

HM GOVERNMENT HORIZON SCANNING PROGRAMME

EMERGING TECHNOLOGIES: BIG DATA

A HORIZON SCANNING RESEARCH PAPER
BY THE
EMERGING TECHNOLOGIES BIG DATA COMMUNITY OF
INTEREST

PUBLISHED DECEMBER 2014

NOTE TO THE READER: THIS ANALYSIS WAS COMPLETED IN 2013 AND HAS NOT BEEN UPDATED WITH MORE RECENT DATA BUT SENSITIVE INFORMATION, SUCH AS POLICY IMPLICATIONS HAVE BEEN REMOVED

INTRODUCTION

1. The Government has already made a commitment to Big Data by classifying it as one of the 'Eight Great Technologies' which will propel the UK to future growth and help it stay ahead in the global race. The 'Information Economy Strategy' reports on the increase in data being generated and the importance of new types of computing power in order to reap the economic value of the data.
2. This paper sets out to cover the following areas:
 - a) Defining Big Data
 - b) High-level trends in Big Data
 - c) Opportunities for Big Data applications
3. In developing this paper, a 'community of interest' has been established comprising policy leads and analysts from across government with an interest in Big Data. This paper draws on their insights, insights from the private sector, academics, and the extensive open source literature on the Big Data topic.

DEFINING BIG DATA

4. Following consultation across HMG, the following is a working definition of Big Data:

"Big Data refers to both large volumes of data with high level of complexity and the analytical methods applied to them which require more advanced techniques and technologies in order to derive meaningful information and insights in real time".

5. Within this definition, there is a fundamental assumption about the power and importance of new techniques and technologies, which are often called 'analytics'. The real value of analytics is that it can draw out new meanings, insights and value from bringing together individual datasets, which on their own might have limited value.
6. The scope of analytics covers three functions:
 - a) *Descriptive analytics* helps us to understand what is happening in the world, using visualisation techniques, some modelling and regression.
 - b) *Predictive analytics* helps us to predict what will happen, using forecasting.
 - c) *Prescriptive analytics* helps us to work out what we want to happen, using optimisation, simulation, scenario testing and Multi-Criteria Decision Analysis.

HIGH LEVEL TRENDS IN BIG DATA AND ANALYSIS OF THEIR IMPLICATIONS

7. With the Big Data universe expanding so quickly, looking to the future is very difficult. But there are 6 key trends that are already happening to a greater or lesser degree that are shaping its development. These trends provide an overview of the fundamental characteristics of Big Data. Trends 1-5 have largely been driven by the private sector, particularly the Internet companies (including Google, Amazon and Yahoo), which are highly active in social analytics.
8. **TREND 1: The emergence of cloud computing.** This allows users to access highly scalable computing and storage resources through the Internet. It allows companies to tap server capacity as needed and expand it rapidly to the

enormous scale required to process big datasets and run complicated mathematical models. Cloud computing lowers the price of data storage because the resources are shared across many users, who pay only for the capacity they actually utilize. Companies can access this capacity much more quickly, without the time and expense needed to set up their own systems, and they do not have to purchase enough capacity to accommodate peak usage.

9. **TREND 2: The development of new software tools and database systems for large, unstructured datasets.** Some of the data now being collected is structured in familiar ways, such as customer purchase histories, inventory levels, and machine energy usage. However, much of it is unstructured, meaning that it does not conform to a pre-defined schema and cannot be easily searched, processed, or queried in traditional database systems. This includes blog posts by users around the Web, most video and audio content, research articles published in academic journals, and handwritten medical records. New software tools make it possible to process these unstructured data in vast quantities. An important development in the commercial market is that it is now possible to mine structured data sources alongside unstructured data sources to merge datasets that were previously incompatible.
10. Independent analysis by EMC Corporation¹ indicates that the vast majority of new data being generated is unstructured. By 2020, the trend towards greater social analytics will encompass embedded medical devices (to capture vital and non-vital biometrics), entertainment and social media, and consumer images that can be analysed in real time.
11. **TREND 3: Development and refining of analytical tools so that they can process vast quantities of data in near-real time.** For example, there have been rapid advances in the ability to synthesise unstructured data from multiple sources (e.g. social media feeds, clicks and usage patterns on web pages, GPS tracking data, crowd-sourced reports on traffic and road incidents) and use this information to make very quick decisions.
12. **TREND 4: Monetisation of Big Data sources - realising the value of the data collected by selling it on the market.** The data collected in the course of companies' everyday activities often has an intrinsic value. For example, any comparison website has customer journey information that can be mined to understand more about how consumers make decisions. Most companies will have transactional information from their own customers, but this is only part of the picture, so adding additional data collected by other companies helps to build 'the story'.
13. **TREND 5: The increasing concerns around privacy of data and intellectual property.** Situations are occurring where big data analytics collide with difficult issues around privacy and ethics. Big Data technologies, and predictive analytics in particular, can empower organisations with astonishing capabilities. Knowing a lot about an individual user, consumer or citizen makes it possible to forecast their specific needs and behaviour, sometimes with a very high degree of precision.
14. **TREND 6: The rise of global smart cities.** Smart Cities² use information and communication technologies to be more intelligent and efficient in the use of

