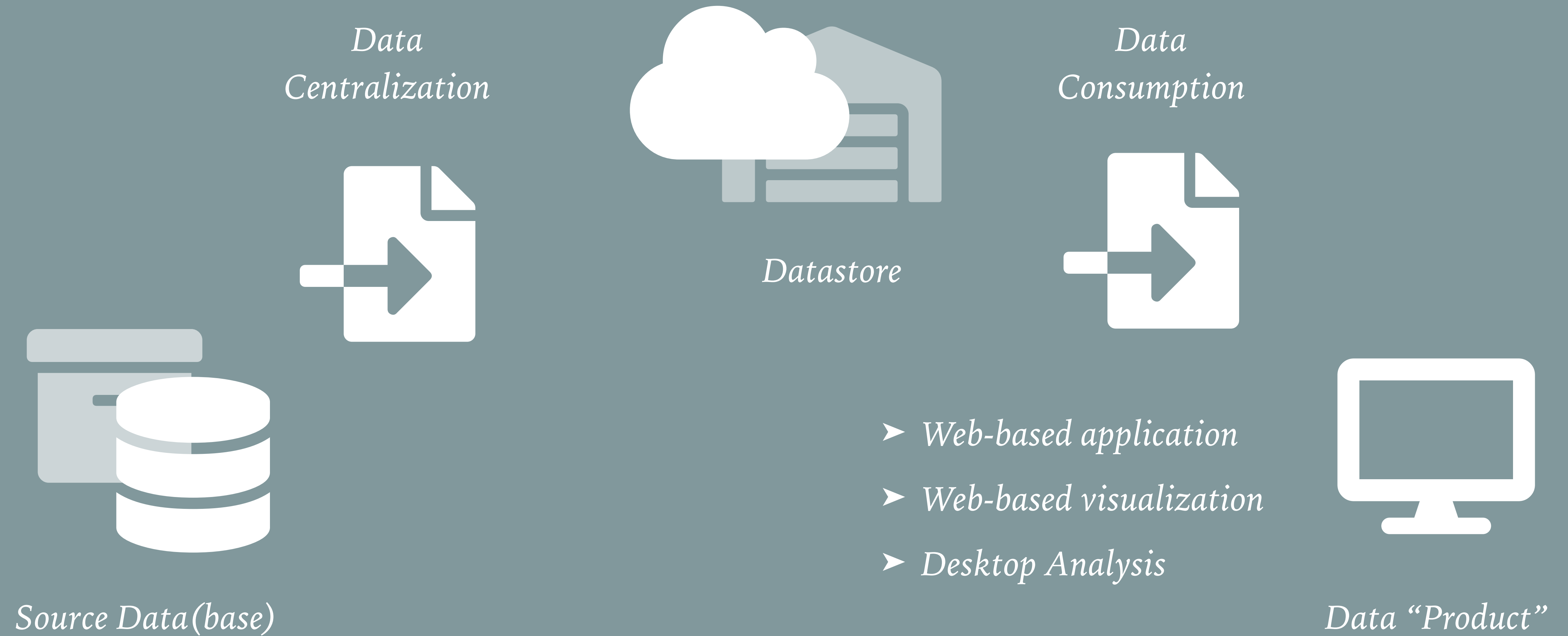


Source Data(base)

TRADITIONAL OPEN DATA WORKFLOW



Data "Product"



*Data
Centralization*



Datastore

*Data
Consumption*



Source Data(base)

