July 2013

BIG INNOVATION CENTRE

Lessons learnt from a public–private big data hackathon

Sonia Sousa

The author would like to thank and acknowledge the contributions made to this report by Carl Allchin (Barclays Bank), Charles Levy (Big Innovation Centre), Chris Widgery (Camden Council), Filomena Laporta (EDF energy), Julie Huxley-Jones (GlaxoSmithKline) and Nicola Searle (Intellectual Property Office). Also, thanks to Conor Murray for event and editorial assistance.

The Big Innovation Centre is an initiative of The Work Foundation and Lancaster University. Launched in September 2011, it brings together a range of companies, trusts, universities and public bodies to research and propose practical reforms with the ambition of making the UK a global open innovation hub as part of the urgent task of rebalancing and growing the UK economy, and with the vision of building a world-class innovation and investment ecosystem by 2025. For further details, please visit www.biginnovationcentre.com.

Executive summary

This report presents the lessons learnt from a recent hackathon organised by the Big Innovation Centre and Camden Council, the local authority of a Central London borough. The hackathon involved many of the Big Innovation Centre's partners including universities, public bodies, and companies; such as Barclays Bank, EDF Energy, GlaxoSmithKline and Google.

The Big Innovation Centre–Camden Council hackathon proved that both companies and members of the public are very keen on having access to UK council data. It creates a win-win situation for all, and a great opportunity which local authorities cannot afford to miss. Local authorities are able to find more cost-effective ways of delivering their usual services as well as developing new ones. Local tax payers receive greater value for money. Companies and entrepreneurs are also able to develop new and better services for their customers by exploring the rich local-level granular data.

The event

The event brought together innovation specialists from different sectors, Camden Council leaders, data engineers, programmers and data analysts. An eclectic group of around 75 people came together on the 10th of May 2013 at Google London Campus to solve three specific problems faced by Camden Council:

- Health: What do Camden's future patients look like? What will they be treated for and how?
- Social housing: How can the council target the repairs budget more effectively to deliver the maximum benefit to its tenants (e.g., lower energy costs, greater energy efficiency, better ventilation, or a combination of these)?
- Street presence teams: How can the teams target their efforts to reduce crime and anti-social behaviour?

The data made available on the day came from two sources:

- From the council, including: (1) housing stock condition and repairs data for about 33,000 properties; (2) Camden-wide prescription and illness data; (3) data on ambulance call-outs by category, broken down to ward level; (4) crime data by type of crime, at the street level; and (5) incidents in parks and open spaces at ward level;

- And publically available data at the council level, including: (1) demographic and socio-economic census data by post-code sector and (2) housing energy efficiency data at the council level.

The practical outcomes

This public-private, cross-industry hackathon demonstrated that an open platform where people with different skills and insights can work collaboratively is an effective way of unlocking the potential of largely unexplored datasets. The teams were able to develop several workable tools for Camden Council to build on:

- A web-based simulation tool to better target social housing repairs;
- A tool to demonstrate the performance of the street cleaning teams;
- An application for the street presence teams that allows them to identify venue licencing infringements;
- The identification of key health issues that Camden Council's health services will need to be prepared for in the next 10-15 years.

The lessons learnt

Camden Council's staff members were able to:

- Identify which data quality issues (e.g., connectors, missing data, data formats) caused the greatest difficulties to the exploitation of the data;
- Receive practical suggestions on realistic ways to fix the data quality issues, e.g. consistency of the format, geographical consistency and the type of connectors needed to link datasets;
- Receive new insights on the utility of the data they gather on a regular basis;
- Learn about diagnostic and analytical tools they were not aware of;
- Develop new conceptual ways of addressing the practical questions they were faced with on the day.

Participating companies were able to:

- Identify the sort of data collected by councils;
- Explore potential economic opportunities from merging / linking council data with their own corporate data.

Policy recommendations

The Big Innovation Centre identified policy recommendations at both local and national level. Local policymakers should:

- Push for a culture shift within councils towards embracing an open data and open innovation agenda;
- Work with other local authorities and the central government in creating common standards on (1) data anonymisation, (2) data quality and (3) data release formats;
- Open local public data.

National policymakers should:

- Put in place a big data-friendly legal and regulatory framework for using, combining, and sharing data without compromising privacy, safety, and intellectual property rights. This is not a straightforward task. The lessons learnt from this hackathon coupled with a recent debate organised by the Big Innovation Centre[1] suggest some possible ways forward:
    - A shift away from governing the data itself to governing the usage of data;
    - Explore whether the UK should build on the Creative Commons legal infrastructure to develop a Privacy Commons;
    - Explore whether the UK should adopt an 'opt-in unless you opt-out' approach to personal data disclosure;
    - Ensure that the regulatory framework for big data avoids loopholes, is easy to understand and is easy to comply with.
- Have an active role in the development of new platforms that allow merging and combining datasets from companies, universities and public bodies without putting privacy and security at risk;
- Open up data across all public bodies, especially publicly funded data;
- Ensure that the right digital infrastructure is in place – building effective and reliable digital networks is a precondition for making the most out of big data;

[1] See, for example, the conclusions of the debate co-organised by the Big Innovation Centre and the Economic and Social Research Centre as part of the Big Data Week 2013 event at: http://us4.campaign-archive2.com/?u=f48c74d9cc73a25267a7bc364&id=2ef911193d

- Adjust the education policy to deliver the right data skills – big data needs a labour force able to analyse, interpret and put the insights extracted from data to work.