¹ <http://www.emc.com/collateral/analyst-reports/idc-the-digital-universe-in-2020.pdf>

² <http://www.fastcoexist.com/1679127/the-top-10-smart-cities-on-the-planet>

resources, resulting in cost and energy savings, improved service delivery and quality of life, and reduced environmental footprint all supporting innovation and the low-carbon economy.

OPPORTUNITIES FOR BIG DATA APPLICATIONS³

15. The previous section provided an analysis of the key trends relating to Big Data while indicating a number of challenges, principally around information security and privacy. However, a balanced evaluation of these trends will also indicate a number of opportunities. There are five categories of opportunities. Under each category, a range of examples from the UK and internationally have been identified.

16. **Creating transparency:** As the trends section indicated, transparency could result in profiles of individuals or social groups being developed which could be deemed as being negative. However, both external stakeholders such as citizens and businesses and internal stakeholders such as government employees and agencies can improve their efficiency when data from large public sector databases are made more accessible. For example, government agencies regularly collect a large amount of data on individuals and businesses through various regulatory and public service activities. Yet citizens and businesses frequently have to fill out forms for which some of the data have already been collected and stored. If agencies were to pre-fill forms for citizens and businesses from data already stored in government databases, this would save time for the submitters of forms as well as government agencies that would not have to re-input data.

The process for obtaining a **UK driving licence** is a good example of data sharing in action. When you apply for a driving licence online, the Driver and Vehicle Licensing Agency (DVLA) requires a photograph and signature for your new licence. If you have a UK passport then the DVLA will try to capture these electronically from the information already held by the Identity and Passport Service (IPS). This is often held up as an example of how simple changes can deliver practical improvements for end-users.⁴

17. **Enabling experimentation to discover needs, expose variability, and improve performance:** One highly valuable contribution that Big Data can make is to uncover tremendous variability in performance within different parts of a government agency that are performing broadly similar functions. Such information can offer a very valuable opportunity for improving the performance of operating units within an agency. There is an opportunity for Government to exploit open data sources to improve monitoring of policy delivery at a local scale. This is particularly importance for delivery-focused departments.

Important sources of Big Data exist outside traditional organisational boundaries⁵ One increasingly important source is the information shared publicly via social media. In the business sphere, organisations like Procter & Gamble are able to scan for

³ http://www.mckinsey.com/insights/business_technology/big_data_the_next_frontier_for_innovation

⁴ <http://www.policyexchange.org.uk/images/publications/the%20big%20data%20opportunity.pdf>

⁵ <http://www.policyexchange.org.uk/images/publications/the%20big%20data%20opportunity.pdf>

relevant feedback and comment, which can then be sent straight to the screen of the individuals that need to see it. This agile, real-time learning is a step change from the traditional approach of reading and replying to pen-and-paper correspondence and management memos as the primary source of data on how an organisation is working.

18. Segmenting populations to customise actions: Using segmentation to tailor services to individuals has long been an accepted practice in the private sector. However, the ethos of the public sector tends to be that governments should provide exactly the same services to all citizens. Research shows that segmenting and tailoring government services to individuals and population cohorts can increase effectiveness, efficiency, and citizen satisfaction.

Bundesagentur für Arbeit (German Federal Labour Agency) analysed its huge amount of historical data on its customers, including histories of unemployed workers, the interventions that it took, and outcomes including data on how long it took people to find a job. The idea was to develop a segmentation based on this analysis so that the agency could tailor its interventions for unemployed workers. This process, along with other initiatives applied over three years, allowed the agency to reduce its spending by €10 billion annually at the same time as cutting the amount of time that unemployed workers took to find employment, and increasing the satisfaction among users of its services.

Amazon applies collaborative filtering, a system that generates highly personalised recommendations based on its knowledge about purchase histories, product ratings and reviews. Similar principles are at work when Facebook and other social networks serve up personalised news feeds based in part on the user's position and relationships in the social graph. The result is content and services that feel more relevant to the user without them having to expend the significant effort that would be required to filter and prioritise a mass of unsorted material.

Mobile devices⁶ will increasingly connect individuals to a wide range of government services, ranging from medical records to NHS services, education resources to benefits, welfare and tax systems. Mobiles will also connect to the Internet of Things – devices in homes, cars, the workplace, and shops, and connect to similar devices carried by others forming a mesh of social interactions and transactions based on Big Data. A foretaste of what this world will look like is the Google Glass project. It is claimed that this prototype is capable of voice-activated translation, route finding, video recording and information retrieval. This new type of product will be as disruptive as smart phones and tablets.

