European Data Market
SMART 2013/0063

D3.2
Data-intensive Policy-Making Impacts on European Governments

June 26, 2014

Author(s)	David Osimo, Katarzyna Szkuta (Open Evidence), Giorgio Micheletti, Gabriella Cattaneo, (IDC)
Deliverable	D 3.2 Quarterly Stories – Story 2
Date of delivery	June 26, 2014
Version	1.0
Addressee officer	Katalin IMREI Policy Officer European Commission - DG CONNECT Unit G3 – Data Value Chain EUFO 1/178, L-2557 Luxembourg/Gasperich katalin.imrei@ec.europa.eu
Contract ref.	N. 30-CE-0599839/00-39

TABLE OF CONTENTS

The present story focuses on the data-intensive policy making impacts on European governments at national and local level. The table below outlines the story's general information and the key elements of the story's description.

The document is structured in three main parts:

- In the first paragraph an overview of the data-driven policy making in Europe will be provided;
- In the second paragraph a few case studies from European governments running projects on data-intensive policy making will be presented: the stakeholders involved in those case studies, the technology used and the impacts obtained by the innovative use of this technologies will be investigated;
- The third paragraph will summarize the key messages and conclusions to be taken from the case studies.

1.1 Short Summary

Big data already plays an important role in government, especially in providing better public services. From garbage collection to crime reduction, from public transport management to social services, there are many innovative initiatives in Europe already delivering concrete quantifiable benefits in terms of cost reduction and better, more targeted service provision. Many EU startups are active in this domain.

Open government data is an important enabler, but the high impact is only achieved when administrative data is analysed and contextualized with other categories of open (meteorological, traffic sensor, etc) data.

However, these initiatives and services still are at the micro level and have not scaled as in the US. Government adoption is patchy and left to individual initiative; providers have not developed commoditized scalable solutions.

A strategic approach is needed to government as a lead user of big data, and as an enabler of greater data sharing, in the context of a wider big data industrial strategy. Very few countries have done so at present.

1.2 General Information

GENERAL INFORMATION	
Title	Data-intensive policy making impacts on European governments - Framework conditions and application cases.
Link with information /Sources	Web Pages, Interviews, Policy Documents, background documents sent by interviewees http://www.prescribinganalytics.com/
Interviews	Pamela Cook, Managing Director, Infoshare. http://infoshare-is.com/ Dr Erik Jenelius, KTH, Royal Institute of Technology, Traffic and Logistics Department. http://www.kth.se/en Dr Alexandre Valette, SNIPS, co-founder. http://snips.net/ Laura Hart, Department for Business, Innovation & Skills, UK Government. https://www.gov.uk/government/publications/uk-data-capability-strategy Suzanne Littlehales, SAU Manager, Policy, Performance and Research – Strategic Analytical Unit (SAU), Nottinghamshire County Council. http://www.nottinghamshire.gov.uk/
STORY DESCRIPTION	
Topic/object of story	Decision-Making processes in government based on data analytics.
Main examples	Single view of multiple databases, data analytics in social work, traffic management, traffic predictions based on past data and context, fraud detection, overall data policy for business and research.
Main impacts identified in this story	Increased operational efficiency, more focused/accurate policy intervention. lower costs, increased benefits for beneficiaries, real-time or near real-time policy interventions based on predictions and real-time data analysis (in road traffic and crime).
Main stakeholders	Innovative SMEs (SNIPS, Infoshare), big companies (ATOS, IBM), government agencies (Nottingham City Council, London Borough, UK government), research institutes (KTH).
Key words	Data-intensive policy making, big data in government.

1.3 Trends in data-intensive policy making

"90% of all the data in the world has been generated over the last two years" - this phrase is cited all over when talking about importance of big data. But this data influx is more than visible in the government sphere.

According to the *Demystifying Big Data* report, the amount of information captured by the US government since the year 2000 has increased exponentially. For example, in 2009, the U.S. Government produced 848 petabytes of data and U.S. healthcare data alone reached 150 exabytes. At this rate, Big Data for U.S. healthcare will soon reach zetabyte (10^21 gigabytes) scale and soon yottabytes (10^24 gigabytes).

