THE BIG DATA OPPORTUNITY

Hetan Shah
Executive Director, the Royal Statistical Society
@HetanShah

What steam was to the 19th century, and oil has been to the 20th, data are to the 21st. They are the driver of prosperity, the revolutionary resource that is transforming the nature of social and economic activity, the capability that differentiates successful from unsuccessful societies.

There is considerable interest in data and how they can improve policy, prosperity and even our democracy. There is particular talk of 'big data'. It is a term that is not well defined – perhaps better seen as 'mood word' which indicates the ubiquity of data and our growing interest in them. Big data hints at the rise of digital data which come from our phones, our supermarket purchases, and soon (with the emergence of the internet of things) our fridges and other household devices. We know we are in an era of big data through the explosion of data visualisation; the increased linking of datasets, interest in new types of data such as digital or administrative data; the rise of new ways of analysing data such as machine learning; and technological change which means we can gather and store more data than we ever dreamt possible.

... the driver of prosperity ...

The opportunities from greater usage of data for better policy are limitless. In health we will see the rise of personalised medicines, as well as improved methodologies for clinical trials. Better local data could lead to 'smarter cities' which can monitor transport needs or improve energy usage through real time data. The Foreign Office is already experimenting with analysis of social media data to find key influencers who can act as bridges between communities in troubled hotspots. The Office for National Statistics is trialling how to use administrative data (the data we give to government for other non statistical purposes – eg applying for a driving licence) to supplement and perhaps even supplant the traditional decennial census in order to provide more real time information about the society we live in.

TAKING DATA SERIOUSLY IN POLICY

To make the most of the big data opportunity, evidence must be taken more seriously in policy formulation and evaluation. Making policy when resources are tight is difficult but decision makers should take into account the probable quantified consequences of alternatives.

The Government has made welcome investment into the 'What Works' centres, which link academics with policymakers to cast evidential light on key issues. The Behavioural Insights Team or 'nudge unit' has focused on the small changes that can make a big difference, such as what letter wording from the tax man leads to more people paying their taxes. To build on this, Government should publish the data and evidence that underpin any new policies it announces, and should also commit to regular and long term evaluation of policies. Where we lack the data to inform choices between options in important policy areas, we should invest in getting it.

DATA SHARING

Greater data sharing between government departments for statistics and research purposes

... Government should publish the data and evidence that underpin any new policies ...

would provide opportunities for a range of public services and policy areas ranging from 'smart cities' to better healthcare. One option would be to follow the Canadian model and allow the Office for National Statistics access for statistical purposes to all major public and private sector datasets – which are already regulated and controlled – to focus not on us as individuals, but on how society is changing as a whole. Increasingly important data are held by the private sector – eg think of what aggregating everyone's mobile phone data or supermarket loyalty card data could help tell us about our society's lifestyle, movements, dietary habits etc. These new data sources can provide more real time evidence than was previously possible. Our research [1] with Ipsos MORI suggests that the public supports data sharing if it is done for the public good by organisations

trusted to safeguard privacy and confidentiality. Safeguards should be built into any sharing of personal data at the outset including anonymisation of the data (as far as is possible), legal penalties for misuse, access only to accredited researchers where appropriate, and gate-keeping access to big linked data sets so that personal identities are protected, as in the UK's Administrative Data Research Centres. There needs to be

... important data are held by the private sector ...

more detailed reporting of cases where personal data is shared with companies, as this is the area on which the public most want reassurance.

OPEN DATA

We can also make considerable gains through the 'open data' agenda which seeks to make data more open, accessible and reusable. The best known examples come from transport data – for example, Transport for London has made its transport data open, and as a result companies have made apps which tell you when your next bus is due. As well as making society more convenient, this is economically productive as it reduces time

... Open data can be a rich source of innovation at relatively little cost ...

that might otherwise be wasted. Open data can be a rich source of innovation at relatively little cost – particularly if it is effectively marked with standard codes for geography, time and other attributes. Geospatial data such as postcode address files are the core reference data upon which society depends, and also act as a catalyst to release economic value from other open datasets. Government has made some progress in this area but can do

more to open these up. And in future if public entities are privatised, their underlying data should not be lost to the public good which is sadly what happened to the Royal Mail's Postcode Address File.

PRIVATE SECTOR DATA

The private sector has an important role to play in sharing and opening their data. Companies should be encouraged to share data with

researchers for research purposes, to share the data they hold about individuals with those individuals, and to publish open data for everyone, for the public good. More and more important data is being held by private companies and much of them are not commercially sensitive. If they were released as open data we would know more about the country we live in – eg the location of supermarkets, ATMs, post boxes and water company boundaries. I would also like to see the rigour shown around official statistics in the public sector, and financial statements in the private sector, extended to other crucial information sources, such as the clinical trials reported by

pharmaceutical companies. And as we move to a wider range of public service providers, schools and hospitals and other public services provided by private providers should adhere to the same data standards and transparency as those in the public sector, so that we are able to monitor services across the board.

SKILLS FOR THE DATA ECONOMY

To prepare for the increasing

'data economy' that we inhabit we need to skill up the nation. More and more people – eg teachers, nurses, will have to handle data to do their jobs. But as a nation we don't have the skills. Our research[2] in association with Nesta and Creative Skillset supports that skills shortages are a pressing issue for the UK's data industries. In part this will need to be addressed in our education system. We should ensure that all young people learn to handle and interpret real data using technology, and should train teachers from primary school through to university lecturers to encourage data literacy from an early age. Basic data handling and quantitative skills should be an integral part of the taught curriculum across most A level subjects. The Nuffield

... skills shortages are a pressing issue for the UK's data industries ...

Foundation found in 2010 that across 24 countries, England, Wales and Northern Ireland had the lowest level of participation in the study of any kind of mathematics post-16. This has a severe impact on higher education and employment. Recent rises in A Level Mathematics participation are welcome, but wider improvements are needed. We should also put resources into making sure that new A levels and AS levels, as well as new Core Maths qualifications for those who do not wish to take A level mathematics, deliver appropriate statistical skills.

As well as improving everyone's skills to a basic level, we need to ensure we have the higher level skills for data analysis that will give our top companies and universities the edge that they need. As a recent report[3] by Nesta and Universities UK indicated, to achieve this Research Councils

UK probably needs to take a cross-cutting interest in data analytics so that action on this issue is properly co-ordinated. We also have to ensure an appropriate balance between capital and revenue spending on science – big data is not just about high powered computers, but highly skilled analysts who can understand the data. And we have to ensure that all universities retain their data analytics capabilities. It is good to see the new Alan Turing Institute as a home for data science, but for good science all universities need to retain a strong data analytics and statistical unit, both for the strength of UK innovation in this field, and as a service to support other scientific disciplines. Otherwise we will continue to hear about the 'reproducibility

crisis' where advances in our science and research base are too limited due to a lack of collaboration on the underlying data.

There is also a wide range of public service professions for which data skills are increasingly important, including in government. Politicians, policymakers and other professionals in the public sector should be given basic training in data handling and statistics. To address these needs and following our pre-election 'Parliament Counts' campaign, the Royal Statistical Society is glad to offer workshops for MPs and their staff on basic statistical concepts.

References

1 http://www.statslife.org.uk/news/1672-new-rss-research-finds-data-trust-deficit-with-lessons-for-policymakers

2 http://www.statslife.org.uk/files/model-workers-report.pdf

3 http://www.nesta.org.uk/publications/analytic-britain-securing-right-skills-data-driven-economy