socitm insight

briefing

... a Socitm *Research* service

Big data: What does it mean to 'do' it?

A recent report by Capital City Foundation, *Big Data in the Big Apple*, advocates a Mayor's Office of Data Analytics for London based upon the New York model. Much of this is a good checklist for doing big data anywhere.

You need to remember at all times that the point of all this effort is to help your city and its people thrive. Keep all this in mind. Just dive in and do it. You may be amazed at what you find.

Mike Flowers, founding director of MODA

Executive support is the foundation

Extract from Executive Order #306: THE MAYOR'S OFFICE OF DATA ANALYTICS (MODA) April 17, 2013

Section 1
There is established in the Office of the Mayor The Mayor's Office Of Data Analytics (the "Office").

Section 2
The Office will be headed by a Chief Analytics Officer. The Chief Analytics Officer shall report to the Mayor's Chief Advisor for Policy and Strategic Planning and shall consult regularly with the Deputy Mayor For Operations, the Deputy Mayor For Economic Development, and the Commissioner Of the Department Of Information Technology And Telecommunications (DoITT).

Section 3
The Office shall develop and work with agencies to implement data-driven solutions to City service delivery issues. The Office's responsibilities shall include but not be limited to the following:

a. Collaborative, Data-Driven Solutions. The Office shall work with City agencies to identify how data held by those agencies can be analyzed and combined with other agencies' data to best fulfill their respective missions, shall develop strategies based on such data, and shall assist agencies in implementing those strategies.

b. Citywide Data Platform. The Office shall develop and implement a citywide data platform that aggregates and updates data from City and other governmental agencies and other sources and that connects and synthesizes data regarding a single address, business, or individual that was previously isolated within individual agencies. In collaboration with DoITT, The Office shall:

Work with city agencies to ensure that they have proper technology to provide and retrieve data from the citywide data platform; Train agency staff to use the citywide data platform; and Develop practices for performing ongoing and new data analytics with each agency.

c. Oversight of Data Projects. The Office shall, as appropriate, oversee agency data projects to ensure that agencies use best practices and such projects are appropriately prioritized by need, impact and feasibility.

d. Data Liaison duties. The Office shall serve as the designated point of contact for outside partners contributing to or using City data.

e. Implementation of the Open Data Law. The Office shall work with DoITT to ensure compliance by City agencies with the Open Data Law. The Chief Analytics Officer will also serve as the City's Chief Open Platform Officer (COPO) as defined in DoITT's Open Data Policy and Technical Standards Manual.

New York's journey started in earnest with the firm, public commitment of Executive Order #306. It made plain a detailed but straightforward mandate from which a series of detailed actions could flow.

The right size is 'just enough'

The Mayor's Office Of Data (MODA) is headed by a Chief Analytics Officer who reports upwards to obviously strategic interests (strategic planning, operations, economic development and ICT). The role is now supported by a Director of Analytics, a Director of Public Affairs and a Director of Strategy + Operations [1].

The Office's core functions include [2]:
- Collaboration with City agencies to implement data-driven solutions to City service delivery issues.
- Building a Citywide data platform to facilitate data sharing.
- Oversight of Citywide Data Projects.
- Implementation of the City's Open Data Law.

CCF report that the MODA team is not large: "In 2013 the roles included: Chief Analytics Officer and Chief Open Platform Officer; Deputy Director; Chief of Staff; Chief Analyst; Chief Programmer; Analyst; Special Advisor to the CAO; Technology Advisor to the CAO; and Senior Advisor to the CAO. Far from creating a significant additional layer of bureaucracy, the New York MODA model is lean and highly efficient."

How does it work in detail?

In MODA's 10-step model we see the classic Deming cycle, plan–do–check–act (PDCA) and its variant observe-plan–do–check–act (OPDCA) exemplified:

	MODA's 10 step model	OPDCA stage
1	Understand how day-to-day operations work	observing
2	Identify areas where data could help	planning
3	Form a project plan	
4	Understand data context	
5	Create a Memorandum of Understanding (MOU)	doing
6	Integrate data	
7	Test hypotheses	
8	Service delivery team review	checking
9	Automate the process	acting
10	Implement solution	
Repeat by moving on to the next opportunity…		

There is an optional 11th step, to 'Delegate responsibility for the data model' when it is appropriate to hand over continued maintenance of a product/service to another party.