This growth of government data calls for better management and analytics in order to efficiently use the data themselves. Interestingly, in the same period of time the incentives for data analytics' use have also increased. Austerity measures, and hence cuts in government spending, push government agencies to identify opportunities for more efficient operations and interventions. Data analytics shows how the results can be delivered at lower costs but also makes the identifications of issues easier and faster and finally allow for more accurate and focused interventions.

McKinsey estimates the potential savings for European public administrations at around €150 billion to €300 billion a year.

But there is a caveat to be made - the 3Vs of data (volume, velocity and variety) should be expanded to 4Vs as the "data veracity", is of outmost importance in the public sector, even more so in policy making where people well-being and sometime lives are at stake). Therefore in our story we will cover also examples of data cleansing and validation since this is the first but very important step where the data-intensive policy making commences.

The impact of data-intensive solutions on government can be structured alongside three levels.

- The first is the level of service provision. Governments remain in charge of many services, either directly or indirectly. Garbage collection, security, public transport, social services, tax collection are all areas where big data solutions are being used.
- The second level concerns public policies, and specifically the possibility to design more effective public policies by anticipating their impact ex ante or better evaluating it ex post. While evidence-based policy making is a long-standing trend in government, the usage of big data to anticipate problems and predict impact is new - and still developing.
- The third level is strategy. Big data is not considered as a tool to support decisions, but as a factor of economic competitiveness, which requires strategic intervention.

The level of public service provision is the most mature, where many applications are visible. Beside security, an obvious example is the application of big data; the most interesting applications areas are related to:

- transport, in particularly traffic jam prevention. The combination of data from sensors, crowdsourcing and contextual open government data help to reduce congestion in roads as well as in public transport.
- crime prevention. There are several examples, especially in the US, of big data services being used by local police to anticipate crime and identify target areas.
- tax collection, another traditional area of big data application. These are used typically to prevent fraud and identify clear possible target to carry out human checks.
- garbage collection, in order to optimise the services and prevent accumulation of garbage.

- disaster management, in terms of providing fast and additional information on the status. For instance, in the case of the Haiti Earthquake, models based on SMS reports by people were able to predict the structural damage of most affected areas.
- social services. This is a more recent and original area of application. Data-intensive tools help identify people at risk earlier, improve chances of success.

Obviously, in none of the examples above big data solutions are able to deliver the impact "by themselves". Impact is always generated by the combination of such tools with human effort and organisational change.

At the level of policies the main problem to be addressed is the increasing complexity and interconnectedness of our world; the goal of applying big data to policy problems is therefore to anticipate problems by better prediction, ensuring fast detection, and gathering real-time feedback on policy implementation. However, most examples remain at the level of research or pilots, with little large-scale deployment.

The main application areas appear to be the following:

- development projects: The World Bank has piloted several big data projects in relation to development policies, in order to use alternative datasets to address wicked problems and data gaps. For instance, several project aim to measure poverty and social indicators based on unpredictable data sources such as night illumination, mobile surveys, tweets, scraping websites.
- anti-fraud and money laundering
- crisis management and anticipation by building up crisis observatories, i.e. laboratories devoted to the collecting and processing of enormous volumes of data on both natural systems and human techno-socio-economic systems, so as to gain early warnings of impending events.

At strategy level, governments can help generating overall socio-economic benefits by facilitating the creation of a data-friendly ecosystem. This requires the typical innovation policy measures in support of a specific sector: from training and competence building on data skills, to infrastructure provision in terms of broadband and data centres, to a supportive regulatory environment.

The UK government has been particularly active in this domain, by launching initiatives that aim to open up government data and create the conditions for optimal data sharing in the economy, by introducing guarantees for consumers' control over personal data owned by businesses, with the final objective of creating "a new marketplace for services and products created to help consumers access and understand their data".

Domain	Application sector	Example
Public services	Transport	Stockholm city congestion charges and real-time traffic management
	Garbage	SNIPS project in the Aquitaine region
	Social services	Nottingham City Council
	Tax collection	London Hackney
	Crime prevention	PredPol solutions in US
Public policies	Development	UN GlobalPulse, World Bank Datadive
	Disaster management	SMS reporting for Haiti Earthquake
	Money laundering	SynerScope (NL)
	Innovation policy	Knight Foundation scientific roadmap based on big data
Government strategy	Open government data	UK Government - Open Data Strategy – Open data platform, Open Data Institute
	Training and infrastructure	UK Government – Data Capability Strategy,
	Data sharing standards and regulation	UK Government – Data Capability Strategy

Public health initiatives are not mentioned here as they are considered part of the specific Health domain, which will be treated separately.

