

The UK big data project playing Moneyball to build smarter cities

The Future Cities Catapult aims to improve policymaking – be it for better flood defences, banking or medicine

[Katie Allen](#)

Sunday 3 January 2016 The Guardian

When London's Soho was hit by a cholera outbreak in the summer of 1854 it took a maverick to pinpoint the cause and come up with a way to contain it.

Doctor John Snow rejected the established belief that cholera was airborne and he set out to disprove the miasma - or "bad air" - theory by [plotting cases on a map](#). The physician's findings proved his theory that the disease was in fact being spread through contaminated water. He traced the outbreak back to a water pump.

Snow's story is one of the oldest examples of someone asking the right questions, analysing a dataset and throwing up new insights into previously stubborn challenges.

As such it is also one of the favourite tales told by Peter Madden, the man heading the government's [Future Cities Catapult](#). The centre was established just a stone's throw from where Snow made his discovery to help modern-day innovators find ways to make smarter decisions about how to run cleaner and more efficient cities in the UK and beyond.

The hope is that in 2016, so-called "smart cities" will map and analyse all manner of statistics, from household waste volumes to social care needs, to better direct their services. For anyone who fancies trying their hand at this blending of big data and big decisions, Future Cities Catapult has created a game for the [Big Bang Data exhibition](#), now running at London's Somerset House until March.

The idea is to demonstrate how data and new modelling methods can help policymakers set priorities and predict the impact of their decisions.

In an era of spending cuts and pressing environmental problems, the potential rewards are huge, but reaping them is by no means simple. There are four obvious challenges.

Firstly, there is the sheer quantity of data. As the Big Bang exhibition's title implies, there has been an information explosion in recent years, as we increasingly use our phones and computers to manage our day-to-day lives and leisure time. With every click, search and selfie, the mass of data expands. Experts predict a [4,300% increase in annual data generation by 2020](#).

The second challenge is the quality of data. In the ever-growing hoard of information, some bits are more useful than others. Some are more reliable. In the past, researchers and policymakers could navigate this patchy world by putting datasets into a sort of hierarchy, ranging from a dodgy sponsored survey at the bottom up to official statistics at the top.

Now, as the way we work and live changes, that old model no longer stands. Sensible businesses have been using their own data to shape decisions for years. Take supermarkets and the vast insight they gain from loyalty card schemes. Official data collection and statistics in the UK have been criticised for failing to keep up.

Sir Charlie Bean is among those who say the UK's official statisticians need to move with the times if their figures are to capture the real state of a modern economy, where people shop online, work from coffee shops and invest in laptops more than lathes.

The government has asked the former deputy governor of the Bank of England to [lead a review of the country's economic statistics](#), and he reports back in March. His [interim report](#) published last month found "statistics have failed to keep pace with the impact of digital technology" and he called for the Office for National Statistics to be more "proactive" in producing figures.

Advertisement

Bean's interim report brings us to accessibility, the third challenge for those wanting to harness the economic power of big data.

One of his key recommendations on future-proofing economic statistics was for the UK to unlock the vast hoard of data that the public sector collects in the course of its day-to-day business.

There are vital privacy and security considerations, but used in the proper way, this data trove could improve GDP estimates, regional statistics and a host of other indicators. As Bean said: "It's nonsensical that different bits of the government don't speak to each other".

The fourth challenge is the toughest to tackle, but also the most important - working out how best to use the growing mine of available data. When Snow found a pattern on his map of cholera cases, he already knew what he was looking for. He had a theory.

Exploiting big data will take investment in training and a serious dose of open-mindedness. It will also require some unlikely collaborations. Pairings such as that [between a biomedical research institution and the number cruncher who led a data revolution in the world of baseball](#).

Paul DePodesta worked as Billy Beane's assistant on the Oakland Athletics baseball team and used data to change the way the team recruited players. His great success was retold in Michael Lewis's bestselling book, [Moneyball](#) and the eponymous [Hollywood film](#).

Now the Harvard economics graduate is turning his analytical hand to medicine by [joining the Scripps Translational Science Institute \(STSI\)](#). As assistant professor of bio-informatics, he will work on new ways to use the growing amount of data being collected from patients through advances such as genetic testing and wirelessly connected devices such as pacemakers and defibrillators.

DePodesta sees no limits to where data can improve decision making, be it in baseball, banking or retail. As he points out, however, such shifts in thinking are far from inevitable. "Medicine is just beginning to explore this opportunity, but it faces many of the same barriers that existed in those other sectors – deeply held traditions, monolithic organisational and operational structures and a psychological resistance to change," he said when the collaboration was announced.

Snow could tell him a thing or two about resistance to change. Sadly for the 19th-century doctor, his findings about water and cholera were not taken seriously until after his death in 1858. That same summer, it took the "Great Stink", when the stench of the Thames at Westminster was so foul that MPs were forced out of parliament, to finally prompt politicians to rush through funding for a new sewer system for London.

As the cleanup continues from devastating flood damage in the north of England and Scotland, people are quite rightly asking what it will take to catalyse serious action from this government. Homeowners and businesses want to know how much of the damage was predictable, and more importantly how much of it was preventable.

Plans for smart cities, new digital strategies and sturdier statistics are all welcome, but any mention of the power of big data was notably absent from David Cameron's [new year message](#).

It certainly is not everything and cannot replace flood defences but, used smartly, it can help predict where they should go. More broadly, for a government intent on doing more with less, Snow and Moneyball, his modern-day successor, offer plenty of lessons.