Data and Analytics: Resources for Smart Cities

A report by the Data and Analytics Task Group for the Smart Cities Forum

'Information is the oil of the 21st century, and analytics is the combustion engine.'– Peter Sondergaard, SVP, Gartner Research. Speech given at Gartner ITxpo.

1. Introduction

This report has been prepared by the Task and Finish Group that is considering "Data and Analytics as resources for Smart Cities", as part of the Smart Cities Forum. Cities are both producers and key consumers of data and analytics due to their role as hubs in the information economy.

Cities are systems with many different functions; for example, transport, trade, administration, innovation, economic and social systems, and culture. They also simultaneously function as part of a number of complex, interdependent systems at different scales; regional, national and international, for example.[1] A 'smart city' therefore, might be best thought of as being 'the effective integration of physical, digital and human systems in the built environment to deliver a sustainable, prosperous and inclusive future for its citizens.' [2] To this end, smart cities mobilise many new technologies, particularly those that are digital, in reaching towards these goals.

The aim of the Forum is to enable Government to understand how it might best support the deployment of smart services to improve life in our cities and to produce export and investment opportunities. This Task Group report offers a short summary of the key challenges and opportunities that 'data and analytics' might present for cities in the future and a set of corresponding recommendations for Government.

2. Methods and approach

The Task Group is made up of representatives from the Smart Cities Forum together with external representatives from cities, academia and industry. This interdisciplinary group reviewed evidence around data and analytics, categorising different kinds of data and the range of analytics needed to facilitate their use and explored the range of users, uses and values. A range of city-relevant case studies was then developed to provide illustrations as to how this applies to an urban environment. The group used this evidence to consider what this might mean for cities in terms of challenges and opportunities for the future bearing in mind the

[1] Foresight Future of Cities, Progress Report, upcoming Jan 2015
[2] British Standards Institution

limitations, uncertainties and inaccuracies of data[3]. Data and technology can never give us "perfect knowledge"; however providing these limitations are accounted for, they give us a better ability to draw insight from and take decisions based on a wider range of sources and more timely data.

A brief summary of the evidence is contained in Annex A

3. Data and analytics: implications for urban living

New data is being generated on an enormous scale and nowhere more so than in cities. Many everyday activities and interactions within urban environments can be harnessed to generate a variety of data, ranging from apps and social media to satellite imagery. It is not just the quantity but also the veracity of data which is increasing – emergent information technologies show what people do, not just what they say they do. However to make sense of this vast data-flow, it is neccesery to develop better ways to analyse it. Just as important then as new data-generating technologies are emergent new methods of analytics, such as machine learning, data mining, pattern recognition, profiling, simulation and optimization algorithms. Cities can increasingly be seen as a key challenge – a modern-day 'enigma code'. Harnessing data and analytics to make sense of the patterns within this code and thus cracking the 'science of cities' is a key future challenge.

A key new source of data emerging is the Internet of things (IoT). Cities will be a major test-bed for the proliferation of the IoT on the managing, planning and operation of assets. The Internet of Things is essentially the general phenomena whereby digital networks are connecting up infrastructure, appliances and people through various digital devices. It is estimated that 50 billion such devices will be connected by 2020.[4] There are three main types of IoT interaction:

i. Machine-to-machine (e.g. sensor-controlled mechanisms)
ii. machine-to-people (e.g. online purchases, RFID cards such as 'Oyster' cards)
iii. People-to-people (e.g. through Skype, Twitter etc)

These interactions have the potential to transform how public, private and community services are delivered and how people interact with each other and their environment.

The capabilities to analyse and use this data and to act on the resulting insight must be in place before this vision of the future can be realised.

[3] '11 reasons computers can't understand or solve our problems without human judgement'. *The Urban Technologist*. Available at http://theurbantechnologist.com/2014/09/07/11-reasons-computers-cant-understand-or-solve-our-problems-without-human-judgement/

[4] Future Cities Catapult: Leading Innovation in the Internet of Things

The move towards increased data gathering, accumulation and ubiquitous connection is a global trend in which Government faces a challenge to be proactive rather than reactive. Government also has a role in ensuring that privacy, trust and ethical standards are adequately safe-guarded as the effects of this trend are felt.