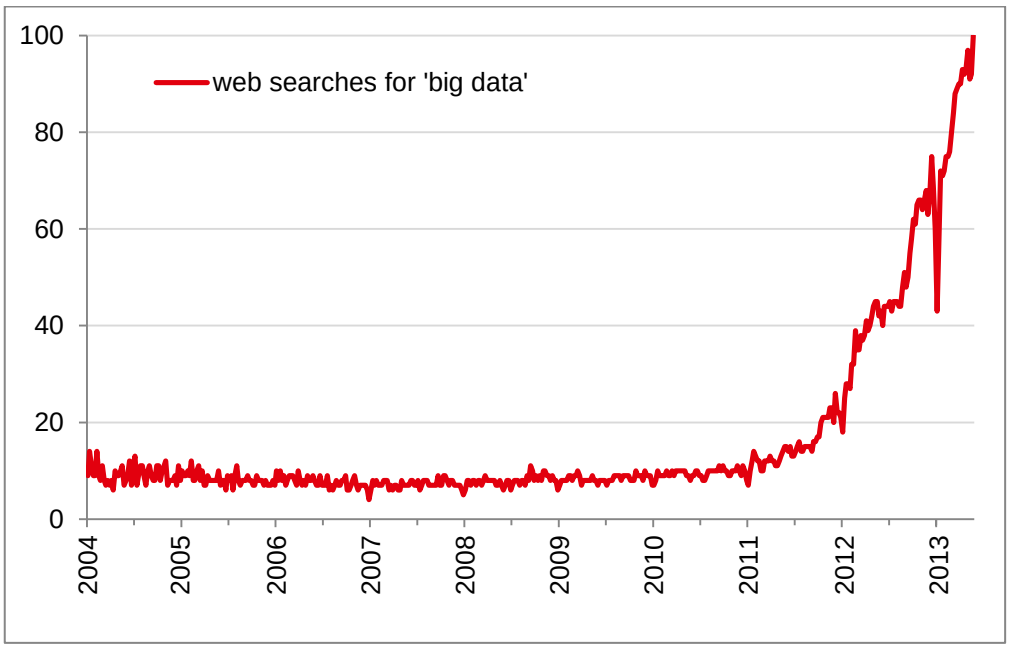
Table of Contents

1. The big data opportunity

Data is growing at an ever-increasing rate and is measured now in petabytes (PB), exabytes (EB) and zettabytes (ZB). The amount of data worldwide is estimated to have more than doubled between 2010 and 2012, from 1.2 ZB to 2.8 ZB, and is expected to be about 14 times bigger by 2020, reaching 40 ZB (EMC[2] 2013).

Businesses, public bodies and other organisations are dealing with fast growing amounts of structured and unstructured data, both quantitative and qualitative. Not surprisingly, the interest in big data is recent but growing exponentially. Take the trend in the web searches for 'big data', as Figure 1 shows. Web searches started to pick up as late as mid-2011 and have skyrocketed since then. This fast-growing, global interest in big data is a manifestation that both companies and public organisations are starting to realise that the insights locked in these vast amounts of data are an asset.

Figure 1: Weekly web searches in a normalised 0-100 scale (1 January 2004 to 18 May 2013)