1.4 Overview of the Case Studies

The case studies below were chosen to demonstrate the impact of data-intensive policy making in different streams of government policy on the level of public services provision, which is most mature trend in this field. We have gathered a wide set of examples from social care and health through transport to crime prevention and tax evasion. We have also singled out examples of one government strategy which focuses on providing framework conditions raising the profile of data and big data use in business and research.

1.4.1 Single and accurate view of citizens' data – Infoshare

Case name	Single and accurate view of citizens' data
Short description	Access to accurate, fit for purpose, up-to-date data, pulling data from multiple sources and identifying mismatches, double entries, linking, cleaning & validating data. Local governments for specific purposes need data from many sources (disparate, often the same citizens with different information on them hence double entries, etc.). • London Borough of Hackney introduced the single repository storing information on Revenues and Benefits, CRM, Housing, Parking and the National Land and Property Gazetteer. • Nottingham City Council - automated matching across multiple agency data sources from the police, youth offending service, health, probation service, education and anti-social behaviour registers for the Troubled Families programme. The Troubled Families Programme was set up in 2011 to help 120,000 troubled families in England turn their lives around by 2015. In return for its support, and reducing the burden of certain families on the public purse, the Council will receive £2.3m during the lifetime of the programme. • Vulnerable information sharing hub. Five public services: local councils, police and health boards in South of Wales are sharing information for better data handling and enhanced analytics (analytics is provided by Atos). • Customer First Programme in Sheffield City Council – single view platform for improved communication with citizens. • East Sussex County Council – multiple agencies data in single view (Education, Social Services, Health, SureStart, Youth Offending) – accurate and up-to-date data on vulnerable children.
Country	UK
Case sub-domain	Social services, crime, tax evasion, fraud.
Leader	Infoshare.
Type of data	Socioeconomic data, financial data, vehicle registration, physical property, etc.
Customers	Local government.
Tech	Infoshare's ClearCore.
Impact type	Saving on operational costs (time and human effort, maintenance fees and the solution self sufficiency), savings on benefits (double allowances, etc.), savings on effectiveness of social intervention thanks to data accuracy, more trust for government (better experience for citizens).
Impact quantification	London Borough of Hackney: • 40% decrease in call handling times due to single data repository

	and single view; • 10 thousands duplicate entries of residents identified (savings in identifying double allowances); • moving online with verification of parking applications and permits (online transactions is a 95% reduction from £7.32 to £0.32). East Sussex County Council: • 6 thousands hours spent on identification, searching and getting the information (decreased match rate) • coherent and accurate view – better handling of cases of vulnerable children. Nottingham County Council: • information from many sources is automatically matched (instead of manually checking thousands of pieces of data by hand by the personnel), time for handling information decreased from 3 months to 3 hours; • long-term indirect impact: successful family intervention reduces the involvement of challenging families in anti-social behaviour by 59%, halves incidents of truancy, exclusion and bad behaviour at school and cuts their involvement in crime by 45%.
Drivers / barriers	Barriers: • government agencies work mostly with large suppliers; • demands cooperation with different agencies/departments and hence breaking the organisational silos. Drivers • spending cuts in government facilitate introduction of innovative solutions.

1.4.2 Predicting passengers flows in the Paris Region Public Transport Network - SNIPS

Case name	**Predicting passengers flows in the Paris Region Public Transport Network - SNIPS**
Short description	SNIPS developed an algorithm that can successfully predict how many people will be boarding and unboarding trains at each station, throughout the day, and up to a week in advance. The algorithm was subsequently integrated in a mobile app (called “Tranquilien”) allowing commuters to see in advance in which train they can find a seat. The ‘seed’ data where delivered by SNCF (from a passengers’ manual count) and matched with contextual data (offices close to the stations, parkings, other points of interest, etc.). Contextual data in this case improve the predictions based on past data and help to predict changes. The data are now enhanced by users (crowdsourcing), i.e. thanks to the users’ feedback on predictions and users data (smartphone as sensors). Most interestingly, the app nudges users’ behaviour -the information on lines congestions have impact an on their decisions to board the trains and therefore spreading the peak hours.