New sensor technology deployed at a range of city scales is going to yield ever-increasing amounts of data from existing and new sources which analytics can mine for intelligence and provide real time situational awareness. We are in the early stages of understanding the processes that could extract value from this data - although demonstrators in cities in the UK and around the world are providing some insights. From autonomous vehicles to intelligent lampposts, to industrial bins which tell the relevant Local Authority when they are full[5], to using online journey planners to inform transport timetabling (see Annex A for details), cities are already successfully demonstrating the value of data and analytics. While networked urbanism has been an evolving phenomenon for over 30 years, the velocity, quantity, and finer granularity of these new forms of data is greater than ever before, and so too is the potential impact on cities.

For this report, the questions are: what are the untapped, undiscovered opportunities which could emerge in the future (section 4)? And what are the challenges created by this type of approach being taken up in all cities, across a range of different services (section 5)?

4. Data and Analytics: opportunities for cities

i. Cities provide a rich source of data on multiple scales and this will increase in the future as the Internet of Things and other data sources become more widely available.

Cities, as dense concentrations of interactions, provide a significant opportunity to gather data, which can then be analysed and used to improve urban living - for example, new transport 'apps' are making cities easier to navigate. There is also an explosion in the ways citizens are generating data, from the use of Amazon market place to the use of smart cards on transport. Increasingly 'apps' are including real-time data from users to improve their service. Sensors in the fabric of buildings and the environment will increasingly supplement people-derived data. In this way, cities provide both the resources (data) and environment to test new approaches at different scales. These approaches have the potential to offer considerable potential benefits for citizens – allowing them to be healthier, happier and more productive – if they are used in applicable and relevant ways.

ii. Improving the efficiency and effectiveness of service provision in complex, interdependent environments.

[5] Future Cities Catapult: Leading Innovation in the Internet of Things

Analysis of data at the local and national level can provide key opportunities for cities to manage resources and provide services more efficiently (see Annex A for specific examples). Greater availability of data, new analytical capabilities and computational capacities will significantly increase the resolution of available information on the performance of city systems and dynamically predict and optimally respond to situations. In this context, new as well as traditionally enriched data sources will provide the momentum for cities to embrace three key opportunities:

- To **plan** better new places, using predictive analytics to forecast the social, economic and environmental impacts that new development will have. City planners will be able to address concerns by pointing to real evidence of benefit. For instance, analytics could show how well connected, mixed-use development can have effective health outcomes - offsetting the increasing burden of obesity on health budgets; or demonstrate how a new development can ease congestion by reducing longer-distance commuting. These approaches could in the long run help tackle intra-city social and health outcome disparities.

- To **construct** new development more efficiently – organising the flow of materials to and from building sites; to build more efficiently and at a higher quality.

- To deliver services and **manage assets** more efficiently – understanding what resources exist and how they can be better governed and managed. For instance optimising and co-ordinating schedules for public work; managing traffic congestion and routing more effectively through linking in-vehicle technologies to traffic signalling; harnessing solar and wind power; using energy in buildings more efficiently both by being aware of how and when it flows and also by flexibly adapting demand on the basis of an understanding of demand and load patterns; enabling new forms of public transport services able to harness data and analytics to adapt flexibly to user demand

In all of this, urban analytics can allow Government to make proactive, preventative interventions earlier than previously possible: for example, by investing in analytics to enable a pro-active, preventative approach to helping children who encounter difficulties in education, Medway Youth Trust estimate that £3,200 of lifetime benefits have been created for every £1 invested in technology[6]. In this way, data and analytics can pervade the everyday operations of every sector of employment and of society generally, with benefits flowing in multiple streams, such as economic output, social vitality and environmental enhancement.

[6]See http://www.computerweekly.com/news/2240036440/Medway-Youth-Trust-uses-text-analytics-to-fight-youth-unemployment and http://www-01.ibm.com/common/ssi/cgi-bin/ssialias?subtype=AB&infotype=PM&appname=SWGE_YT_YV_GBEN&htmlfid=YTC03359GBEN&attachment=YTC03359GBEN.PDF#loaded]

iii. Urban analytics will enable increasingly integrated, responsive modes of urban governance

Increasing availability of data will change the role that citizens can play in the design and operation of their cities. The provision of open data platforms will not only increase the levels of transparency and accountability of city governance, but will also facilitate more direct forms of engagement and involvement in decision-making. It will also enable reactions to policies and announcements to be gauged closer to real-time, feeding citizen opinion and behaviours into more open and arguably democratic approaches to policy-making. The current debate surrounding devolution has seen a broad consensus emerge that cities need greater autonomy; however how best to achieve this has not been clear. The very act of generating data which informs policy will ensure that citizens have new opportunities to influence Government decisions.

