OPINION 10
February 2014

Sustainable Data Science for Sustainable Cities: Big Data and the Challenge of Urban Development

By Linnet Taylor

Digital data for spatial planning are an essential tool for urban development, especially in newly urbanising countries where new sources of digital data can both help to make sense of changing needs and demographics, and can enable interactive urban planning and governance. These facets of urban digital data have been distinct from each other until recently: 'clean', well-behaved data that are amenable to analysis versus 'raw', less structured data such as mobile phone records, social media and other 'big data' sources. Today, however, **the boundary between data subject and data user is becoming permeable**. As citizens increasingly emit spatial data through their use of technology, and as data science becomes a more accessible tool for advocacy groups[1], faster and more extensive data can potentially allow city authorities to see and respond to factors such as mobility, crime and public feedback in real time.

In high-income countries these new sources of data are already the subject of interest and experiment[2]. However, they demand **new configurations of actors and institutions** which may be difficult to recreate in low- or middle-income countries (LMICs). There are two issues cities have to resolve in order to benefit from spatial big data. First, the problem of access. Some big data are generated and channelled by city authorities, for example via traffic sensors, e-government applications, or electronic travelcard records, and create no access problem (though resources are needed for storage). Many big data, however, are channelled by the private sector – for example real-time details of people's movements via mobile calling records, or details of economic transactions performed over mobile networks. Some, such as public feedback projects, are channelled by civil society organisations, often in combination with academia. These are designed to inform the city about gaps in service provision and needs for housing, healthcare or crime control, but may involve data flowing from citizens through other authorities than the city itself. Sometimes the private sector acts as mediator for such projects, for example Microsoft Research's project in Cambridge, UK, which aims to poll citizens for their opinions on neighbourhood issues of development and spatial planning[3].

New flows of data such as these make it necessary for the city to **negotiate and cooperate with new partners** to gain access to information that can greatly enrich urban planning and dialogue. They also raise questions of power and representation: under what conditions should the private sector

1 One example is Tactical Technology Collective's data visualisation toolkit for advocacy https://tacticaltech.org/projects?type=8

2 Examples include the use of predictive traffic management technology in Lyon (http://mashable.com/2012/11/14/ibm-technology-traffic/), and New York City office of Policy and Strategic Planning's use of big data analytics to locate illegally converted housing (http://www.cio.com/article/719926/How_Big_Data_Save_Lives_in_New_York_City?page=1&taxonomyId=600010)

3 http://tenisonroad.com/

share data with city authorities; who should regulate such exchanges, and what will the private sector expect in return? What does it mean for the private sector to become the mediator between the citizen and the city? How does advocacy organisations' ownership of data on public needs and priorities change their relationship with city authorities? Does the citizen gain leverage as a producer of big data, and can the process be participatory despite the technical skills involved?

The second problem to be solved is how to analyse the data. **Big databases require big resources.** Not every big data problem requires a supercomputer, but significant software and hardware resources are necessary to manage and analyse big data due to its richness and real-time aspects. Similarly to the issues with access, new partnerships and alliances are also necessary for big data analysis since few city authorities, even in high-income countries, have data scientists on standby but must borrow them from the private sector or academia. **Options for facilitating data science in fast-developing cities** include the one-off 'data challenge' model, where private-sector and academic data scientists compete to produce the best analysis of a previously unavailable dataset. Another option is for data scientists to be hired into existing city departments, as occurred with the New York City Mayor's office (see footnote 2). A third path is pro-bono work by external experts, as performed by organisations such as DataKind[4]. Finally, some applications are being developed (for example Amsterdam's predictive policing tool Datadetective) which translate big data into simpler formats and analytical methods so that nonspecialists within the city government can run predefined forms of analysis.

These options suggest, however, that **data science for city development may have a sustainability problem.** Even in high-income countries (HICs) where data, resources and computer power are abundant, data science can be expensive, resource-heavy and private-sector dominated. The sustainability challenge is much greater in LMICs because there, data science tends to be an external phenomenon. Although there has been a recent sharp increase worldwide in access to the kinds of communications technology which generate big data about cities, the computing resources, skills and tacit knowledge remain hard to find in less well-resourced locations. Despite this, there are LMICs, such as Rwanda, who are pursuing international collaborations to build analytic capacity[5].

