Case Study:

New York City Open Data / New York City Analytics

Type: Team / Website

Organisation(s): government of New York City

Tags: open data, process, metadata, standards

NYC Analytics   **NYC** OpenData

New York City Open Data (NYC Open Data) and the Mayor's Office of Data Analytics (NYC Analytics) were both established in 2012. NYC Open Data has since become one of the largest and most well-used open data platforms in the world.

NYC Open Data is one of the main programmes run by NYC Analytics, but the organisation has a wider remit and also explores and delivers practical applications for city data.

As of 2019, it had published 2,167 data sets in total, of which 14% were automated (i.e. self-updating), whilst the platform achieved 119,000 unique visitors per month.

NYC Analytics has been made a political and strategic priority and is now a beacon of best practice for open data portal design and governance, and for real-world application of open data.

Background

The New York City Open Data portal was founded in 2012 as part of the remit of the new Mayor's Office of Data Analytics, a dedicated city department with a team of permanent employees working on city data. This was at a time when open data was on the ascendancy in the USA, with Barack Obama issuing a 'Memorandum on Transparency and Open Government' on his first day in office. The New York City Mayor at the time, Michael Bloomberg, was also heavily supportive of this agenda.

Other bodies have also been strongly supportive of open government data in the USA, including the Sunlight Foundation, which supports data transparency to drive effective government.

NYC Analytics is underpinned by a local 2012 law that obligates all agencies to openly publish data according to clear standards by the end of 2018. A clear

GREATER MANCHESTER
DOING THINGS DIFFERENTLY

framework has also been established for publication, including ['Examinations and Verifications'](#) to ensure disclosure of all open data requests.

In December 2018, the Mayor's Office of Data Analytics [(MODA) was permanently coded into the New York Charter](#), establishing new charter mandates for MODA. This included a formalisation of the relationship with the Mayor's Office of Operations, the Mayor's Office of Information Privacy, and the Mayor's Office for Economic Opportunity. This obligated MODA to facilitate effective data sharing between departments, advise on use cases, and to develop an open data public education strategy.

The [MODA team](#) has thirteen members in total, including a combination of professionals, some with extensive operational experience in New York City government, but also a policy analyst, project manager, and several data scientists.

MODA aims to raise awareness of the opportunities presented by data and analytical tools across the civil service, and plays a wider strategic role by [allocating staff to advise individual agencies](#) on using data more effectively.

Open Data Values

The NYODA has established [six 'Open Data values'](#) which aim to foster an open data culture. The first principle 'start with users', reflects a strong commitment to user-friendliness, including meeting the needs of users with disabilities through measures such as screen readability.

The second principle is 'treat the publication of the dataset as its debut' and recognises the importance the rigorous checks for quality and accuracy before publication of each dataset, to avoid publishing poor quality, uncleaned datasets ad hoc.

The other four principles are:

- Encourage purposeful and easy engagement
- Empower agencies
- Integrate open data into citywide processes
- Learn, test, standardize – and learn again

Important considerations

Revised strategy and next steps

The latest strategy, the [Open Data for All Strategic Plan](#), focuses on refining further the user-friendliness of the platform, improving accessibility and providing a helpdesk to support users with any queries or suggestions. It also increases standardisation through a [formal publishing process](#) and includes clear KPIs to track progress over time, ensuring continuous improvement.

GREATER MANCHESTER
DOING THINGS DIFFERENTLY

New York City is also increasingly working to [link up its data with data held by other cities](#) across the United States on the [US data.gov website](#). They have already published 900 datasets on the national data portal, including SAT results, air pollution rasters, and the Hurricane Sandy Inundation Zone.

Content and quality

The website is easy to navigate. It is possible to filter by most recently published data, most popular data, by agency, and by thematic category. This allows even an unfamiliar user to quickly find relevant datasets.

Every available dataset includes useful tabs to the side, including: 'About', which gives details on all downloadable formats for each dataset; metadata; a discussion forum on the specific dataset; and recommendations for related datasets. Another tab links to [Socrata APIs](#), which are readily available for most relevant datasets, and found as a link to JSON data and an API interface. Developers have to go through this route to link with NYC Open Data APIs, although developers can also access the NYC Open Data APIs through [Programmable Web.](#)

The largest publisher of all NYC departments is the Department of Education, which provides in-depth data on school districts, SAT results, attendance and enrolment, whilst the most common file type available is XLSX.

