March, 11, 2021 - Open Data Week

Hi, I’m Devin Balkind

- A software producer who organizes teams to deliver solutions for startups, nonprofits and government agencies.
  - Sarapis: nonprofit
  - Pruvn: business
  - BetaNYC: contractor for city agencies (past)
- I was born & raised in NYC.
- Plan to do WeGov work for 30+ years.

@devinbalkind on Twitter

devin@sarapis.org

Let’s Make NYC the World’s Best Run City

New York City is literally the world’s “favorite” city.

We’re the richest city in the richest country. We have the resources to be the best.

We aren’t. And that’s bad for us and bad for the world.

To billion of people around the world, NYC represents the concept of urban life, a global city, a diverse “melting-pot”, immigration and opportunity.

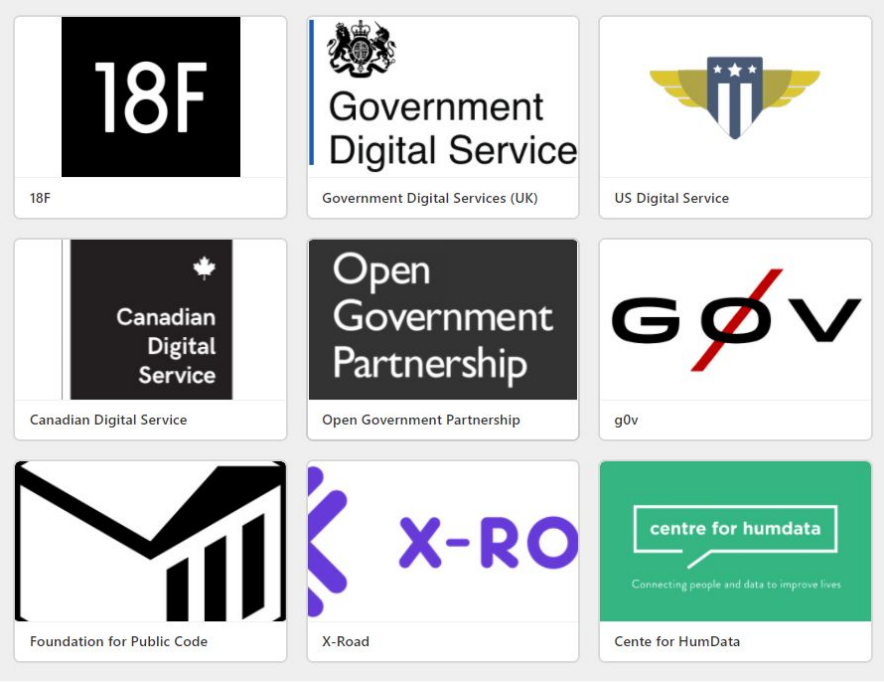
We must do better.

What Does Better Look Like?

Over the last decade, people have figured out how to go government tech better.

New York City doesn't have a general purpose digital service organization applying best practices in digital transformation.

MODA does great work but it's scope is limited.



How do we help?

Phase 1: Tools

We create and support open source software & data projects.

Phase 2: Community

We offer news, knowledge and networking opportunities to like minded people.

Phase 3: Advocacy

We build public awareness and support for best practices and solutions.

Tools

We help technology projects that making NYC more open, effective & participatory.

We've build almost a dozen experimental apps: a City Agency Directory, Capital Project Explorer, City Services API, Laws on Your Phone, Political Directory & Data Maps.

The DataBook is our first real offering.

We also offer resources to projects that align with our values, such as Reported App.

And we provide long-term maintenance for projects, like Capital Projects Directory.

Resources for Reported App

“Reported” helps New Yorkers report dangerous drivers via an information collection app that sends reports to 311, TLC, Twitter and more.

Volunteer developers got stuck and couldn’t get to 3.0 version and the app was broken for half its users.

We gave them $3,000 worth of a Sarapis developer’s time to get 3.0 launched and are now helping them raise money to keep that developer onboard indefinitely.

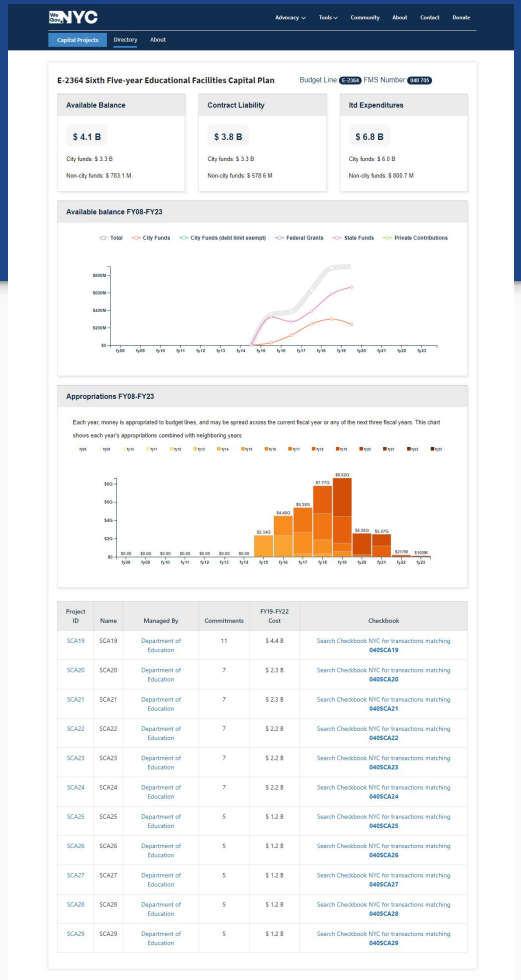
Maintenance for Capital Projects Directory

Chris Whong famously scrapped the capital projects PDFs in 2017 and put up a site.