The 'Data for Development' challenge run in 2012 by Orange[6] in Cote d'Ivoire is a case in point: 168 teams of data scientists worldwide were invited to compete to produce the most innovative analysis of mobile data from Cote d'Ivoire, producing conclusions about economic development, epidemiology, transport planning and population mobility (NetMob 2013). However the teams worked remotely from universities and firms in HICs, unconnected with Abidjan city authorities who could have identified the most pressing issues for them. After the challenge, Orange connected with city and national authorities in Cote d'Ivoire to see how the research could be applied. The 'development' aspect of the project was used to attract the best data scientists with the promise of doing good for a country with few resources, but no information on local needs was offered in advance to guide the research. In contrast, a Dutch-Ugandan collaboration, externally funded, was responsible for the 'TracFM'[7] public service feedback project in Kampala, which engages strongly with local political debates by polling radio listeners, visualising the results and inviting policymakers to discuss them on local radio stations. The TracFM project contrasts with Orange's challenge in several ways: it was generated by findings from development research in Uganda, the collaborators were private but noncommercial (foundations and NGOs) and the data science, rather than being an end in itself, was a tool to generate two-way engagement between government and citizens.

These examples demonstrate that the most important factors in the sustainability of city data science worldwide are undoubtedly local resources and mutual engagement between the producers and users of data. If a city's governance model involves seeking citizen participation in terms of volunteering data, planning and feedback, data science will be seen as an important tool and cities will seek out the necessary resources and alliances. If public accountability is not a priority, external data science may take place, but it may have little impact on local governance and planning processes. Inclusion is not simple: findings ways for the poor and marginalised to participate in providing data

4 http://www.datakind.org/

5 http://www.cmu.edu/rwanda/degree-program/index.html

6 http://www.d4d.orange.com/home

7 http://www.tracfm.org/s/about/

has been shown to be a huge challenge, as has the inclusion of women, and successfully engaging with government often requires persistent and prolonged effort which presents a barrier to citizens with a shortage of time and resources (Mount 2013). For data science to be sustainable, it must be locally meaningful, must draw on local resources, and must demonstrate benefits to local residents as well as city authorities.

References

Mount (2013) How New Technologies Can Make Every Voice Count in Government Activities. Accessed 24.2.2014 at http://www.ictworks.org/2013/07/16/how-can-new-technologies-be-used-to-make-every-voice-count-in-government-decisions/.

NetMob (2013) 'Analysis of mobile phone datasets for the development of Ivory Coast'. NetMob Conference proceedings, May 1-3 2013, MIT, Boston.

EADI
Kaiser-Friedrich-Strasse 11
D-53113 Bonn
Tel.: (+49) 228 . 2 61 81 01
info@chance2sustain.eu
www.eadi.org
www. chance2sustain.eu

ISSN 2308-0965

Chance2Sustain examines how governments and citizens in cities with differing patterns of economic growth and socio-spatial inequality make use of participatory (or integrated) spatial knowledge management to direct urban governance towards more sustainable development.

Consortium partners: European Association of Development Research and Training Institutes (EADI, Germany), Governance for Inclusive Development (GID) at the Amsterdam Institute for Social Science Research (AISSR-UvA, Netherlands), Centre National de la Recherche Scientifique (CNRS, France), Centro Brasileiro de Análise e Planejamento (CEBRAP, Brazil), Cities for Life Forum (FORO, Peru), Norwegian Institute for Urban and Regional Research (NIBR, Norway), School of Planning and Architecture (SPA, India), University of KwaZulu-Natal (UKZN, South Africa)

EUROPEAN COMMISSION
European Research Area

SEVENTH FRAMEWORK PROGRAMME

Funded under
Socio-economic
Sciences & Humanities