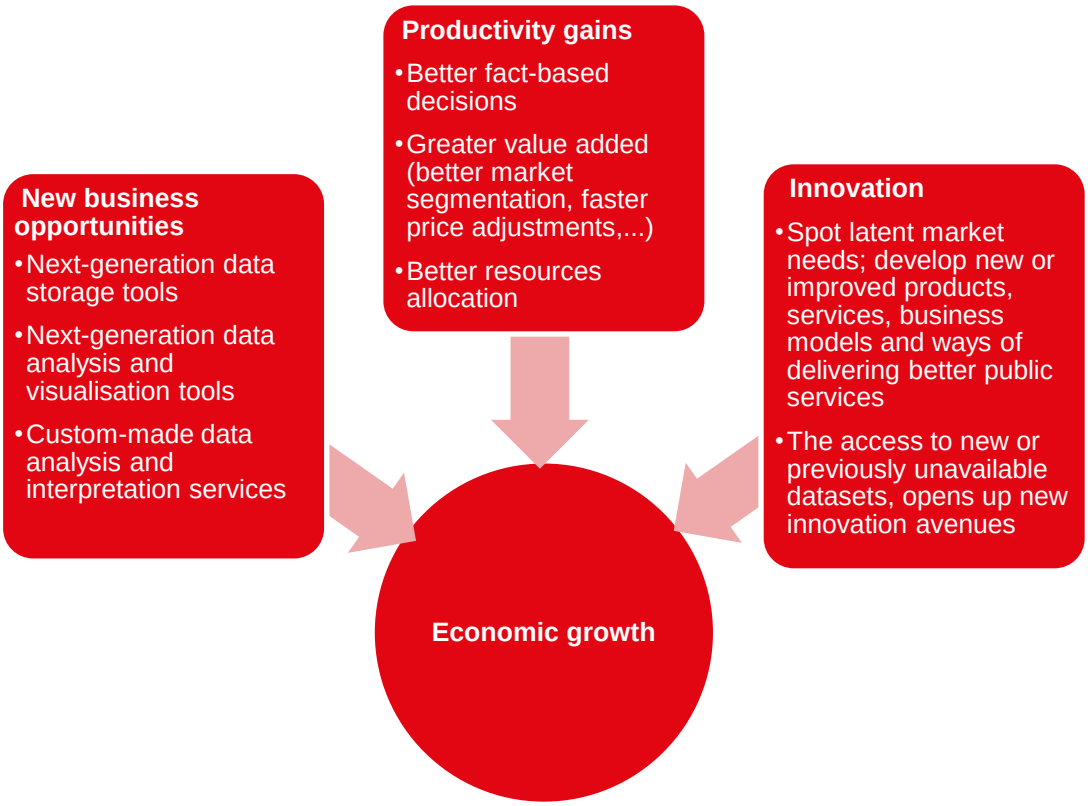


Source: Google trends
Note: The values represent the number of weekly web searches for the keyword 'big data', relative to the highest number of observed searches for this keyword over this time period. The values are presented in a normalised 0-100 scale. The highest number of web searches for 'big data' occurred in the week of 18 May 2013, which resulted in a normalised value equal to 100.

At the Big Innovation Centre, we consider that the knowledge extracted from these vast amounts of data creates unprecedented economic opportunities; ranging from disrupting the way we do everyday tasks to creating completely new product and service markets, new business models and new, more efficient ways of delivering public services. As the highly transformative power of the knowledge extracted from big data spans all industries, private and public sectors alike; we at the Big Innovation Centre consider big data a new general purpose technology – the second phase of the digital revolution.[2]

Conceptually, the economic opportunities brought to light by big data can be grouped in three categories, as depicted in Figure 2: new business opportunities, productivity gains and innovation.[3]

Figure 2: Types of economic opportunities created by big data



Source: Author's construction based on Sousa (2013).

In practice, however, the real-world examples of economic benefits extracted from exploiting the potential of big data is scant.[4] Practical projects like the big data hackathon discussed in this report give an important contribution to fill this gap.

---

[2] This was also recently acknowledged by Stephan Shakespeare in Shakespeare (May 2013).

[3] for a detailed discussion on the economic opportunities of big data, see (Sousa 2013)

[4] See: Einav and Levin (2013), OECD (2013), Mandel (2012), and MGI (2011).

2. Open innovation: the way forward in exploiting the big data opportunity

The economic opportunities summarised in Figure 2 are better exploited within an open innovation ecosystem where companies of different industries, public bodies, universities and other organisations come together to share and combine data and insights in order to address common challenges. Only by sharing, combining and opening-up data (once privacy, confidentiality and security are safeguarded), across industries and across sectors, we can extract the greatest economic potential of big data. This is better accomplished in a collaborative, open innovation ecosystem where all the parts involved benefit from the cross-fertilisation of ideas and data (Andersen and Wong 2013, Wong 2012, Perkman and Walsh 2007).

Embracing open innovation to break the data silos and exploiting the economic potential of big data is still in its infancy. The research, both theoretical and empirical, on how to make the most of big data[5] shows that for this shift to materialise we need to close three data-related gaps:

- Firstly, the real-world examples of the economic benefits extracted from big data tend to come from the private sector, within-silo data analyses. There are few examples of value created via combining and sharing data across companies of different industries and across public and private sectors. However, as Einav and Levin (2013) put it, the greatest potential of big data resides in sharing and combining data across industries and across private and public sectors;

- Secondly, most public bodies are lagging behind the private sector in making use of their own data. Public bodies have done little to reach out for insights and expertise on how to make the most of their own data. To be fair to the sector, there are a few good examples of public-led big data exploitation experiments (which will be discussed below), but these are the exceptions rather than the rule;

- Thirdly, the examples available so far show that, despite some recent progress, both private companies and public bodies are still largely unaware of the role that open innovation can play in extracting value from big data.

[5] Some recent examples include: Einav and Levin (2013), OECD (2013), WEF (2013), Willetts (2013), Mandel (2012), and MGI (2011).

3. The role of open innovation in unlocking the value of local authorities' data

Local authorities collect large amounts of data as part of their daily activities and they have been doing so for quite some time. Most of this data is unique because of its level of disaggregation, time span, and variety of subjects covered. Examples include micro-level (individual-, household-, or small businesses-level) data on social care, healthcare, social housing, schools, street cleaning services and street presence uniformed-teams (namely, parking enforcement and parking patrolling). This data has been largely under-used by public authorities. Due to limited and restricted access, local authorities' data has also been insufficiently used by private sector and universities, who could examine this data from a different perspective and uncover new opportunities.