With his blessing, we forked the site, restyled it and added some features.

This allowed Chris to focus on newer projects, like his newer geocoded capital projects map.

We'll help ensure that is maintain long term too.



Open for Business

Know of tool that aligns with our vision?

Tell us about it.

Maybe we can help!

Plus an insight: we ask too much of volunteer software developers! They can't be expected to launch and maintain a project indefinitely. Better is for them to start and structure projects, do as much work as they want, and then bring in a paid (outsourced) software developer to do the annoying stuff, keep things maintained and the project rolling. We love to fund that!

The WeGov.**NYC Databook**

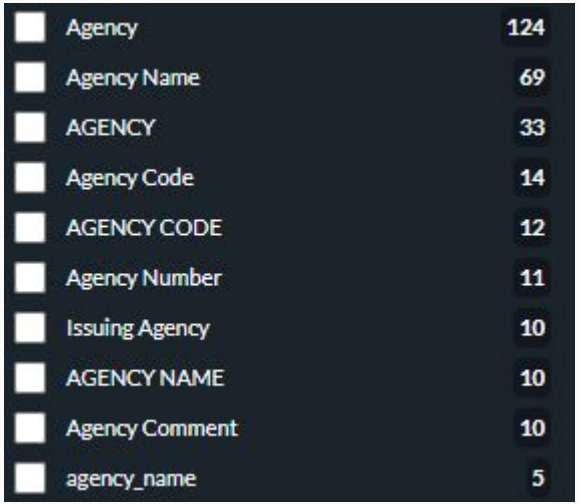
Open Data is Good

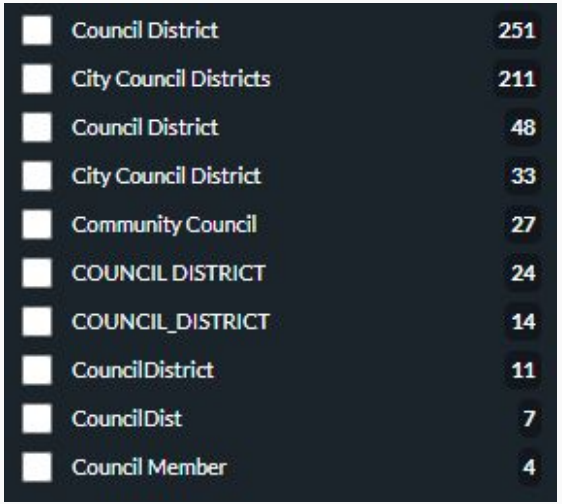
- One of NYC's only technology success stories is NYC Open Data.
- Congrats to MODA for this. It's a significant and genuine achievement.
- Great data tables, published regularly, with useful metadata. Awesome.

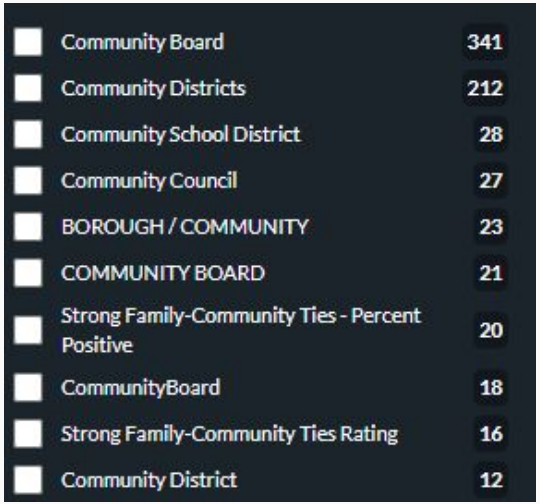
Normalizing the data will unlock its true potential to power more useful and usable public app.

What to Normalize?

