[NYCpedia](#) aims to be a comprehensive encyclopedia of all data NYC. It culls a lot of data from public and private sources into hyperlocalized, open knowledge.

Our ambition is that it becomes a true open data commons federating all these datasets. And for it to truly become one - the Wikipedia of Urban Data, we want to take Open Data beyond just governments publishing data. We want to engage the community to contribute open data as well.

We have a big backlog of PDF-trapped datasets that we want to onboard, and we'd like to take the community's help to liberate these datasets so we can share it with everybody else.

For all NYC-related datasets liberated during the Hackathon, we will host the resulting scraper and the data at [nycdata.pediacities.com](#), which will then in turn feed [NYCpedia.com](#).

==

NYCEDC Monthly Economic Snapshot

A monthly report is compiled by NYCEDC that is loaded with all kinds of interesting data. Its first page has employment figures broken down by industry and borough.

It also has standard indicators for tourism and travel embedded in prose (e.g. air traffic, broadway sales, hotel occupancy, transit ridership, etc.)

These indicators will be added to the various borough and district pages in NYCpedia.

http://www.nycedc.com/resources/economic-research-analysis/economic-snapshots

PDFs for the City Council & Community Board Dashboard

BetaNYC - the largest civic hacker community in the US, is embarking on a project to build tools for government representatives at the City Council and Community Board in its aptly named #BeyondTransparency initiative.

#BeyondTransparency aims to use the Open Data that BetaNYC helped liberate to give our representatives better decision-making tools.

During the [kick-off #BeyondTransparency meetup](#), representatives from the Council and the Boards expressed that they didn't have "actionable", at-a-glance data that they can use to inform their decisions. A lot of their work is also driven by constituents calling after the fact to complain about day-to-day issues (e.g. potholes, noise, construction, etc.). Having more proactive, at-a-glance information would allow them to do a better job.

These indicators will be added to the City Council/Community Board dashboards being built by the BetaNYC community. Already, 2 days after the kick-off, [high-resolution, machine-readable crime data has been liberated](#) from the [NYC Crime Map](#) which will be used in the dashboard.

Some sample NYC PDFs that have been prioritized as valuable sources for the dashboard are:

- Scorecard/Street Cleanliness Ratings
  http://www.nyc.gov/html/ops/html/data/street_scorecard.shtml
- DOT Street Construction Schedule
  http://www.nyc.gov/html/dot/html/infrastructure/construction.shtml
- Department of Aging Senior Profile Data
  http://www.nyc.gov/html/dfta/downloads/pdf/demographic/profile_olderNYRS_0911_newsummaries_update.pdf and other Dept of Aging PDFs

If you want to have a Dashboard Indicator that you care about from other PDFs, the Dashboard effort is also a community effort. Make your case by liberating the PDF data first!

Data from Environmental Impact Statement PDFs

Environmental Impact Assessments have been mandated by the Federal Government for certain types of projects since 1969. Many states and one city (New York City) have their own processes to measure the potential impact that projects may have on the environment, society, or the economy. Projects that go through an Environmental Impact Assessment process collect and analyze data and publish it in a PDF report. There is often unique data generated in these reports that may be useful to others. In New York City many of these reports are gathered at this website:
http://www.nyc.gov/html/dcp/html/env_review/eis.shtml

The "DEIS" includes full reports that contain many tables that could be extracted. These reports are not standardized, but there are common tables of data that are generated for many of them. Interesting examples include Noise and Traffic.

We will aggregate these valuable, high-resolution, privately-generated data that is now in the public domain but trapped in PDFs, and publish it in nycdata.pediacities.com.