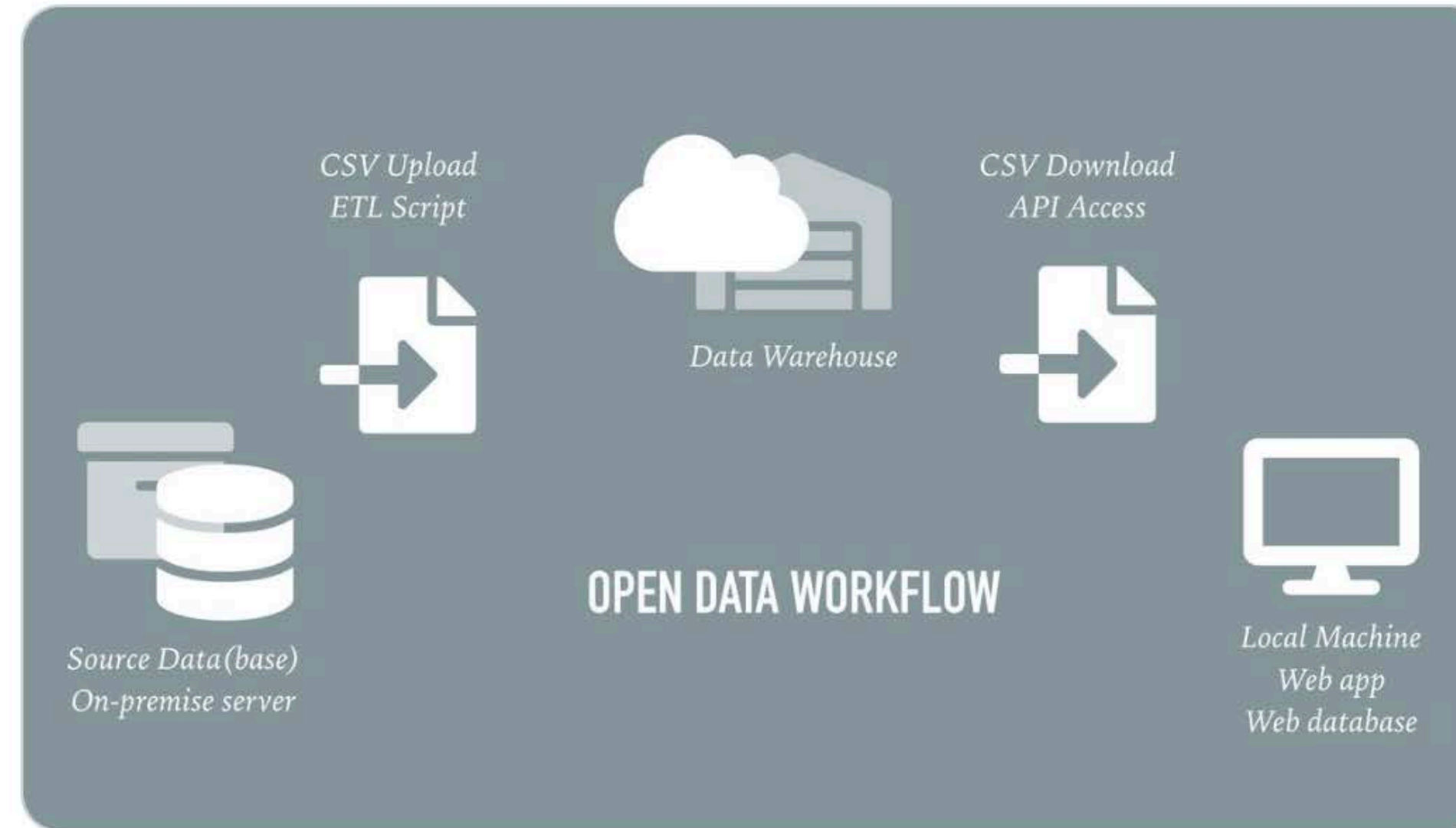
TRADITIONAL OPEN DATA WORKFLOW



Data "Product"



abhi nemani [@abhinemani](#) · 54m
Simple sketch of [#opendata](#) workflow



3 2 2



Michael Schnuerle [@LouDataOfficer](#) · 21m
That's Louisville, except the data warehouse is an S3 bucket.

1 1



Michael Schnuerle
[@LouDataOfficer](#)

Replying to [@LouDataOfficer](#) and [@abhinemani](#)

And the first ETL are broken SSIS 2008 scripts. And the second ETL are shell scripts on the open data site. Each managed by just one different person. We are working on a better solution...

**THE REALITY OF OPEN DATA IS
MUCH, MUCH MORE
COMPLICATED.**



COMPLICATIONS IN OPEN DATA

- **Source data** access often requires creation of DMZs (Demilitarized Zones)
 - Cloud hosting costs money, procurements, etc
- ETLs for **data centralization** typically require technical, need to be maintained, and are dataset specific
- Different types of data might require different **data warehouses** (eg spatial)
- Effective **data consumption** demands good documentation and metadata
- Governments cannot guarantee the accuracy/fairness of data in user-generated **data products**

LECTURE 8 ASSIGNMENT (ASSIGNED 4/24)

➤ Readings

- *Pioneering open data standards* by Bibiana McHugh
- *Data and design* by Cyd Harrell
- *Why good hackers make good citizens* by Catherine Bracy
- *Small (city) pieces, loosely joined*

➤ Project work

- Finish JKAN customization
 - Submission details on Canvas
 - Requirements: 1) Local instance; 2) Web deployment; 3) Custom.css; 4) 3 datasets, categorized and detailed