iv. Increased urban resilience

Having better information and being able to use data to inform decision-making more effectively will help cities prevent, or at the least recover quicker from environmental shocks, such as air pollution and flooding, as well as identifying severe threats to their functioning. Increased resilience with respect to pollution and weather could help cities cope with climate change much more effectively in the decades to come. Moreover, cities face important challenges with respect to how they develop ever better ways of interaction and movement. Coordinating physical vehicles by controlling and integrating demand and supply is advancing apace as information is delivered to controllers, drivers and passengers that enable them to respond intelligently to potential disruptions. New ways of empowering local communities through the delivery of information in real time via the web or through apps, email, social media while drawing responses to problems and crises through the crowd all provide ways in which cities might be made more resilient. Terrorist activities can also be monitored and countered using various big data that is associated with new ways of sensing the unusual activities, notwithstanding the enormous privacy concerns that are raised and must be addressed.

v. Revealing hidden synergies

In recent years dramatic improvements in information and communication technology - especially social media, mobile devices, e-commerce and analytics - have made it easier for people and organisations to make contact and interact. Information about supply and demand has become more freely available; and it is increasingly easy to reach consumers through online channels.

In response, online peer-to-peer marketplaces have emerged to compete with traditional models of business in many industries – Apple's iTunes famously changed the music industry in this way; YouTube has transformed the market for video

content, e-Bay and the 'freecycle' movement have transformed markets for second-hand goods, Zopa have created a market for peer-to-peer lending, and "sharing economy" businesses such as Lyft and Air b'n'b are proliferating. As technologies such as 3D printing and small-scale energy generation improve, these ideas will spread to other industries as it becomes possible to carry out activities that previously required expensive, large-scale infrastructure at a smaller scale. It is certainly true that policy and regulatory environments need to adapt, as highlighted by the recent independent report for governance. But nevertheless, these themes represent a significant opportunity for social and economic growth enabled by data.[7]

5. Data and Analytics: challenges for cities

The challenges facing UK cities are enormous and pressing, including those associated with climate change, demographic change, economic development, and resource scarcity. How will we provide infrastructure for future cities as resources become scarcer and energy more expensive, as our population ages, and the climate becomes less predictable? How will cities cope with the risk of increased flooding and disruption? How will we generate new jobs in an era when global productivity appears to be declining and economic polarization increasing?[8] What will future cities look like in the coming decades, and how will different rates of growth affect this? Will work and living become hyper-local as knowledge workers leave the office, and goods and services move with them? How will car ownership be affected by the likely arrival of autonomous vehicles? How will social behaviour change?

For cities to realise the potential of data and analytics to address these challenges, the following issues need attention:

i. **Access to data** needs to be improved and more widely machine-readable, so that more relevant data is open to decision makers across sectors. Too often by the time data is received it is out of date. Much of the data that describes cities is locked in private systems and may not be in a readily accessible or usable form[9]. Private utility and transport companies have data which would be useful to local authorities and other service providers, but are often reluctant to share, even when the results could feasibly deliver benefits for all Government needs to engage with the private sector to unlock its vast data supplies where they are relevant to communal and collective action without

[7] 'Unlocking the Sharing Economy: Independent Review'. BIS. https://www.gov.uk/government/publications/unlocking-the-sharing-economy-independent-review
[8] http://www.worldfinance.com/strategy/corporate-governance-strategy/a-global-productivity-slump
[9] 'Why Open City Data is the Brownfield Regeneration Challenge of the Information Age'. *The Urban Technologist.* Available at http://theurbantechnologist.com/2012/10/18/why-open-city-data-is-the-brownfield-regeneration-challenge-of-the-information-age/

compromising competition. Some existing initiatives are already showing the potential of sharing data across sectors[10] – but much more remains to be done. Even when open, compatibility and inter-operability issues often limit the sharing of data between organisations. Common standards are needed to make open data of practical use.

ii. **Analysing and acting on Data –** it essential to capture and analyse the right data at the right time for city decisions to be effective. The 'science of Cities' is still in its infancy, and the models produced are currently very simple and primarily descriptive rather than predictive. There is much work to be done over the coming years before urban analytics produces useful models to inform policy-makers. In the mean-time, decision makers need to know how to interpret data and where data-based interventions will be most impactful in order to act on this analysis (see point iv. below). Conflicts will arise with new services and changes in behaviour from the deluge of data. An agile response that does not choke opportunity or innovation is needed.