	The company applied also its algorithm to predict risk of car accidents in London and San Francisco (SafeSignal). It works also with La Poste (French mail service) to predict the clients flux in post offices.
Country	FR
Case sub-domain	Public transport.
Leader	SNIPS and SNCF.
Type of data	Transport data and contextual (open government) data.
Customers	Public agencies - national railway company, French mail service.
Tech	Machine learning.
Impact type	• resources efficiency: better planning – train schedules adapted to traffic; • passengers as sensors - less costly than installing sensors on trains (costs of sensors + maintenance); • socio-economic: spreading the peak hours – passengers' comfort, more balanced use of public transport.
Impact quantification	The impact is highly dependent on number of users. Tranquilien was launched in June 2013. It had 35 thousands users in the first quarter. Currently, the applications has been downloaded 100 000 times.
Drivers / barriers	Barriers: • low awareness of data value in government; • fear of big private companies disposing of important data and acting as intermediaries between citizens and public agencies (e.g. Google transit).

1.4.3 Real-time traffic management in cities - KTH big data deployment

Case name	**Real-time traffic management in cities - KTH big data deployment**
Short description	Stockholm has a lot of experience in using data in traffic management. Its traffic charging system developed with IBM helped to decrease the traffic and nudge citizens to choose public transport. The system varies tolls by time and day to influence traffic patterns and congestion levels. The KTH division on traffic and logistics research team works with City of Stockholm and Swedish Transport Authority on an efficient real-time traffic management. Data from radar sensors on motorways are combined with probe vehicle GPS data (from a largest Stockholm taxi company). This would allow the city and traffic management authority to understand how traffic congestions appear as well as react in real time to situations at hand. The algorithm allows also for short-term prediction s(in next 30 minutes), which has impact on traffic management (slowing down some motorway

	lanes for example). It will allow decreasing the number of traffic congestions instead of only knowing about them and informing the drivers. A similar pilot project in Australia in 2008 (over 15km by Vicroads with the use of coordinated ramp signals) resulted in 4.9% increase in average flow (pcu/h/lane), 34.9% increase in travel speed (from 48.9 to 66 km/hr) and 65.3% reduction in delay (min/km).
Country	SE
Case sub-domain	Traffic management.
Leader	Royal Institute of Technology of Sweden (KTH).
Type of data	GPS from large numbers of vehicles, radar sensors on motorways, toll charges, weather.
Customers	Local government.
Tech	Stream analytics.
Impact type	Reduction in traffic and emissions.
Impact quantification	The first pilot with variable toll charges reduced traffic by nearly 25 percent, while train and transit passengers increased by 40,000 per day. After three years, the reduction in traffic wait time was 50%, 60,000 more passengers were taking public transport and vehicle emissions had dropped by 14 to 18 percent in the inner city. Many more Stockholm's residents were combining auto use with extra walking and bicycling. The KTH algorithm will allow for even more fluent traffic, reduced travel time and reduced emission by real-time reaction to traffic congestions.
Drivers / barriers	Drivers: • Very good collaboration with public authorities. Local government understands the need for good analytics and prefers getting raw data as opposed to only procuring highly processed insights. • The project inscribes into larger local and national policy.

1.4.4 Other interesting examples of big data use in public services

The PredPol tool was developed over the course of six years by a team of PhD mathematicians and social scientists at UCLA, Santa Clara University, and UC Irvine in close collaboration with crime analysts and line level officers at the Los Angeles and Santa Cruz Police Departments. The tool processes crime data in order to:

- assign probabilities of future crime events to regions of space and time;
- present estimated crime risk in a useable framework to law enforcement decision makers;
- lead to more efficient & more accurate resource deployment by local law enforcement agencies.

The use of this data analytics tool resulted in better accuracy of prediction than that of experienced crime analysts (50% more matches in a 6 month randomized controlled trial). It also resulted in

reduction of crime itself. In Los Angeles' Foothill Division, crimes were down 13% in the 4 months following the rollout compared to an increase of 0.4% in the rest of the city where the rollout had not happened.