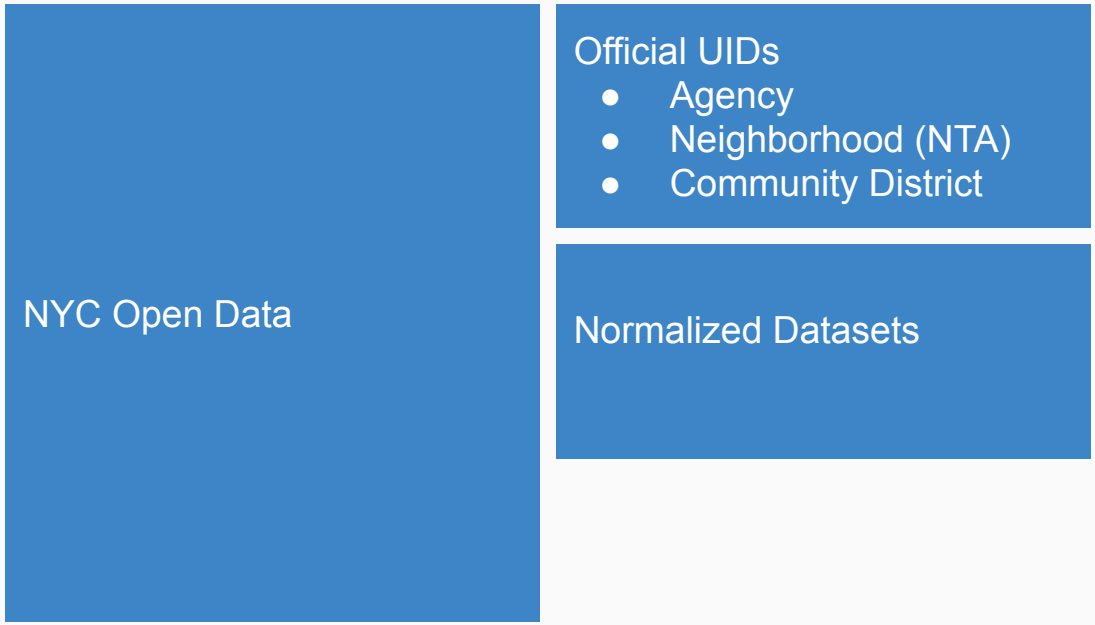
- Data Clinic's Scout Tool helps us answer this question.







General Proposal



NYC should:

- Establish UID indexes
- Build a data pipeline that applies those UIDs to open data sets.
- Republish those normalized datasets

We've Done a Little Bit of It

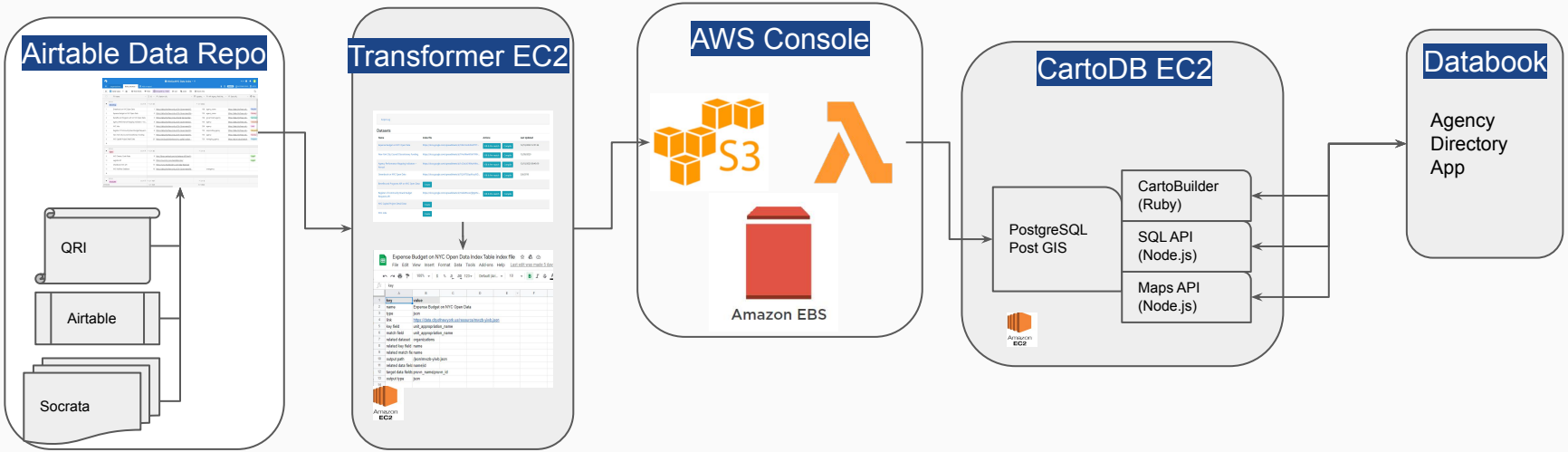
Find data open data sets, scrapers, data entry.

Normalize the data by adding consistent UIDs over all the datasets.

Archive the data in S3 so it's easy to share in bulk and in AWS environment.

Upload data into Carto so its accessible via the Carto SQL API.

Develop apps powered by the SQL API.