Inspiring examples

Some forward-thinking local authorities across the UK and abroad are starting to look at this data as an asset to reduce costs and deliver new and improved services to tax payers. One approach has been opening-up local authorities' data to the public in the hope that researchers, companies, and universities will explore these datasets in ways that would help local authorities to deliver better services. Manchester, New York, Chicago, San Francisco, and Philadelphia are examples of cities around the world that have embraced, or who are about to embrace, an open data agenda. Some concrete outcomes have already come into light. For example, using the data made available by the city of New York, Ho (2012) found that restaurants hygiene grades had little consistency over time and suggested improvements for the city's hygiene inspection services. In Chicago, software developers have built on the city's open data to develop applications that show, for example, the status of requests to fix potholes or which streets have been cleaned after a snow storm.[6]

Local authorities across the UK and abroad are also learning how to use open innovation to build on the data they have available to develop new and better services. In the UK, Manchester is one of the earliest examples. As part of the Manchester open data project,[7] in November 2012 the city council organised a hackathon to explore the potential of the data made available by the city council and other public partners. According to the statement of

[6] These examples were found in an article published on *The Economist* (US edition) on 27th April 2013 titled "By the numbers: Cities are finding useful ways of handling a torrent of data" (http://www.economist.com/news/united-states/21576694-cities-are-finding-useful-ways-handling-torrent-data-numbers)

[7] This project was partially funded by the European Community's ICT Policy Support Programme.

the event,[8] several workable tools emerged from this experience including: (1) a visualisation tool that allows users to plot different councils' datasets on a map (e.g., location of road traffic accidents and speed cameras); (2) an application that allows people to find their nearest bus stop and its timetable; and (3) an application to find details about the nearest public toilet, such as its opening times.

In the US, New York is a good example. By looking at its historic data and merging data from different departments, New York City was able to pinpoint for example: (1) the houses with greatest probability of a catastrophic fire; (2) what businesses are more likely to be cigarettes smugglers and, as such, should be targeted by the cigarette-tax city's inspectors; or (3) what restaurants are more likely to be dumping grease down a manhole and thereby should be targeted by the restaurant inspectors.[9]

Although these are inspiring examples, this review shows that they are more the exception than the rule (Einav and Levin 2013). This is particularly true in the UK. In fact, as Lucy Watt and Quentin Jones noted in a *The Guardian* article on the 26th of April 2013,[10] UK councils seem to be increasingly keen on embracing open innovation and open data as a way of addressing local problems while coping with budget cuts. However, only few have put it into practice so far.

The case of Camden Council, London, UK

Camden Council in central London is one of these rare cases. A forward-thinking leadership recognised that the council had both the need to address challenges in a more cost-effective way, and an opportunity to do so by making better use out of the data they had available.

Recognising that this will be difficult to do alone,[11] Camden Council teamed up with the Big Innovation Centre, a London-based open innovation hub that brings together large companies from different industries, universities and several public bodies. By bringing together a unique private-public, cross-industry network, the Big Innovation Centre offers the sort of diversified pool of insights and skills that the Council need in order to learn how to make the most of its data through open innovation.

Rather than adopting a holistic approach towards a comprehensive open data agenda

---

[8] Available at: http://www.manchester.gov.uk/news/article/6484/inaugural_manchester_hackathon_h...

[9] These examples were provided in an article by Gillian Tett in an *Financial Times* article on 1 April 2013 (http://www.ft.com/cms/s/2/a284331a-9751-11e2-a77c-00144feabdc0.html#axzz2VG2hD0lV)

[10] Available at: http://www.guardian.co.uk/local-government-network/2013/apr/26/councils-hack-day-geek-squad-problem-solving

[11] This was acknowledged publicly by Camden Councillor Theo Blackwell in a comment posted to his Blog on the 13th of May 2013 (available at: http://theoblackwell.blogspot.co.uk/2013/05/camden-hosts-hackday-to-use-data-for.html)

(similar to that adopted by Manchester city), the Big Innovation Centre and Camden Council followed a hands-on, bottom-up approach. The approach started with a few practical problems and open the council-level datasets that could help address these problems. The lessons learnt from opening up and experimenting with a small number of datasets from different departments within the council could, then, be used to put together a more comprehensive council-level open data and data exploitation agenda.

For about five months, The Big Innovation Centre and Camden Council worked together to put in place a big data practical experiment based on the principles of open innovation, seeking to achieve four objectives:

- Deliver practical tools that allowed Camden Council to make more efficient use of its data and deliver great value for money to the local taxpayers;
- Identify opportunities for developing innovative services across public organisations and private companies from different industries ;
- Identify obstacles and problems in using, combining and sharing public and private data;
- Obtain insights into sensible policy recommendations in order to influence a big data-friendly policy agenda.

4. Methodology

The Camden Hack Day took place on the 10th of May 2013 at Google's Campus London and followed a 5-step methodology, as systematised in Figure 3 and explained below.

Figure 3: Methodology for an open innovation-based big data practical experiment

Source: Author's construction

Step 1: Define the practical questions

The first step in setting up the Big Innovation Centre-Camden council big data practical experiment was to define a small set of practical questions able to meet three criteria:

- Deliver better and more cost-effective services to the local taxpayers;
- Solve problems by exploring council-level data, possibly combined with other data sources;
- Be similar or complementary to questions faced by some of the Big Innovation Centre's partners, with a view at stimulating their participation in this practical experiment.

The combination of these three criteria resulted in the following three practical questions:

- What do Camden's future patients look like? What will they be treated for and how?
- How can the council target the repairs budget more effectively to deliver the

maximum benefit to its tenants (e.g., lower energy costs, greater energy efficiency, better ventilation, or a combination of these)?

- How can the teams target their efforts to reduce crime and anti-social behaviour?