iii. **Ethics, Privacy and trust issues –** increased data use will impact on citizens and their willingness to embrace smart city initiatives. As more and more aspects of daily life are captured as data, privacy concerns are likely to grow. Laws and regulations guide organisations, particularly around privacy and the use of data, defining the current "no-go" areas for an organisation. However, recent advances in analytics and big data technology have widened the gap between what is possible and what is legally allowed, changing the balance of power between citizens and the data collectors. Within this gap are new opportunities alongside the risks of public relations disasters and unintended consequences. And it is within this gap where the ethical questions around what is acceptable are raised[11]. NHS England's postponement of the launch of Care.data due to public fears is an example of how this may present barriers for city initiatives. To allay these concerns-, new governance regimes and ethical codes for data use will have to be created.

iv. **New security threats –** While the IoT has the potential to make cities smarter, it also runs the risk of making cities 'buggy, brittle and hack-able' (Kitchin, 2014)[12]. Viruses, malware and hacking are all increasingly dangerous as more and more urban infrastructure relies on networked connections. This could make the 'smart city' hugely vulnerable to systems errors and cyber-attacks unless security becomes a key priority. Government has a role to play in ensuring robust and large-scale IoT infrastructure is put in place.

[10] For instance, the Administrative Data Research Network - http://www.adrn.ac.uk

[11] *Ethics for big data and analytics.* IBM White Paper. www.ibmbigdatahub.com/whitepaper/ethics-big-data-and-analytics

[12] Kitchin, R. 2014. 'The real-time city? Big data and smart urbanism'. *GeoJournal,* vol 79.

v. **Skills/Capability** in both local and central government and in the private sector need to be enhanced – being able to assess the long term value and impact of smart city initiatives as well as coordinate work across many players, both public and private (for example, access to mobile phone data). New professions and roles will need to emerge, including urban analysts and data-driven urban practitioners with access to predictive analytics, and existing professionals, from asset managers and investors to artists will have to update their skills in order to use data to inform their work. In their report "Cities Outlook 1901", the Centre for Cities surveyed the economic performance of UK cities relative to the size of their population throughout the 20th Century[13]. They found that the most strongly correlated influence on this performance was the ability of cities to provide their population with the skills most in demand as technology and the economy evolved over time. In the early 21st Century that challenge will be dominated by data and technology and citizens who do not have the skills to exploit them will be excluded from many of the opportunities of the era.

vi. **The velocity and value offer of data –** increasingly the market drivers for data have shifted from prioritising volume and variety of data to velocity and veracity of data. Data in which the user can have a high degree of confidence is needed almost in real-time to be used effectively. The increased velocity available from advanced analytics will have major unintended repercussions, such as making some existing jobs obsolete, if adequate skills and capabilities are not put in place[14]. While traditional data analysis will become increasingly automated, jobs concerned with manipulating and making inferences from data will likely be created to compensate, if the requisite skill-base is in place.

[13] *Cities Outlook 1901.* Centre for Cities. http://www.centreforcities.org/research/2012/07/12/outlook1901/
[14] Forrester's *15 emerging technologies to watch before 2020*. blogs.forrester.com/brian_hopkins/14-10-31-forresters_top_emerging_technologies_to_watch_now_through_2020

6. Recommendations:

i. **Recognising Interdependencies** – much quality work is being done applying data and analytics to make cities smarter, but the data landscape remains highly fragmented, with many blind spots between institutions, disciplines and data-sets. Cities are sites of complex interdependencies, and it is in this context that the below recommendations must be considered, so that possibilities to collaborate and partner are realised, and ideas do not fall through these gaps.

ii. **Opening and sharing data** - The Government should continue to develop its open data policy and to facilitate and incentivise others to contribute on an appropriate basis. A suitable Government agency -perhaps the UK Statistics Authority (UKSA)or the Office for National Statistics (ONS) should maintain an overview and catalogue of applications to facilitate the sharing of information and expertise. It should facilitate a watching brief on data that is collected by the private sector with a view to mobilising that data that is relevant to public policy.

iii. **Giving cities a voice in the IoT** -The Internet of Things Advisory Board (proposed by the Government Office for Science Internet of Things Review) should have a representative from the Task Group or Smart Cities Forum to ensure that an urban perspective, and the opportunities and challenges it presents, are taken into account. The UK needs to keep pace with other nations in the development of large-scale test beds of IoT (e.g. SmartSantander[15]). Developing urban IoT test beds requires aligned research and development investment. These test beds should be linked directly to incubators and targeted investment programs for overcoming barriers to implementing IoT in the urban environment.