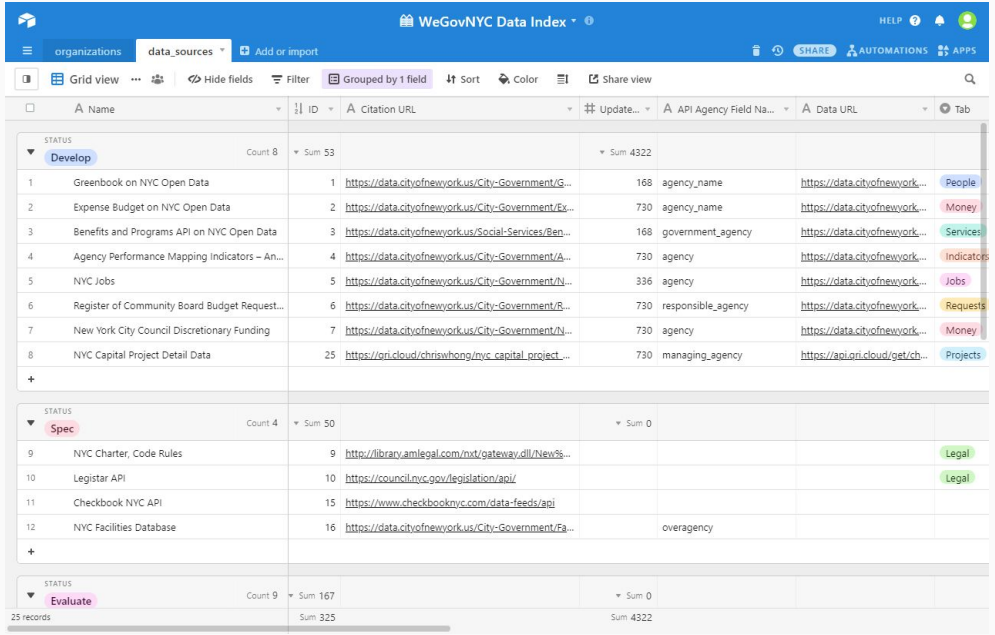


How the Transformer Works

1. Log Datasets in Airtable

Datasets are logged in the [WeGovNYC Data Index Airtable]'s data_sources table. Fields:

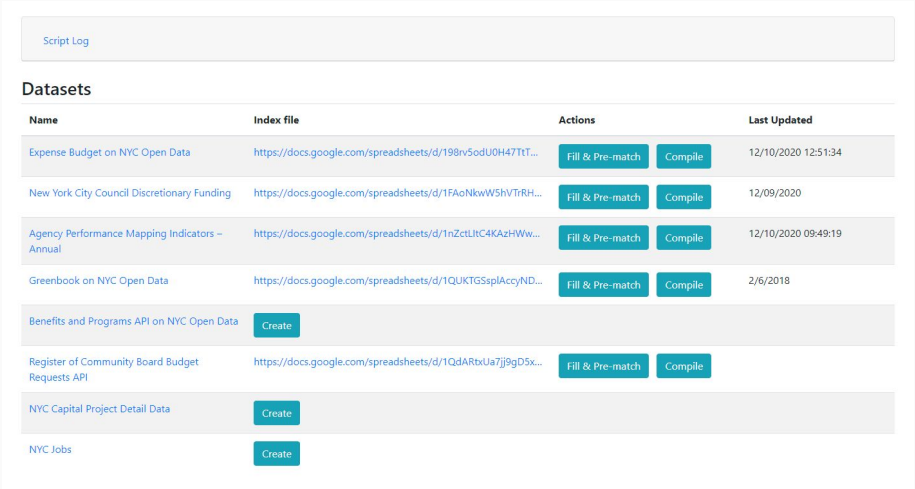
- Name: Title of the dataset
- ID: Internal identifier
- Citation URL: Where dataset links to.
- Update Schedule: Frequency of data update in hours.
- API Agency Field Name: name of field in the dataset containing a city agency.
- Data URL: Link to JSON data API
- Description: Text to appear in apps describing dataset.
- Source: Location of dataset
- Status: What to do with it
- Index Table ID: ?
- Last Updated: When the data was last transformed.
- Index Filling Data: When data was last indexed.
- Output URL: Location of transformed data.



2. Create, Fill & Prematch in Transformer App

The Transformer pulls data sets from the Data Index and allows the user to:

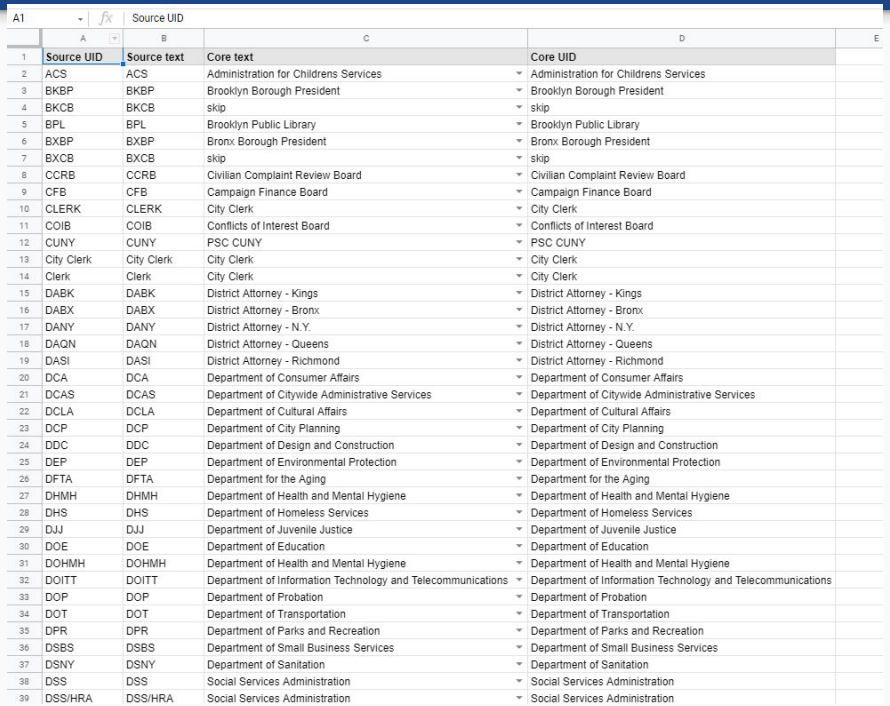
- **Create** an “Index file” Google Sheets for the file.
- **Fill & Pre-match** the file. This pulls the Airtable Agency Index, matches it to the agencies in the sheet, and allows the user to manually change those matches.
- **Compile** then applies the matched fields to the new dataset and publishes it at the URL accessible via the “Last Updated” field.