The following steps guide each stage in the process:

Step One
- What is the service being provided?
- How is the service allocated, scheduled and delivered?
- What factors go into the prioritisation of delivery?
- How is the delivery recorded in the organisation's ICT system(s)?

Step Two
- What type of problem is this?
- What data exists around the operation?
- What other data would be helpful (hypotheses)?
- What is the desired end goal of the data use?
- What's the commitment from the agency and MODA?

Step Three
- What data will be used, and what new data is needed?
- What is the timeline for the project?
- What are the checkpoints during the project?

Step Four
- In what context should the data be understood?
- What are the datasets and what are they measuring?
- How is the data generated?
- What road bumps should MODA anticipate?
- How is the data interpreted?
- How is the data set stored?

Step Five
- What is the purpose of the project?
- What data security guarantees are provided?

Step Six
- What sort of system records the data?
- What is the most appropriate method for transmitting the data to MODA?

Step Seven
- What variables will we test?
- What's the most appropriate analytical technique for this analysis?
- What do the preliminary results show?
- How do we communicate these results to the delivery team?

Step Eight
- Does any of the analysis surprise the delivery team? If so, why?
- What agency procedures could account for data surprises?
- How can the analysis be altered to produce a more accurate result?
- How should the analysis be tested in the field (pilot)?
- How will we know if the pilot is successful?
- What systems are necessary to support the pilot and pilot measurement?

[1] http://www.nyc.gov/html/analytics/html/about/leadership.shtml   [2] http://www.nyc.gov/html/analytics/html/about/about.shtml

Step Nine
- What system needs to be changed and how?
- How often will the solution be reviewed and calibrated?
- Are we confident that this is not disruptive to the field?
- How do we maintain the solution on an ongoing basis?

Step Ten
- What education needs to be provided to staff in the field?
- How will success be measured over time?

DataBridge and DEEP infrastructure

MODA has a series of agreements with agencies concerning data exchange with itself. Data from the city's data element exchange programme (DEEP) is gathered together and made available for reuse in the DataBridge data warehouse system.

CCF report that the DataBridge permits cross-agency data analysis from over 40 different agencies: "Through detailed analytics, MODA finds previously unknown patterns and relationships that lead to better decisions and resource allocation."

It's worth noting that New York's Open Data Law is undoubtedly helpful in promoting an ethos that data is at least potentially sharable (see Box).

> *There is a strong steer from central government to have more public open data, and many organisations have complied to various extents. But there is no UK or regional open data law per se to back that up. The New York City Council passed its own local law which preceded the MODA initiative. Have any Socitm members got such a mandate? Would it be a good thing if there was?*

Box: Local Law 11 of 2012 – Publishing Open Data
Be it enacted by the Council as follows:

Section 1. Declaration of legislative findings and intent
The council hereby finds and declares that it is in the best interest of New York city that its agencies and departments make their data available online using open standards. Making city data available online using open standards will make the operation of city government more transparent, effective and accountable to the public. It will streamline intra-governmental and inter-governmental communication and interoperability, permit the public to assist in identifying efficient solutions for government, promote innovative strategies for social progress, and create economic opportunities.
http://goo.gl/ldwWpH

Why do the same or similar here?

The CCF report advocates establishing a similar function in the London Mayor's office and describes the kinds of work it could perform:
- Intelligently designing shared services.
- Combating illegal conversions.
- Identifying empty homes.
- Fighting tax and benefits fraud.
- Targeting food safety inspections.
- Boosting new business growth.

> *What comes to mind in a similar vein in your area?*

As reported by CCF, the lessons learned by the MODA team have many resonances for experienced Socitm members:

1. Strong executive support is essential – "Modern technology makes it much easier to access and analyse city data, but it will only make a difference if there is the political will to use it."