iv. **Fostering privacy, trust and security** - The Government should ensure that an appropriate body – the UKSA and/or ONS – take responsibility for the oversight of privacy issues associated with data availability. Moreover, with the increasing use of IoT in the physical realm, cyber physical security must become a key area of research and investment.

v. **Delivering the skills and capability to analyse and act on data** - The Government should ensure that, in the public sector and beyond, there is sufficient expertise to provide the analytics that will ensure maximum business use of data. It should provide advice on the quality of new data sources that are very different from traditional ones. In all sectors, the Government should encourage learning and skills generation, from the primary school to the

[15] http://www.smartsantander.eu/

workplace, to avoid a deepening of the ‘digital divide’. Emerging information technology has the potential to address social and economic disparities, but only if positive action is taken to ensure everyone has the ability to benefit from it. The government should consider incentive schemes for the creation of Chief Data Officers within entities that act as a focal point for the opening up and analysis of data.

vi. **Delivering long term solutions at pace** - Given the prominence of capability as a key barrier, it is proposed that the Task Group focus on work that will enhance understanding of the long term value of smart city projects – how to make cities and their citizens smart in the long run. Notwithstanding the above, the focus should be on identifying the highest value requirements first and delivering the narrow set of capabilities needed to realise the benefits whilst having an eye how the overall architectural and analytical vision will be realised. In other words, cities need to work out what is needed and do it quickly, but in a way that the next requirements can build on progressively to avoid reworking the vision later.

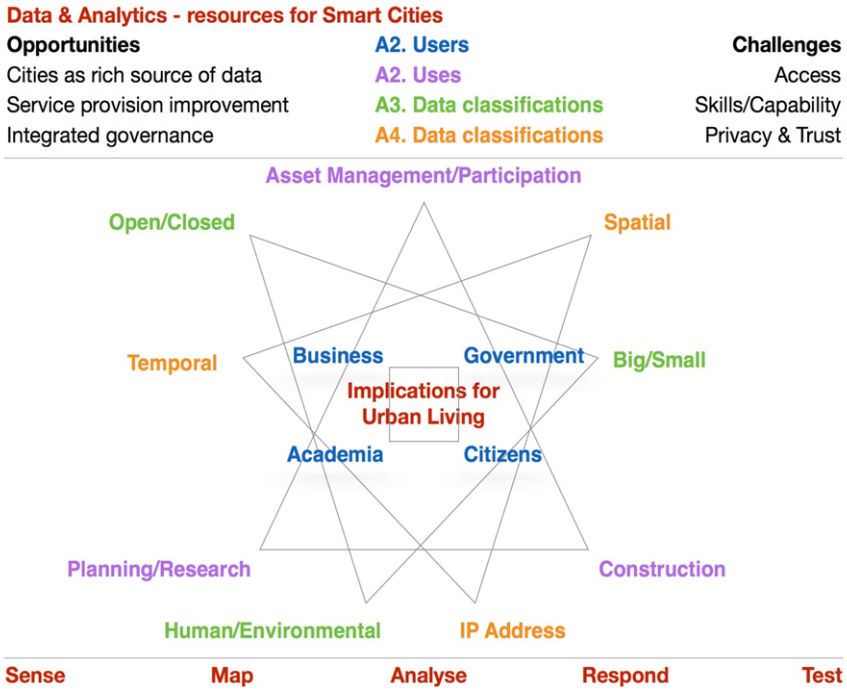
Annex A. The data and analytics challenge: a structured approach

A1. Outline

'Big data' refers to the increased volume, velocity and exhaustiveness of data being generated from a great variety of new sources that are now available and which will continue to expand in the future. This is data that is streamed from sensors and mobile devices in real time as well as traditional data usually pertaining to individuals that is collected through survey.[16] It is big in that it requires new tools to extract structured meaning and patterns from it. There are many facets to this picture and in this Annex we aim to provide a structure to enable conclusions to be drawn about the means of taking full advantage of this development. We begin by categorising users and uses and then classify the types of data available.

To be useful, data has to be processed in various ways for different circumstances and we review the analytics that underpin the task of driving value from data. There is already a wide range of uses of new data, making formal classification very difficult and we proceed by offering a range of case studies. Smart cities as we define them here, of course, are driven by data availability and associated analytics.