3. Match in Google Sheet

A Google Sheets is generated with three sheets:

- Configuration Info from Airtable>Data_Sources
- Organization List from Airtable>Organizations, which is the canonical list of organizations used for matching.
- **Dataset Match**, which displays a list of all agencies in the dataset and an editable fuzzy match with all our official organizations.

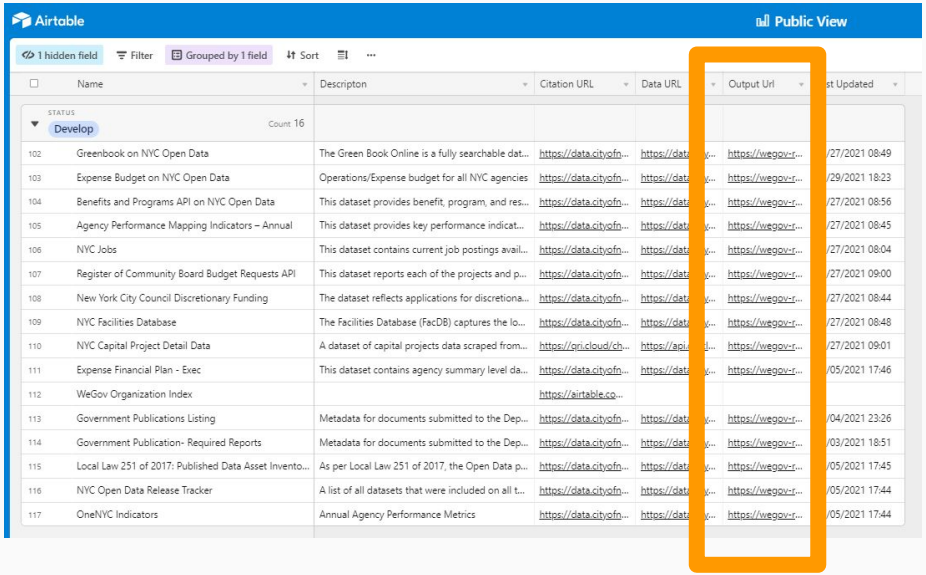


A1 | Source UID

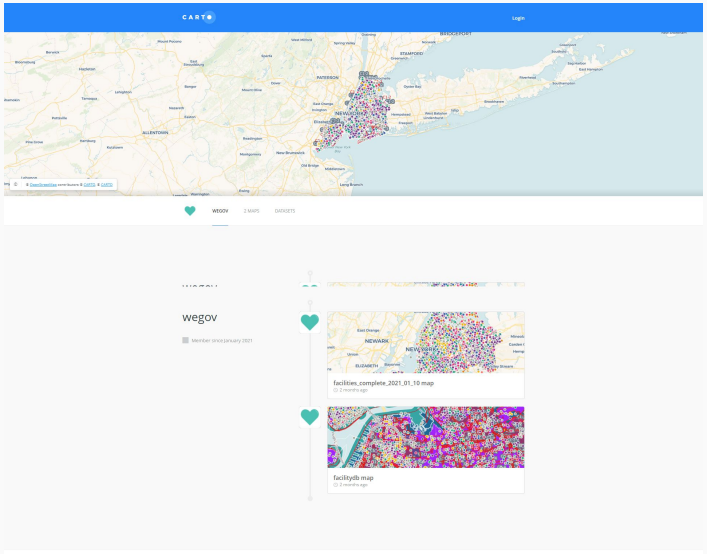
	A	B	C	D
1	Source UID	Source text	Core text	Core UID
2	ACS	ACS	Administration for Childrens Services	Administration for Childrens Services
3	BKBP	BKBP	Brooklyn Borough President	Brooklyn Borough President
4	BKCB	BKCB	skip	skip
5	BPL	BPL	Brooklyn Public Library	Brooklyn Public Library
6	BXBP	BXBP	Bronx Borough President	Bronx Borough President
7	BXCB	BXCB	skip	skip
8	CCRB	CCRB	Civilian Complaint Review Board	Civilian Complaint Review Board
9	CFB	CFB	Campaign Finance Board	Campaign Finance Board
10	CLERK	CLERK	City Clerk	City Clerk
11	COIB	COIB	Conflicts of Interest Board	Conflicts of Interest Board
12	CUNY	CUNY	PSC CUNY	PSC CUNY
13	City Clerk	City Clerk	City Clerk	City Clerk
14	Clerk	Clerk	City Clerk	City Clerk
15	DABK	DABK	District Attorney - Kings	District Attorney - Kings
16	DABX	DABX	District Attorney - Bronx	District Attorney - Bronx
17	DANY	DANY	District Attorney - N.Y.	District Attorney - N.Y.
18	DAQN	DAQN	District Attorney - Queens	District Attorney - Queens
19	DASI	DASI	District Attorney - Richmond	District Attorney - Richmond
20	DCA	DCA	Department of Consumer Affairs	Department of Consumer Affairs
21	DCAS	DCAS	Department of Citywide Administrative Services	Department of Citywide Administrative Services