Another good example of big data analytics is taken from the health sector, in the UK where Mastodon C, a start-up big data company collaborated with Open Healthcare UK and Dr Ben Goldacre. The team looked at the regional prescribing patterns of statins (drugs used to lower cholesterol levels) by NHS doctors in England. The analysis of aggregated prescription data and costs of all the drugs from the market (which have equal effects and safety levels) showed an important variation in prescription rates of the most expensive drugs. The cost of an individual prescription item varied from 81p for a generic drug, to more than £20 for licensed drugs. Based on these results, the team estimates that up to £200m could have been saved if the research had been conducted a year ago. This data would have helped NHS to better plan for funding as well as the clinicians on effective drug prescription. The study results are available on http://www.prescribinganalytics.com/ platform and the team plans to analyse data from other drug categories.

1.4.5 UK's Data Capability Strategy

Case name	UK government data capability strategy
Short description	The UK data capability strategy published in October 2013 is part of the wider Information Economy Strategy published in June 2013 as well as of Using Industrial Strategy to help the UK economy and business compete and grow. The strategy has three pillars: • Skills; • Infrastructure, software and collaborative R&D; • Sharing and linking data securely and appropriately. The main objective of the strategy is raising the profile of data through building a strong skills base, supporting the growth of infrastructure and R&D research as well as creating conditions for sharing data across sectors and disciplines. Most interesting for this case initiatives are: • Creation of the **Administrative Data Research Network** to promote research on linked data from government departments, allowing the analysis of massive and mixed datasets. • The **midata initiative** enabling consumers to access data that companies such as energy providers or banks hold about them, which they can use for their own benefit. • The **Connected Digital Economy Catapult (CDEC)** is supported with over £50 million of funding from the Technology Strategy Board, bringing together innovators from industry – both large companies and SMEs – research and academia, promoting alignment and collaboration amongst the many players in the digital innovation community

	• **Trusted Data Accelerator** - a platform where companies, universities, public bodies and other organisations can share, combine, experiment and engage innovators with data without fearing compromising IP, data privacy and security
Country	UK
Case sub-domain	Policy
Leader	UK government, Cabinet Office and Department for Business, Innovation and Skills.
Type of data	Open government data, open data in research, personal data.
Customers	Business, academia.
Impact type	Creating framework conditions for big data use – building skills base, supporting infrastructure and R&D, reducing licensing friction, setting up standards, building trusted environments for collaboration of innovative SMEs and large companies that hold data
Impact quantification	Research by Nesta shows that UK data-driven firms are 40% more likely to report launching products and services ahead of their competitors than those who aren't data-driven firms. The Trusted Data Accelerator aims to link innovative SMEs that have ideas on how to use data but lack access to it and large data holders who are not knowledgeable about the potential new usage of what they hold. UK's digital market is valued at more than £100bn, driven by growth of digital services on the internet, which represented 23% of total UK GDP growth between 2004 and 2009.
Drivers / barriers	Information not available due to recent start of the project.

1.5 The technological innovation and the role of stakeholders

Government, in its institutional roles, has traditionally been a holder of large datasets such as anagraphic and fiscal data. Analytics solutions have always been used in government departments, and in particular in information intensive sector such as security, transport and tax. However, the cases presented here share a common level of innovation because of the usage on large scale, real time and diverse datasets. In particular, the diversity of datasets implies the joint analysis of different government datasets (e.g. historical analysis of school records and crime reports) merged with contextual data (weather, events, other contextual conditions) in order to identify unexpected correlations that could help predictions. The goal is to better predict and target issues of interest to government.

Suppliers are often start-ups companies that encounter difficulties in providing their services to government, but they also include large vendors (IBM), consultancies (Atos) and research centres (KTH). Providers are typically selling their services in terms of analytical tools and actual analytical work, often in partnership with other consultancies.

To involve start-ups, most projects start off as experimental partnership or research project between users in government and providers, rather than as traditional procurement. The market appears still immature and fastly evolving, as well as the demonstrated impact. However it seems that in the US there are more companies providing commoditized services on the market (vertical applications of "big data as a service") in the government domain, such as Palantir, Fiscalnote and PredPol.

Open government data act as an enabler, by allowing for contextual analysis to identify correlations. However the core datasets used are typically not public, mainly because they contain sensitive information. They include administrative data such as tax, health, school, crime records; sensor-produced data; and crowdsourced data from people.

1.6 The data-driven innovation benefits and impacts

As already highlighted, the data-intensive policy making examples are still early pilots and therefore the impact data are mostly the outcomes of the projects themselves. Still, the early numbers are already promising and visible in different sectors.