2. Data models must be shaped by front-line experience and expertise – "Data is meaningless without context."

3. Focus on outcomes that provide a proven return on investment – "MODA is clear that what matter are outcomes. According to its model, data initiatives do not proceed beyond a trial stage unless there is clear evidence that they deliver tangible benefits and improvements in service."

4. Start small and with measures everyone can support – "Fires and rats have no political supporters."

5. Do not try to change the work of front-line staff – "[Collaborate] with front-line staff to make their work even more effective, increasing employees' job satisfaction."

6. Using data does not require vast numbers of specialised personnel or new layers of bureaucracy – "The NYC MODA team was led by a former lawyer who recruited a small team of data analysts via Craigslist. Today MODA is made up of just nine people."

7. Using data does not require procuring high-end technology – "The team began with nothing more than old spreadsheets and analysis conducted in Microsoft Excel. …This 'just build something that works' approach reduces the technical obstacles that could prevent getting things done. In cases where modest IT changes are required by an organisation to connect to MODA, a further principle of the New York model is that those changes are paid for from a central budget. This removes the financial barriers to data sharing."

8. Any organisation that wants to access MODA's data must first share their own – You have to give to get. No free riders there.

9. All data must be geo-tagged (geo-coded) – "While adopting open standards (i.e. common ways of recording information) can make this process more straightforward, it is not a necessary condition for the MODA model to work. Organisations in New York City use several different methods for geo-coding and MODA does the hard work of matching them."

10. No part of the data extraction or analytics process should require human action – "For data-driven analytics to reliably improve a service, the processes created by the MODA team must eventually be automated. Requiring a person to complete a step in the chain before it reaches front-line staff creates single points of failure."

What's happening around the Socitm community?

As yet there's nothing that parallels a Mayor's Office Of Data headed by a Chief Analytics Officer, but a start is being made in "data" per se. In due course an ODA would be a logical consequence of such projects as:

Leeds
(http://leedsdatamill.org/background/)
Leeds Data Mill is a partnership between Leeds City Council and Leeds-based digital content, data insight and storytelling specialists, Hebe Works. The partners want to "enable people and organisations to explore the different complex relationships between the city's services and businesses, by collecting Open Data from multiple sources in a single website, offering a greater insight into the workings of the city than ever before."

Greater Manchester
(http://data.gmdsp.org.uk/themes)
The Greater Manchester Data Synchronisation Programme coordinates the release of open data between the Manchester-based innovation hub, FutureEverything, and Greater Manchester's local authorities. The plan is to: build an RDF linked open data store; enable skills transfer into the public sector; create capacity through placement of Code Fellows into local authorities; build peer support structures; and encourage the creation of new public/private digital services and insights.

Birmingham
(http://data.birmingham.gov.uk/)
Birmingham's Data Factory is administered by Digital Birmingham on behalf of Birmingham City Council: "…the place where organisations in the Greater Birmingham Local Enterprise Partnership area can publish their open data."

Further information

- *Big Data in the Big Apple, The lessons London can learn from New York's data-driven approach to smart cities*, Eddie Copeland, Capital Cities Foundation, Foreword by Mike Flowers, founding director of MODA, June 2015: http://capitalcityfoundation.london/big-data-in-the-big-apple/

- NYC by the Numbers, Annual Report – 2013, Michael R. Bloomberg, Mayor, City of New York and Michael Flowers, Chief Analytics Officer, Mayor's Office of Data Analytics, December 2013: http://www.nyc.gov/html/analytics/downloads/pdf/annual_report_2013.pdf

Your feedback and sharing additional information
We encourage members to share any feedback or additional information about the subject of this briefing via the Insight group in the Socitm Network at: https://khub.net/web/socitm-insight.

If you have not already signed up for the Knowledge Hub and enabled system notifications of *Insight* group activity (by user-selected periodic email to your Inbox), then please do so or contact the Socitm office for assistance if you are having difficulties getting onto the Network.

Insight publications are published via the Socitm *Insight* Group in the Socitm Network on Knowledge Hub.

socitm insight