This methodology is represented in the conceptual framework below:

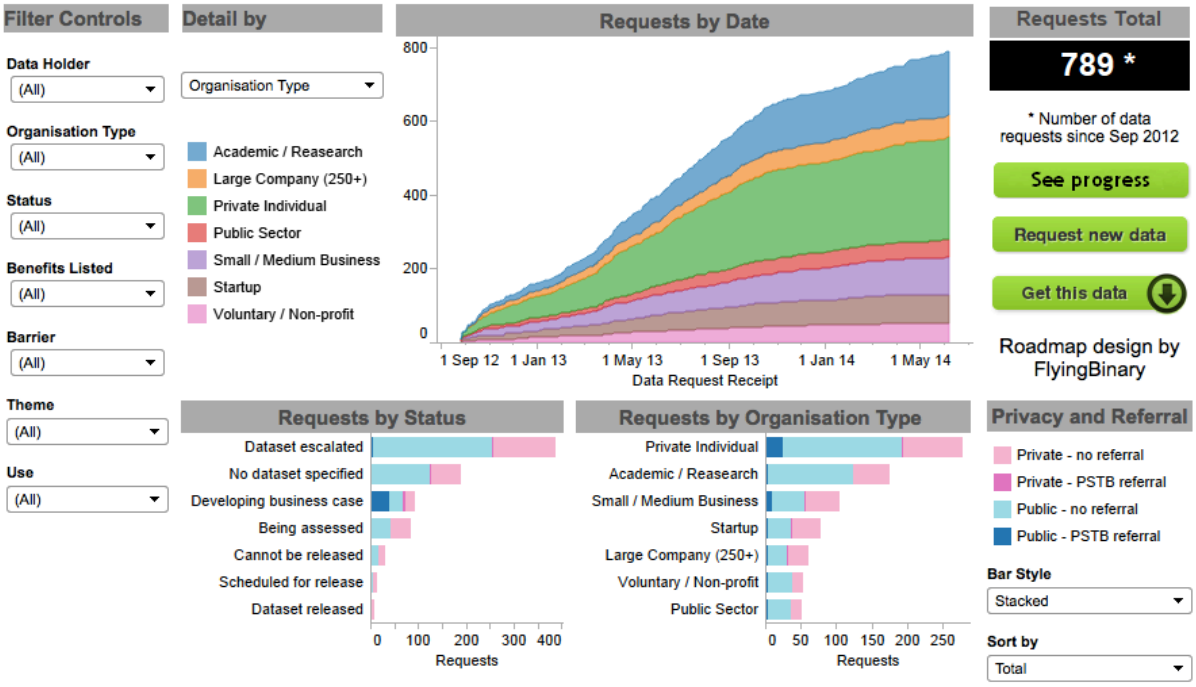


[16] Kitchin, R. (2014) *The Data Revolution: Big Data, Open Data, Data Infrastructures and Their Consequences* by Sage Publications, London.

A2. Users and uses

The four key categories of data users considered by the report are Government, citizens, business and academia. Government (at all levels) has a range of management and planning applications, as do businesses. Academia has an expanding research base, as do many businesses in the technology sector. Citizens' uses have been characterised as 'participation' and this in part reflects the possibility of wider engagement with government but it also embraces the contribution of data in day-to-day living.

The open data user group monitors the number of data requests from these categories of data users showing increases across all categories, with the larges categories being research, private individuals and small businesses/ start-ups.



Source: ODUG[17]

The Future Cities Catapult has conducted 18 ethnographic interviews with a variety of data users from across sectors, exploring their current and potential usage of city data. This research focuses on the barriers for the categories for the use of data. The research shows many of the cultural and behavioural issues that need to be considered in the development of skills and driving an evidence based decision culture within cities:

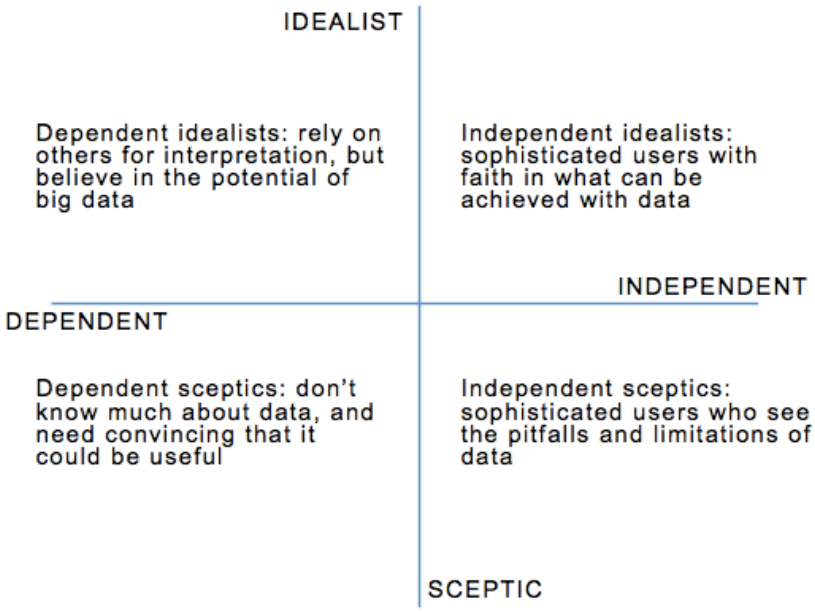
[17] http://www.slideshare.net/jacquitaylorfb/final-bristol-deck?utm_source=slideshow02&utm_medium=ssemail&utm_campaign=share_slideshow_loggedout

- **Faith in data**: the extent to which they believe in big data, and the potential to change things using data, versus how sceptical they are about the limitations of data
- **Sophistication in data use**: the extent to which users can obtain, make sense of and analyse data
- **What's in it for me?:** the extent to which users believe that data can help them
- **Need for interpretation**: the degree of interpretation they require from data
- **User or provider**: are they are a data user, or a data provider, or a mediator in the journey between the two
- **Head or heart**: whether a user prefers to act on hunches, or empirical data

There's a strong sense that city data is a growing trend, which users believe will have a major impact. Three challenges to achieving value from the data are identified:

1. Opening data is only the first step: city data in its current forms often seems unilluminating, intimidating and open to misinterpretation. The focus needs to shift from opening data to getting it used.
2. Data doesn't tell the whole story: users worry that city data can often miss the informal, the excluded, the new and the emerging; that it can sometimes obscure important qualitative insights. There are also concerns that too much of the most valuable data remains in private hands, and that the market for trading data is very immature.
3. Users within each category vary significantly around aptitude and attitude.

In addressing challenge 3, the research provides a categorisation of users, not based on sector, but on faith in data (idealist versus sceptic) and ability to analyse and work with data (independent versus dependent users):



The research identifies four opportunities:

1. Increase skills and training across sectors in how to use and interpret data. Identify clear city challenges and engage the community (across business and individuals) to develop clearer, more usable stories. Connect users in different fields and sectors to collaborate on problems.
2. Measure the data that cities can currently miss: the socially excluded, the informal economy; emerging trends. Enrich big data with more qualitative forms of insight.
3. Create a more effective marketplace that encourages private data holders to trade and share
4. Develop services and tools (e.g. within city data stores) that target people by attitude and aptitude.

A3. Types of data

To facilitate analysis and to provide clarity, it is important to unpick and classify the basic categories of 'big data'.

A first dimension is the extent to which data is open or closed. There is a spectrum: 'open' can refer to 'open within departments or government', or 'open to all' –and everything in between. Ownership and related governance arrangements will determine the extent to which data is open or not. *[DN: More about devolution…]*

Other dimensions relate to time, space and the way various digital identifiers define social and administrative data. A core classification follows.

Temporal

On this dimension, there is a spectrum from fast to slow in that data that is produced in real time shows change that occurs on anything from a second by second scale to change that might occur over a week or so. As the temporal interval at which data is available increases, then data which is recorded yearly clearly shows slower change while data that pertains to every decade such as Population Census data might be regarded as slow. Some activities that Government have responsibility for require fast data such as controlling traffic while much of it relates to longer term change such as that related to strategic planning of infrastructure, education, healthcare and so on. Data which is sampled on a second by second real time basis will eventually be available for much longer periods and then this will inform on slower trends. Sometimes, Government may be forced to choose between the velocity and the veracity of data – for instance with disease outbreak or purchasing habits, where new digital technologies are producing real-time user-generated data which however remains of dubious quality compared to traditional survey and monitoring-derived data.