The NYC Analytics team actively explore new ways to use data to achieve practical benefits. One project the NYC Analytics team worked on was to identify where [discrimination by landlords](#) was most likely to happen, using housing, crime, and education data. This data then fed through to criminal justice action in which actors posed as tenants trying to rent apartments. The action led to several arrests of landlords who discriminated against tenants based on their source of income, such as those receiving housing vouchers from the state.

Another NYC Analytics project related to housing was a tool developed to help building inspectors identify [particularly dangerous locations](#) they need to visit where, for instance, illegal subdivisions have been built. By analysing data on individual tax lots, MODA found that certain features were strongly associated with the incidence of fires. This included building age, age of complaint, property value, property size, whether the building had a history of unpaid taxes and whether the Department of Buildings had received previous complaints. These data points were then used to inform NYC Analytics in developing predictive models for their tool.

Awareness Raising

NYC Analytics puts a great deal of emphasis on openness and flexibility. For instance, users can submit requests online for additional datasets to be published using the ['send us a note' function](#). Moreover, the city holds competitions to encourage freelance developers and analysts to create apps and tools to enhance

the website. The website does not try to differentiate users, and so developers will access the data through the same means (or via Socrata APIs) as other users.

One good example is the [NYC Connector](#), which helps connect people to local charities, community centres and other social facilities that are looking for volunteers. The tool uses three datasets on government facilities, senior centres and an NYC Women's Resource Network database – all available on NYC Open Data – to identify locations for volunteer work.

Usage

Detailed information is available on the usage of NYC Open Data. The most popular datasets include data on vehicles for hire (1.4m views, 229k downloads); calls to the 311 phone line for information about local services (398k views, 394k downloads); and planning applications to the Department of Buildings (2.25m views, 33.1k downloads).

This latter example might have proved confusing, as the description references jobs applications – an important economic indicator – despite this being equivalent to local authority planning applications (e.g. house conversions, extensions or buildings) in the UK. This might explain, in comparison to other datasets, the difference between the high number of dataset views and the low number of downloads.

There is still a visible emphasis on data about local government transparency and payroll, with many of the datasets covering public sector operational information – 734 datasets under the 'City Government' theme, and 928 under the 'Education' theme.

Blockers and challenges

User experience

Whilst the website is easy to use and attractive, there are also some apparent weaknesses. For instance, unlike in many other public datastores, easily navigable maps or other geospatial data are not a major part of NYC Open Data. This might make the website less attractive to non-expert members of the public. However, the large number of views and downloads of datasets from the platform despite this indicates that visualisations are not particularly important for a significant segment of the user base.

There is also at times a lack of integration between NYC Analytics and some other useful data sources available about the city. For instance, useful city dashboards linked to New York public data, such as the [NYC Construction Dashboard](#) (which is built on NYC Analytics Department of Building Data) are available in separate

locations and not linked on the NYC Open Data website. This dashboard, for instance, is hosted by the nyc.gov website.

What can Greater Manchester take from this?

- Clearly articulated principles, KPIs, and regulatory frameworks can play a major role in driving the success of city data projects. When designed and implemented properly, these do not act as bureaucratic hurdles, and can instead help foster an innovation-driven culture.
- It is important to embed a culture of open data across departments and organisations to ensure widespread participation in open data projects.
- Constant and active reflection on innovative ways data might be used and how data can improve the lives of city residents keep the project on track and provides a sense of purpose for users and publishers alike.
- Putting user engagement at the centre of an open data strategy helps ensure any quality or usability issues are flagged early and dynamically instead of relying on post hoc evaluations.
- Visualisations may not be as important as is often assumed in engaging large numbers of data users.

Find out more:

2019 Open Data for All Strategic Plan

Analytics Vidhya - Accessing NYC Open Data APIs with Application Tokens

Gov Insider - How New York City used analytics to solve urban challenges

Government Technology - New York City Open Data: A Brief History

Civic Hall - how the NYC commission on human rights uses open data against discriminatory landlords

GREATER
MANCHESTER
DOING THINGS DIFFERENTLY