1.6.1 Increased operational efficiency due to data usage, use of clean and validated data

Local governments and national governments when planning specific interventions often use data from many sources and many departments and agencies. This data needs to be accurate and validated but as coming from many sources often demands specific technologies to obtain a single view of citizens' data. Unified validated databases that can be used for many purposes results in saving on operational costs, mostly in time and human effort but also in fees for consultancy and duplicated costs for IT solutions. Data accuracy helps to save on government spending due to decreased social benefits fraud.

For example, in the London Borough of Hackney the single data repository and single view of citizens' data resulted in 40% decrease in call handling. Identification of 10 thousands duplicate entries incurred important savings thanks to double allowances previously paid. Nottingham County Council where the Troubled Families national programme is being run the time for identifying families fit for the programme decreased from 3 months to 3 hours.

1.6.2 More focused and accurate policy intervention thanks to better data analytics

Single and validated data repositories are not the goal in itself. The analytics potential is huge and some early cases show what kind of impact can be achieved in social care. The analytics applied to multivariable databases allows for more focused social interventions, identification of cases that slip through the social care net and hence create a difficult to quantify but very important socio-economic impact. According to Infoshare, successful family intervention reduces the involvement of challenging families in anti-social behaviour by 59%, halves incidents of truancy, exclusion and bad behaviour at school and cuts their involvement in crime by 45%.

1.6.3 Lower costs of policy intervention thanks to analysis of past data

The analysis of past data allow for very good quality predictions that in result enable for real-time or nearly real-time decisions. Transport data are good examples thanks to possibilities of installing sensors, GPS devices as well as ubiquitous smartphones that are more and more used as sensors. The new traffic management system designed in Stockholm with the collaboration of KTH ill allow for more fluent traffic, reduced travel time and reduced emission by real-time reaction to traffic congestions. The impact on traffic is not yet known but a similar pilot project in Australia in 2008 (with

the use of coordinated ramp signals) resulted in 4.9% increase in average flow (pcu/h/lane), 34.9% increase in travel speed (from 48.9 to 66 km/hr) and 65.3% reduction in delay (min/km). The Tranquilien application for bigger Paris region adds to the past data predictions also the contextual data that increases the quality of predictions when changes occur. The application has been downloaded 100.000 times and counts already 56 thousands check-ins.
The Mastodon C and Open Health Predictive Analytics projects show how analysis of past data can bring £200m saving on one single category of drugs by reducing prescription of more expensive medicines. Similar analysis of electronic health records in cardiovascular disease in the United States achieved an estimated $1 billion in savings from reduced office visits and lab tests.

Finally, the usage of data analytics in crime analysis results in high accuracy of prediction, more accurate than that of experienced crime analysts (50% more matches in a 6 month randomized controlled trial). This allows for more effective policy interventions, which drives down the crimes numbers themselves. In Los Angeles' Fooxthill Division, crimes were down 13% in the 4 months following the rollout of the programme compared to an increase of 0.4% in the rest of the city.

1.6.4 Increased benefits for beneficiaries of government services

Thanks to accurate and consolidated administrative data, citizens have better experience in contacting the government thanks to less administrative burdens linked to filling out the same information ever again, less time in calls, and quick identification. In the result they develop more trust towards the government.
Needless to say that increased efficacy in policy interventions have direct impact on their beneficiaries receiving better and more timely service.

1.6.5 Nudging citizens behaviour based on real-time information

Data analytics allows also for specific policy interventions based on behavioural science. The access to information itself has impact on citizens' behaviour thus driving positive changes without other policy interventions. For example, the variable toll charges incentivise citizens to switch to public transport, cycle or walk shorter distances therefore decreasing emission in cities. The city of Stockholm with its pilot programme introducing variable toll charges reduced car traffic nearly by 25 percent whereas train and transit passengers increased by 40,000 per day. After three years, the reduction in traffic wait time was 50%, 60,000 more passengers were taking public transport and vehicle emissions had dropped by 14 to 18 percent in the inner city. In the same traffic management case, the application Tranqulien thanks to information on trains' capacity, spreads the peak hours and allows for a more balanced use of public transport. In the health area information for clinicians on over or under prescriptions in their area may also have impact on their future prescription costs.