22	DCLA	DCLA	Department of Cultural Affairs	Department of Cultural Affairs
23	DCP	DCP	Department of City Planning	Department of City Planning
24	DDC	DDC	Department of Design and Construction	Department of Design and Construction
25	DEP	DEP	Department of Environmental Protection	Department of Environmental Protection
26	DFTA	DFTA	Department for the Aging	Department for the Aging
27	DHMH	DHMH	Department of Health and Mental Hygiene	Department of Health and Mental Hygiene
28	DHS	DHS	Department of Homeless Services	Department of Homeless Services
29	DJJ	DJJ	Department of Juvenile Justice	Department of Juvenile Justice
30	DOE	DOE	Department of Education	Department of Education
31	DOHMH	DOHMH	Department of Health and Mental Hygiene	Department of Health and Mental Hygiene
32	DOITT	DOITT	Department of Information Technology and Telecommunications	Department of Information Technology and Telecommunications
33	DOP	DOP	Department of Probation	Department of Probation
34	DOT	DOT	Department of Transportation	Department of Transportation
35	DPR	DPR	Department of Parks and Recreation	Department of Parks and Recreation
36	DSBS	DSBS	Department of Small Business Services	Department of Small Business Services
37	DSNY	DSNY	Department of Sanitation	Department of Sanitation
38	DSS	DSS	Social Services Administration	Social Services Administration
39	DSS/HRA	DSS/HRA	Social Services Administration	Social Services Administration

4. Publish to S3 and Carto

S3 data in our public data repo Airtable as “Output URL”



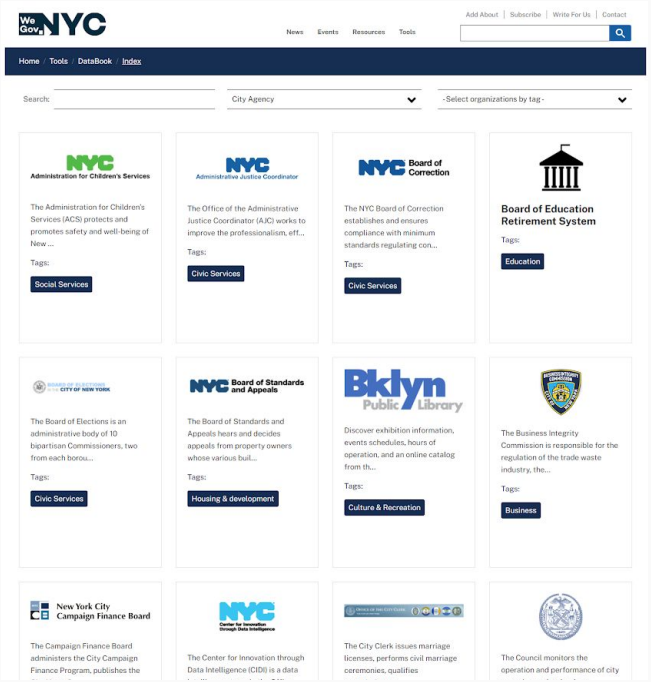
Lamda’d into carto.wegov.nyc



5. Displayed in the Databook

Our first display is a City Agency Directory, with the following datasets for each:

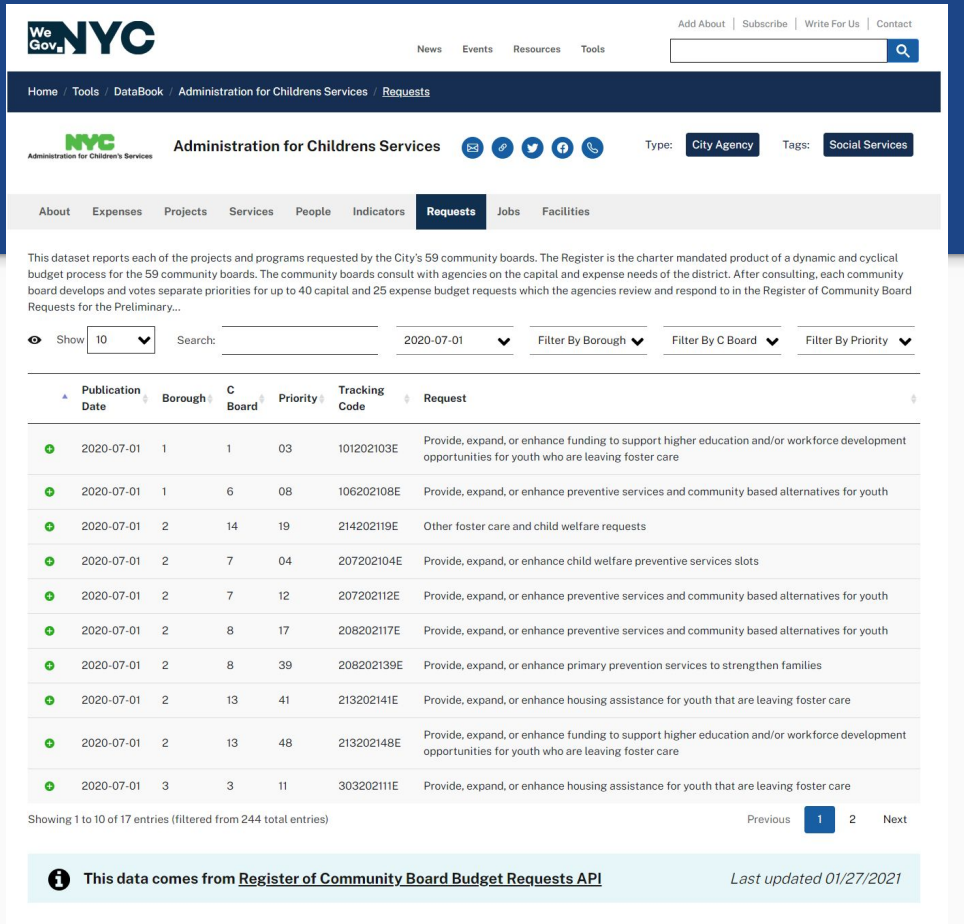
- People - Greenbook on NYC Open Data
- Projects - NYC Capital Project Detail Data
- Services - Benefits and Programs API on NYC Open Data
- Requests - Register of Community Board Budget Requests API
- Indicators - Agency Performance Mapping Indicators – Annual
- Finances - Expense Budget on NYC Open Data
- Jobs - NYC Jobs
- Facilities - NYC Facilities Database
- Organizations - WeGov Organization Index



Marginally Very Cool

Timing is Everything:

- 15-30 minutes to normalize a dataset, no coding involved.
- 1 hour to add a new dataset using very popular Laravel and DataTables.net tools.
- Design follows USDS Federal Design System
- Datasets are updated automatically. If an agency name changes or isn't matched, we get an alert to update the matching table.



We Gov. NYC

Add About | Subscribe | Write For Us | Contact

News Events Resources Tools

Home / Tools / DataBook / Administration for Childrens Services / Requests

NYC Administration for Children's Services

Administration for Childrens Services

Type: City Agency Tags: Social Services

About Expenses Projects Services People Indicators **Requests** Jobs Facilities

This dataset reports each of the projects and programs requested by the City's 59 community boards. The Register is the charter mandated product of a dynamic and cyclical budget process for the 59 community boards. The community boards consult with agencies on the capital and expense needs of the district. After consulting, each community board develops and votes separate priorities for up to 40 capital and 25 expense budget requests which the agencies review and respond to in the Register of Community Board Requests for the Preliminary...

Show 10 Search: 2020-07-01 Filter By Borough Filter By C Board Filter By Priority

Publication Date	Borough	C Board	Priority	Tracking Code	Request
2020-07-01	1	1	03	101202103E	Provide, expand, or enhance funding to support higher education and/or workforce development opportunities for youth who are leaving foster care
2020-07-01	1	6	08	106202108E	Provide, expand, or enhance preventive services and community based alternatives for youth
2020-07-01	2	14	19	214202119E	Other foster care and child welfare requests
2020-07-01	2	7	04	207202104E	Provide, expand, or enhance child welfare preventive services slots
2020-07-01	2	7	12	207202112E	Provide, expand, or enhance preventive services and community based alternatives for youth
2020-07-01	2	8	17	208202117E	Provide, expand, or enhance preventive services and community based alternatives for youth
2020-07-01	2	8	39	208202139E	Provide, expand, or enhance primary prevention services to strengthen families
2020-07-01	2	13	41	213202141E	Provide, expand, or enhance housing assistance for youth that are leaving foster care
2020-07-01	2	13	48	213202148E	Provide, expand, or enhance funding to support higher education and/or workforce development opportunities for youth who are leaving foster care
2020-07-01	3	3	11	303202111E	Provide, expand, or enhance housing assistance for youth that are leaving foster care

Showing 1 to 10 of 17 entries (filtered from 244 total entries) Previous 1 2 Next