Step 2: Get the data ready

The second step involved selecting available data in different departments of Camden Council, data pertaining to any of these questions. One major issue in getting the council data open to the public was to make sure that it was properly anonymised.

We also searched for datasets outside the council that could be combined with their data to answer these questions. We looked for publicly available datasets, as well as private data that could but could be made accessible once properly anonymised.

Numerous legal and internal hurdles made it impossible for the Big Innovation Centre's corporate partners to bring along any of their data. Legal hurdles also limited the datasets that Camden Council was able to bring to the event. All in all, the data used on the day came from two sources:

- Data made available by Camden Council, which included:
    - Housing stock condition and repairs data for about 33,000 properties;
    - Camden-wide prescription and illness data;
    - Data on ambulance call-outs by category, broken down to ward level;
    - Crime data by type of crime, at the street level;
    - Incidents in parks and open spaces at ward level.
- Publically available data at council-level with different levels of geographical desegregation, including:
    - Demographic and socio-economic census data by post-code sector;
    - Housing energy efficiency at council-level.

Step 3: Find the right mix of skills

The third step was to ensure that we would have the right mix of skills for exploring the data, which should include: innovation specialists from different industries, experts from Camden

Council, data engineers, programmers, data analysts and policymakers.

In order to accomplish this, we made use of the network of companies, universities and public bodies at the Big Innovation Centre. Many partners responded favourably and enthusiastically, so it was quite easy to find the right amount and combination of skills by relying primarily on the partners of the Big Innovation Centre. Nevertheless, we reached out for interested people outside this network. As we had more registrations than places available, among those registrations from outside the Big Innovation Centre's network we selected only those able to fill the skills gaps.

As a result, we had an eclectic group of around 75 people (selected from a pool of 130 registrations) with diverse skillsets: programmers, data analysts, data scientists, thoughtful leaders, policymakers, entrepreneurs, and graduate students from three types of organisations, mostly Big Innovation Centre's partners:

- Companies: Barclays Bank, EDF Energy, GlaxoSmithKline and Experian;
- Universities, including University of Oxford and University College of London;
- Public organisations: Intellectual Property Office, JISC, Islington Council and, of course, Camden Council.

Out of the 13 teams formed on the day, 5 worked on the Health problem, 3 on the social housing problem, and 5 on the street presence problem.

Step 4: Find a trusted hub

Having a trusted hub behind an open innovation-based big data practical experiment is important for two reasons. Firstly, skilled people are more likely to be willing to work collaboratively, share ideas and insights if they trust that their ideas and insights will not be misused. Secondly, organisations, both public and private, are easily prone to bring data to an open innovation practical event if they are sure that this data will be used for good. In this big data experiment the Big Innovation Centre was the trusted hub which made this event possible.

Step 5: Choose the open innovation tool

Among the open innovation tools appropriate for exploring the possibilities of big data, a hackathon was considered the way to go. A short, yet intensive team work effort was considered well suited to kick off a big data experiment aiming at examining large unknown data.

Figure 4: The Big Innovation Centre and Camden council hackathon on 10th of May 2013

The Big Innovation Centre–Camden Council hackathon, although sharing some features with other hackathons, was unique in several ways. Like other hack days, the Big Innovation Centre–Camden Council hackathon was an event in which programmers, data scientists, data analysts and innovation specialists worked collaboratively to address a few real-world challenges in a short period of time, which in this case was one day. Our hackathon differed from most similar events in, at least three ways:

- Firstly, the event pooled insights, skills and data from a full range of organisations, mostly Big Innovation Centre's partners. They included: (1) companies from a variety of industries; (2) world-class universities; and (3) public authorities, including the local councils of both Camden and Islington;

- Secondly, as the leading organiser, the Big Innovation Centre used its convening capacity to provide a trusted and neutral hub;

- Thirdly, this hackathon was also a policy-driven learning experience. The Big Innovation Centre aimed at using this practical experience as a way of identifying practical obstacles in opening, using, sharing, and combining private and public data in order to set sensible and evidence-based policy recommendations.

5. Lessons learnt

The major struggle with local authorities' data is not the size of the datasets. Rather, the real struggle is dealing with hundreds of small datasets most of them suffering from data systems legacy problems.

The results of the co-organised Big Innovation Centre and Camden Council big data hackathon demonstrated that an open innovation ecosystem is an effective way of understanding the issues around these disjointed bits of data and finding ways forward in unlocking their economic potential. The event also showed that a collaborative, open and innovative environment is a great learning opportunity for policy purposes. The lessons learnt from this practical experience can be grouped in 3 categories: (1) practical outcomes, (2) what is needed to break the data silos and (3) legal and regulatory hurdles.

5.1 Practical outcomes

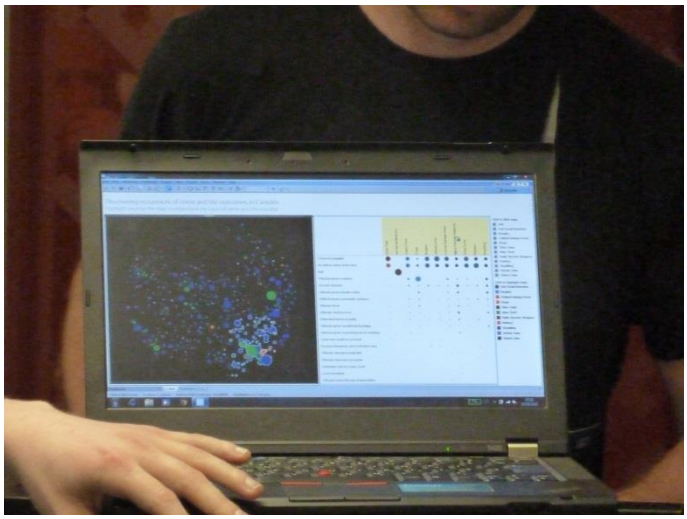
The practical outcomes from the hackathon organised by the Big Innovation Centre and Camden Council are summarised below.