- Real time data: from sensors, mobile phones, Twitter feeds etc.
    - Government 'admininstrative' data – central and local
    - Retail, marketing and other company-held data
    - 'Small data' – generated by individuals
    - Organisational data:
        - Performance data (league tables etc)
        - Financial accounts
- Slow (in terms of frequency of collection)
    - e.g. Population Census data (only gathered every 10 years), Labour market data
    - Planning data: Geometry data from Ordnance Survey, for example, GIS-related data – at all granularities – from macro to micro
- Managing the shift from slow to fast e.g. data from DECC on energy consumption by postcode, management of shift to smart meters and the impact this will have on cities.

Geospatially referenced data

- Many applications depend on spatial location and such locations can be at a very fine spatial scale, at the level of metres or below. The geometry for the scale is usually produced using digital photogrammetry at the finest scale, GPS at a coarser scale and by remote sensing at the largest earth level scale.
- Much data is geocoded to these locations once it has been collected as long as it has some common key or spatial reference

- Flow data such as transport movements are collected using a variety of systems from smart cards, to embedded sensors with GPS, to traditional forms of survey

Digitally Identified Individual and Administrative Data

- Social media data which pertains to individual responses
- Content sources (e.g. reports from the National Audit Office , Hansard and the wider web)
- Video and related pictorial data

A4. Analytics

Our definition of analytics embraces the processing of data, the building of enhanced data basis for analysis and modelling, the development of new techniques for extracting pattern in such data (though data mining) and the visualisation and deployment in decision support systems. We discuss each in turn.

- Processing data
    - Raw data nearly always has to be processed in a variety of ways to make it useful. These activities range from basic preparation through linking and integrating data to the generation of synthetic data. Examples follow.
- Preparation
    i. Data quality + examples (RFID card users tapping in and not tapping out to a transport system, social media spamming)
    ii. Data provenance – where it has come from, what is done to it etc before use
    iii. Applications of data might not be possible because of privacy issues (refer forwards to full section on privacy for more detail)
    iv. Spatial referencing for linking data, and temporal matching. Common linkage points such as common keys e.g. (IP) address matching. Spatial and temporal referencing and difficulties in linking
    v. The currency of the data, and the frequency with which the data is updated, also need to be understood.
2. Enhanced data bases for analytics and modelling
    i. IoT and transport modelling – using sensors as a basis for estimating origin-destination flows by mode
    ii. The geodemographics of populations, as well as life cycle change and aging.
    iii. Microsimulation of individual populations

iv. More effective forms of modelling ranging from descriptive data-driven models, analytical where understanding is key, through to predictive and thence to design or prescriptive models
v. The role of algorithms and data mining in exploring data

3. Visualization of data and new forms of visual analysis
4. Decision support systems – the persuasive power of visualisation in decision making scenarios for example in traffic movements and flows

A5. Deriving value from data

All four categories of data users are also, in various ways, generators and providers of data. Government has data on all of its activities, including data on individuals as users of services but also on the performance of delivery units. The same applies to business; citizens generate much of their own data – position data through mobile phones to Twitter data while academics conduct surveys but are also responsible for much of the processing. Much of the value of data is generated within organisations for their own use. However, there are new opportunities that arise for example from the linking of data[18]. New opportunities arise in this context from the cultural shift towards transparency and increasing 'openness' of data – a striking example being the Government's shift to openness following the Berners-Lee-Shadbolt proposals which led to the Open Data Institute.

i. Business and commercial - contributions to economic change and growth which result from available data and analytics
ii. Public policy and planning
iii. Research
iv. Citizens – citizen science and public participation: measurement and display
    i. Building information systems that can be interrogated to provide 'pictures' of a city through time
    ii. And that can articulate planning issues to a wider public
    iii. cf. 'Enquiry by design'
    iv. Crowdsourcing
v. Discuss future possibilities – e.g. better data integration etc
vi. Equity as well as economic value

A6. Case studies/applications:

We illustrate our analysis through a number of case studies – albeit a small sample from what is already a very large pool. Case studies demonstrate:

[18] For instance the ESRC's Business and Local Government Research Centres - http://www.esrc.ac.uk/research/major-investments/Big-Data/BDN-Phase2.aspx

- positive impact and effectiveness - that new value is emerging from the use of data and analytics
- value in the field of city management and planning – that is the optimization of existing city infrastructures and service systems (management) and the planning of future city infrastructures and service systems. This includes infrastructure and service systems such as transport, energy, housing, education, waste, water and healthcare.
- applications only made possible with recent advances in data collection/ processing/ analytics - rather than long-standing uses of data that have provided value (e.g. use of census data)