The roll out of **Smart Meters** by DECC will allow energy companies to have a real time view of energy consumption across the country and could be used to manage supply and demand. For consumers, a key benefit could be the ability to assess their personal energy consumption and tailor the most effective way to run their appliances.

19. Another major opportunity relating to segmenting population is around the creation of Smart Cities.

⁶ Foresight Futures Report, Computing and Data: Exploiting Abundance by Professor Nigel Shadbolt

The Technology Strategy Board⁷ is investing £24 million to fund a large scale **Future Cities demonstrator** in Glasgow, £3 million each to Bristol, London and Peterborough, a further £50 million over five years to create a Future Cities Catapult in London and £5 million in an SBRI competition to support innovative companies to create new solutions to challenges identified by UK cities.

As a vehicle for GDP growth, Singapore has sought in the past to establish itself as the **Banking hub, medical hub and a “hub” of Southeast Asia** in several other areas. Along that line of thought, the Infocomm Development Authority has a working group tasked with establishing Singapore as an International Data & Analytics Hub.⁸

IBM has been working with a number of cities (including Rio de Janeiro, Dublin, New York, the Chinese city of Zhenjiang, Chicago and Stockholm) in areas like **emergency services response, the movement of traffic and public transportation systems, power and water supplies** in order to make them work more efficiently.⁹

20. Replacing/supporting human decision making with automated algorithms:

Some of the more sophisticated applications of Big Data use automated algorithms to analyse large datasets in order to help make better decisions. These types of techniques can often be very effectively applied in compliance activities, for example, among public sector agencies that need to find anomalies in payments such as in tax collections or benefit payments from labour or social security departments. Sophisticated tax agencies apply automated algorithms that perform systematic, multilevel checks on tax returns and automatically flag returns that require further examination or auditing.

The **Office of Policy and Strategic Planning in the Office of the Mayor of New York City** is applying predictive data analytics¹⁰ to discover and act upon patterns in databases that may indicate financial fraud. This is being applied to areas like tax evasion on cigarette sales, waste disposal and large-scale illegal occupancy in dwellings. The office employs approximately 5 data analysts. Given the success of the New York project, a joint working group is being set up on data analytics between New York City, Boston, Chicago and Philadelphia.

HMRC has a world-class data analytics capability including the **Connect and ADEPT systems**. Connect uses a range of complex and varied datasets, over one billion records, and applies analysis and profiling techniques to tackle a range of customer behaviours, including fraud and error. One example of how HMRC has used this system was as part of a campaign aimed at medical professionals, where Connect helped make the links between tax records and data from hospitals, pharmaceutical companies and insurers.

21. Innovating new business models, products and services with Big Data: Big Data from government can unlock innovation both inside and outside the public

⁷ <https://www.innovateuk.org/web/modernbuiltkn/article-view/-/blogs/%C2%A324-million-investment-will-make-glasgow-a-city-of-the-future>

⁸ Reporting from the UK Science Innovation Network

⁹ http://www.ibm.com/smarterplanet/us/en/smarter_cities/overview/

¹⁰ <http://strata.oreilly.com/2012/06/predictive-data-analytics-big-data-nyc.html>

sector. Providing readily accessible Big Data tools and analytics can allow commercial organisations, civil society and individuals to create new value for the public sector in a variety of ways. These could include feedback on services, insight into better management practices, and suggestions for improving new and existing programs. Big Data innovation can lead to experiments in public policy and public sector programs to improve government performance.

In the United Kingdom, the non-profit **Open Knowledge Foundation** used databases made available through the government's open data initiative to develop wheredoesmymoneygo.org, a site that makes it easier for citizens to view and understand UK public spending through analysis and visualization of the data.

Mastodon C¹¹ is an example of a big data start-up company incubated at the Open Data Institute and Open Healthcare UK. It worked with publicly available NHS prescription data to identify NHS prescription savings using big data. By using data on prescribing practice across England, variations in spending on different classes of drugs can be identified. It is then possible to calculate the potential savings to be achieved by moving from prescribing branded to generic drugs. For statins alone, the NHS could save around £200 million per year by reducing prescriptions of branded in favour of generic versions. When extended to all classes of drugs, the total potential savings could amount to £1.4 billion per year.

The Office of National Statistics is interested in the potential of Big Data to complement, augment or replace conventional data sources to improve the quality of official statistics. For instance, ONS has used Google trends analysis to provide early warning of migration trends (e.g. Time trends in the volume of Google searches for Polish terms correlates well with trends in the number of Polish migrants to the UK after the EU8 countries joined Europe in 2006-08).

¹¹ https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/198905/bis-13-743-market-assessment-of-public-sector-information.pdf