- Camden council will be able to build upon the several workable tools developed on the day to address the three practical questions posed to the delegates on health, social housing and the management of street presence teams. These tools included:
    - A web-based simulation tool for social housing to identify the houses with greatest savings on their energy bills through low-investment repairs (lower than £10,000/year);

Figure 5: Social housing simulation tool

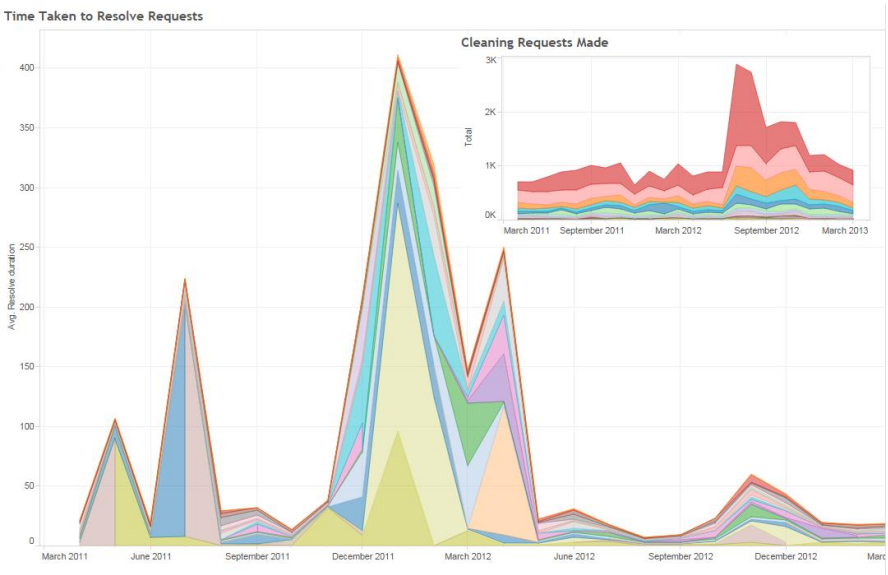
- A data model to identify the wards in which street presence teams should focus to deliver higher services quality without increasing their workload beyond realistic levels;

Figure 6: Data model to identify the wards that street presence teams should target



- A tool to demonstrate the performance of the Street Cleanliness teams;

Figure 7: Performance tool for the street cleanliness teams

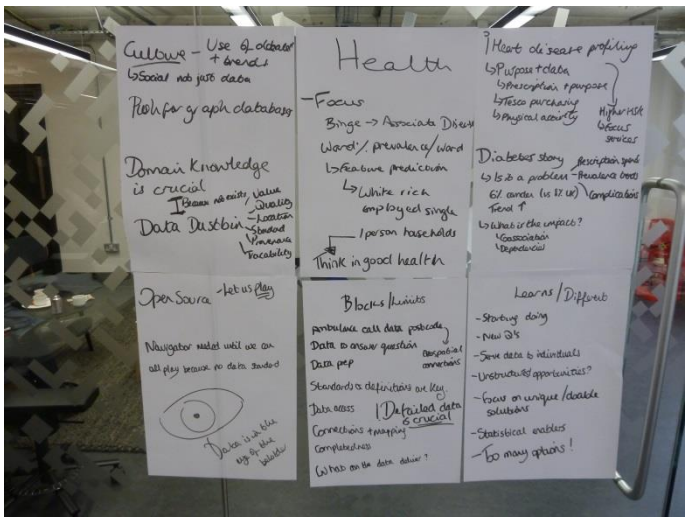


- An application for the Street Presence Teams to carry and identify licensing infringements by a venue;

Figure 8: Application for the identification of venues' licensing infringements

- A framework to identify the key health issues that Camden council's health services will need to be prepared to address in the next 10-15 years.

Figure 9: Framework to identify the key health concerns in 10-15 years' time



- Camden council's staff members considered the event a great learning experience because they were able to:

  - Get new insights into the utility of the data they gather on a regular basis;

- Learn about diagnostic and analytical tools they were not aware of;
- Get a practical understanding of what data quality issues caused greatest difficulties to exploitation of data
- Collect practical suggestions on realistic ways of fixing these issues (more on this below);
- Develop new conceptual ways of addressing the practical questions the teams were committed to address on the day. This was possible because each team included delegates from companies that work on similar problems regularly. Their experience and insights were crucial to frame the problems, develop practical ways of approaching them and suggest different courses of action when the analysis hit a roadblock.

- Companies and entrepreneurs were able to:
  - Identify the sort of data collected by councils;
  - Explore potential new economic opportunities coming from crossing the council data with corporate data.