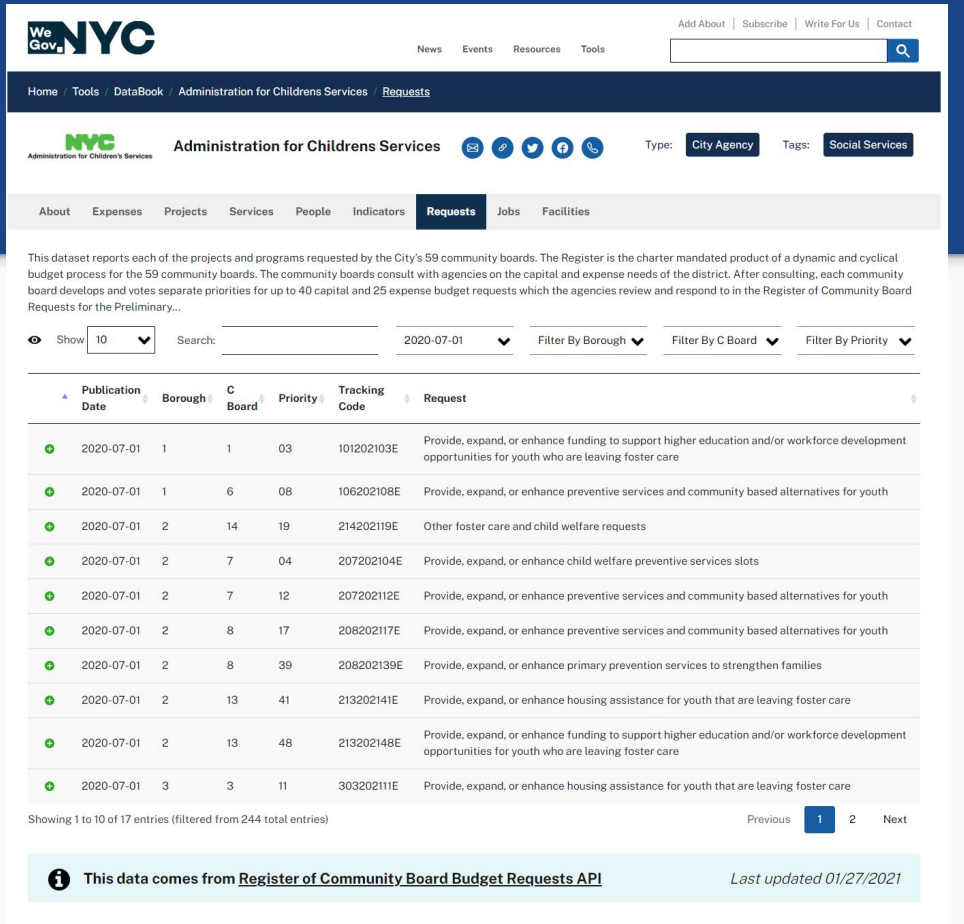
This data comes from Register of Community Board Budget Requests API *Last updated 01/27/2021*

What's Next?

- 10+ more agency datasets
- Adding charts and maps
- Normalizing geographies: council districts, community districts, NTAs
- Adding interactivity with comments and other user generated content
- Integrating/incorporating other projects

Then hopefully:

- Partnerships with good government orgs
- Advocacy for more, better, normalized open data and API-driven web apps



Zooming Out: We’re Ready to Build Community

We want to create spaces where people creating technology in the public interest can raise awareness about issues, showcase their projects and onboard collaborators.

To do that, we’re gearing up to offer the following:

- News: stay up to date with our blog, social media, newsletter and Slack.
- Events: meetup to learn and share new things, and find collaborators.
- Knowledge: a dynamic directory of relevant people, places and things.
- Services: a fiscal sponsor enabling projects to access nonprofit benefits including tax-deductible contributions, credits, and more.

Open for Business

We’re hiring a data manager & community builder.

We’re also looking for sponsors that believe in our mission.

Get Involved: https://tinyurl.com/wegovnyc

Contact Me: Devin Balkind, devin@sarapis.org, @devinbalkind on Twitter

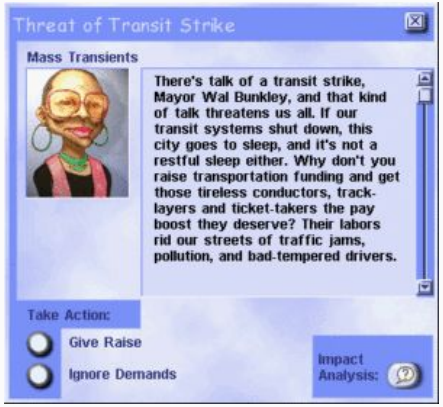
Notes

Eventually: SIM**NYC**ity

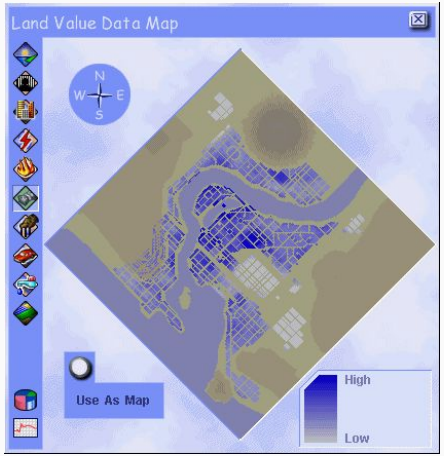
Imagine an interactive “game” where you could see city data on real maps, follow agency activities, give feedback on issues, make your own budget and more.

Citizen-Driven Budgets

Decision-Making



Data Maps



Geospatial Awareness