1.6.6 Creating a favourable environment for innovation adoption and diffusion

Government is not only a data holder of importance and nor its only focus in big data is on data analytics for its own use. The important role of government is also to created favourable environment for big data use by citizens and companies.
UK's very comprehensive Data Capability strategy shows that there are three levels where the government can intervene – by building a skills base to answer to the increasing skills gap, by building data infrastructure and funding further research and finally by enabling and stimulating further collaboration between different stakeholders in the data value chain. Interesting example is the Connected Digital Economy Catapult (CDEC) that brings together industry (both SMEs and large data holders), researchers and academia and creates a secure environment for innovation. The Midata Programme is also worth mentioning: here consumers are able to access data that companies such as energy providers or banks hold about them, which they can use for their own benefit (for example

energy providers by sharing the data on consumption allow their clients to have better insights on their energy use and type of contract they should look for).

Overall framework	Government impact	Data
Increased operational efficiency due to data usage, use of clean and validated data	Reduced costs	40% decrease in call handling decrease in time spent on identifying families in need from 3 months to 3 hours.
More focused and accurate policy intervention thanks to better data analytics	Better (more evidence-based) identification of beneficiaries of social interventions.	Successful family intervention reduces the involvement of challenging families in anti-social behaviour by 59%, halves incidents of truancy, exclusion and bad behaviour at school and cuts their involvement in crime by 45%.
Lower costs of policy intervention thanks to analysis of past data	Increased effectiveness of policy measures. Reduction in crimes. Reduced traffic.	£200m saving on one single category of drugs by reducing prescription of more expensive medicines Crimes were down 13% in the 4 months after the crime prediction programme roll-out.
Increased benefits for beneficiaries of government services	Reduced time spent on filling out forms and identification.	
Nudging citizens behaviour based on real-time information	Reduced traffic. More balanced use of public transport.	Reduced car traffic nearly by 25 percent whereas train and transit passengers increased by 40,000 per day.
Introduction and adoption/diffusion of innovation	More collaboration between innovative SMEs and large data holders (also in government). More innovative products	Research by Nesta shows that UK data-driven firms are 40% more likely to report launching products and services ahead of their competitors than

	and services.	those who aren't data-driven firms.
Quantitative impacts such as: turnover/revenues generated	Overall savings in government spending.	Estimated savings in European government from €150 billion to €300 billion a year.

1.7 Conclusions

The adoption of data-intensive solutions in government appears to be very important in addressing some of the key challenges that governments face, mainly in dealing with complex and multi-faceted problems.

Impacts are already visible at the micro-level and on information-intensive public services, while they are still emerging at the policy level. Crime prevention and traffic management appear as the most mature sectors.

There are many innovative EU-based businesses in this domain, but not at the level of market development of their EU counterparts. Projects are often ad-hoc collaborations between government and companies, while in the US there are several vertical applications with commoditized "big data as a service" solution.

Government as a lead user could stimulate the growth of the sector provided the well-known barriers to collaboration between government and start-ups are addressed: public funding and public procurement are largely inaccessible to start-ups in this domain. Government can play a strategic role also as a catalyser of better data sharing in the economy, by developing the appropriate incentives and checks and balances.

Main Sources

- Australian Government Information Management Office – The Australian Public Service Big Data Strategy: Improved Understanding through Enhanced Data-analytics Capability Strategy Report
- David Bollier (2010) – The Promise and Peril of Big Data, the Aspen Institute,
- danah boyd and Kate Crawford – Six Provocations for Big Data, *A Decade in Internet Time: Symposium on the Dynamics of the Internet and Society, September 2011*
- The Economist Intelligence Unit – Big Data and the Democratisation of Decisions
- James Manyika, Michael Chui, Brad Brown, Jacques Bughin, Richard Dobbs, Charles Roxburgh, and Angela Hung Byers (2011) – Big Data: The Next Frontier for Innovation, Competition, and Productivity, McKinsey Global Institute
- The Partnership for Public Service and the IBM Center for The Business of Government – From Data to Decisions II: Building an Analytics Culture
- TechAmerica Foundation's Federal Big Data Commission – Demystifying Big Data: A Practical Guide to Transforming the Business of Government
- World Economic Forum – Big Data, Big Impact: New Possibilities for International Development