5.2 What is needed to break the data silos

Just like any other local authority(as well as many companies), Camden Council's legacy systems of collecting and storing data, many of them several years old, created serious challenges to the data analysis process. Historically, the council data has been kept in separate and incompatible files, within departments as well as across different ones. In many cases, data sets were incomplete and missing key variables as, for example, time range or the geographical unit.

The hackathon showed that to break the data silos Camden Council has to:

- Identify connectors among datasets and make them available via search tools. For example, in order to link datasets at the street level (latitude and longitude), 6-digit postcodes, 4-digit postcodes, postcode sectors, wards, and council level; it is necessary to make available a user-friendly search tool which allow users to easily aggregate lower-level data into higher and compatible geographical levels;
- Conduct some data audit and data exploitation to assess the quality and completeness of the datasets. This should include:
  - Ensure consistency of geography, as much as possible;

    - Ensure consistency of format, as much as possible;

- Ensure that newly collected data is complete and fill-in missing key data in current datasets (wherever possible). Simple examples are:

    - Include the appropriate geographical locator (e.g., street address for crime data, postcode, postcode sector, wards,…) in future data records; fill-in the missing geographical locator for those datasets where it is cost-effective to do so;

    - And do the same with time ranges;

- Store the data in standardised format files as, for example, relational databases and web-based access services.

5.3 Legal and regulatory hurdles

Releasing data for the day proved to be a very hard task, more so for private companies than for Camden Council.

Camden Council was able to release some properly anonymised data. However, the Council was unable to release some relevant datasets that could directly or indirectly have been useful on the day. The Council's data was also only temporarily open, which meant that participants had access to the data on a temporary data cloud and the data was deleted at the end of the day. Every delegate signed a Non-disclosure Agreement where they committed themselves to delete the data from their own computers before leaving the room.

Broadly speaking, the major hurdles that both companies and Camden Council faced were around purpose specifications and use limitations. The current rules and regulations, namely the UK Data Protection Act, require the purposes and type of data uses to be specified when individuals provide consent for data use. Therefore, a new or different type of data use would require companies and the council to obtain a new consent from each individual to use the same data previously collected.

6. Conclusions and policy recommendations

The Big Innovation Centre–Camden Council hackathon proved that both companies and members of the public are very keen on having access to UK council data. It creates a win-win situation for all, and a great opportunity which local authorities cannot afford to miss. Local authorities are able to find more cost-effective ways of delivering their usual services as well as developing new ones. Local tax payers receive greater value for money. Companies and entrepreneurs are also able to develop new and better services for their customers by exploring the rich local-level granular data.

For all of these sweeping benefits to materialise, policymakers need to address several issues. Below we highlight the most prominent actions that both local-level and national-level policymakers should take.

6.1 Recommendations for local-level policymakers

Policymakers at local-level should:

- Push for an open data and open innovation culture among the councils' staff members. Councils' data tend to be closed in departmental silos, which need to be broken down. This might require, like in the case of New York,[12] that council leaders make open data and breaking departmental silos an explicit priority;

- Work with other local authorities and also with the central government in setting forward common standards on (1) data anonymisation, (2) data quality and (3) data release formats. The recently published Shakespeare Review created the right momentum for this to be done, by calling for the UK to define a National Core Reference Dataset of all public-sourced data (Shakespeare 2013). Yet, as the author recognises, what is missing now is the definition of what a Reference Dataset should look like. The best way of accomplishing this is by learning from experiences like our big data hackathon with Camden Council;

- Open local public data. The benefits we can reap from data are only as great as the amount, quality, and richness of the data that general public, entrepreneurs and companies can get access to. Some existing government data initiatives[13] are a step in the right direction but more needs to be done.

[12] http://www.ft.com/cms/s/2/a284331a-9751-11e2-a77c-00144feabdc0.html#ixzz2VG6zVGKk
[13] One example is: http://data.gov.uk/

6.2 Recommendations for national-level policymakers

Policymakers at national-level should:

- Put in place a big data-friendly legal and regulatory framework for using, combining, and sharing data without compromising privacy, safety, and intellectual property rights. This is not a straightforward task. The lessons learnt from this hackathon, coupled with the lessons learnt from a recent debate on the topic organised by the Big Innovation Centre[14] suggest some possible ways forward:

    - Rethinking the traditional approach to data governance, shifting from the focus on controlling the data itself to governing the uses of data.

      The traditional data-protection approach where individuals consent on specified purposes, usually at the time of data collection, is no longer fit for purpose. This traditional approach fails to account for the nature of today's hyper-connected data architecture. It also fails to account for the opportunities created by allowing data to flow and combine with other data. A new approach which enables value-creating data purposes but prevents intrusive and damaging ones is needed. This requires putting in place a set of lastly permissions, controls and trustworthy data practices to govern the uses of data rather than the data itself;

    - Exploring whether the UK should build on the Creative Commons legal infrastructure to develop a Privacy Commons;

    - Exploring whether the UK should adopt an 'opt-in unless you opt-out' approach to personal data disclosure;

    - Ensuring that the regulatory framework for big data avoids loopholes, is easy to understand and is easy to comply with.

- Support the development of new platforms that allow merging and combining datasets from companies, universities and public bodies without putting privacy at risk. These platforms are crucial to carrying out trusted big data 'test-bed' experiments. The greatest economic potential of big data comes from cross-linking data from different organisations, public and private. But this can only be accomplished within a secure environment where companies, entrepreneurs, academics and public organisations can share and experiment with public and

[14] See, for example, the conclusions of the debate co-organised by the Big Innovation Centre and the Economic and Social Research Centre as part of the Big Data Week 2013 event at: http://us4.campaign-archive2.com/?u=f48c74d9cc73a25267a7bc364&id=2ef911193d

private data without fear of losing commercial advantage or breaking the law. Like other practical experiments with big data, trusted 'test-beds' are also effective ways of learning how a fit-for-purpose legal and regulatory regime for big data should look like;

- Open-up data across all public bodies, particularly publicly funded data. The UK has some of the world's best and most complete public datasets in, for example, health, demographics, agriculture and meteorology (Willetts 2013). For the UK economy to reap the benefits of these and other important public datasets, the most innovative and creative companies and entrepreneurs have to have full access to the data. Only then can companies and entrepreneurs build on the opportunities of public data to develop new market-ready products, services and organisational methods, thereby promoting economic growth and job creation;

- Ensure that the right digital infrastructure is in place. Building effective and reliable digital networks is a precondition for making the most out of big data. This includes:

    - Investing in broadband capacity, 4G wireless networks and data storage capacity;

    - Ensuring digital networks are safe and secure;

- Adjust the education policy to tackle the potential skills mismatch. Making the most out of big data calls for a labour force able to analyse, interpret and put the insights extracted from data to work. The UK education policy must make the necessary adjustments to deliver:

    - The next generation of data scientists and data analysts;

    - Mid-level professionals able to interpret and make meaningful use of the data analysis results.

  A failure to adapt to these skill requirements will limit the UK's ability to capture the full potential of big data. It can also raise structural unemployment in the UK.

References

Andersen, Birgitte and David Wong. 2013. *The new normal: Competitive advantage in the digital economy*, February, London: Big Innovation Centre. Available at: http://www.biginnovationcentre.com/Assets/Docs/The%20New%20Normal.pdf

Einav, Liran and Jonathan D. Levin. 2013. The Data Revolution and Economic Analysis. *NBER Working Papers No. 19035*, May. Available at: http://www.nber.org/papers/w19035

EMC^2. 2013. *EMC 2012 Annual Report*. Available at: http://www.emc.com/corporate/annual-report/big-data.htm (accessed 31st May, 2013)

Ho, Daniel E. 2012. Fudging the nudge: Information disclosure and restaurant grading. *Yale Law Journal* 122.

Mandel, Michael. 2012. *Beyond goods and services: The (unmeasured) rise of the data-driven economy*. Washington, DC: Progressive Policy Institute, October. Available at: http://www.progressivepolicy.org/wp-content/uploads/2012/10/10.2012-Mandel_Beyond-Goods-and-Services_The-Unmeasured-Rise-of-the-Data-Driven-Economy.pdf

MGI. June 2011. *Big Data: The next frontier for innovation, productivity and competition*. Mackinsey Global Institute, May. Available at: http://www.mckinsey.com/insights/business_technology/big_data_the_next_frontier_for_innovation

OECD. 2013. Exploring data-driven innovatopn as a new source of growth. *OECD Working Papers DSTI/ICCP(2013)9/Final*, April. Available at: http://search.oecd.org/officialdocuments/publicdisplaydocumentpdf/?cote=DSTI/ICCP(2012)9/FINAL&docLanguage=En

Perkmann, Markus and Kathryn Walsh. 2007. University–industry relationships and open innovation: Towards a research agenda. *International Journal of Management Reviews* 9 (4): 259-280.

Shakespeare, Stephan. 2013. *Shakespeare Review: An independent review of public sector information*. London: Department for Businesses, Innovation & Skills, May. Available at: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/198752/13-744-shakespeare-review-of-public-sector-information.pdf

Sousa, Sonia. 2013. *How should the government approach the big data challenge? Realising economic opportunities and building an enterprising state*. London: Big Innovation Centre, February. Available at: http://www.biginnovationcentre.com/Assets/Docs/Big%20Data%20challenges%20for%20gov't-ed28Fb.pdf

Willetts, David. 2013. *Eight Great Technologies*. London: Policy Exchange. Available at: http://www.policyexchange.org.uk/images/publications/eight%20great%20technologies.pdf

Wong, David. 2012. *Data is the Next Frontier, Analytics the New Tool*. London: Big Innovation Centre, November. Available at: http://www.biginnovationcentre.com/Publications/21/Data-is-the-next-frontier-Analytics-the-new-tool

WEF. 2013. *Unlocking the value of personal data: From collection to usage*. World Economic Forum, February. Available at: http://www3.weforum.org/docs/WEF_IT_UnlockingValuePersonalData_CollectionUsage_Report_2013.pdf

Acknowledgements

This report is a publication from the Big Innovation Centre, an initiative from The Work Foundation and Lancaster University. The content of this report reflects the opinions of its authors and not necessarily the views of the Big Innovation Centre or its supporters. The Big Innovation Centre is supported by the following companies, public bodies, universities and private trusts.

Contact details

Big Innovation Centre
The Work Foundation
21 Palmer Street
London SW1H 0AD

info@biginnovationcentre.com
www.biginnovationcentre.com
www.theworkfoundation.com