STATEMENT OF INTERESTS

Mark Birkin is Professor of Spatial Analysis and Policy in the School of Geography, University of Leeds.

Kevin Macnish is a Teaching Fellow and Consultant in Applied Ethics at the University of Leeds.

Matthew Robson is Professor of Marketing, Leeds University Business School.

Professor Mick Yates is Visiting at the Leeds University Business School. Until recently, he was Head of International Markets at dunnhumby, a UK based global leader in customer science using Big Data and real time analytics. Prior to that, he spent almost 30 years at Procter & Gamble and Johnson & Johnson. His last role as Company Group Chairman held responsibility for all aspects of the Asia Pacific Consumer businesses.

EXECUTIVE SUMMARY

- Big data and social media analysis will allow UK businesses to gain and maintain an edge in understanding and responding to consumer demand.
- Government investment is needed to overcome a dearth of training, providing direction in educational programmes which emphasizes the combination of data analytics with business skills.
- Ethical concerns include the gaining of informed consent, security, and use of the data in the public interest.

BACKGROUND

The University of Leeds hosts a Consumer Data Research Centre in the Economic and Social Research Council’s Big Data Network. The mission of the Centre is to negotiate and make available diverse commercial and other sources of data relating to individual behaviour, consumption and decision-making. As contributors to this programme, the authors of this document therefore have a direct interest in the safe and ethical provision of social media and other data in real or near to real-time.

RESPONSE

How can real-time analysis of social media data benefit the UK?

1. Big data has the potential to power investigation from the macro level of global attitudes and trends to the personal level of individual psychology. The data collected go beyond demographics to include people’s private life via social media. Combining this information with other sources such as credit card data, supermarkets loyalty card data, and various others, allows the linkage of private life with consumer behaviour at a deep level. The consumer’s decision-making process can potentially be traced, identifying patterns of decision-making and anticipating imminent purchasing needs.

2. Big data will allow businesses to establish that purchases of certain products might be predictive of behaviour in other sectors. Better and more effective consumer outcomes could be obtained by optimising consumer choices in light of individual lifestyles.
3. Organisations may also employ big data to offer tailored experiences and exploit individual contacts. Detailed knowledge of individuals (either directly or through statistical analysis) will provide the potential to adapt processes to suit the individual.

4. Historical storage of data will further allow for the tracking of personal data over time. Coupled with the level of individual data, organisations will be able to match product to need more rapidly and effectively. This will increase the value of customers to these organisations over time.

5. Analysis of real-time sentiments will allow businesses to effectively influence their customers. The ability to profile users’ decision patterns may help to facilitate and direct purchasing decisions. This would occur through offering information to the customer that is relevant and timely to their concerns, and responsive to needs in real time. Monitoring social media allows companies to view brand mentions which may be positive or negative. Communications can then be geared to “respond” to these sentiments, tailored to the user in tone, content and appeal. The user is singled out not only by information such as interests, age, gender and other socio-demographics, but by mood or nuances in communication. This feedback loop will drive ahead the efficiency and effectiveness of marketing communications.

6. Social media is a rich source of consumer creativity which may benefit business through crowdsourcing and constructive exchanges between peers, as well as through more controlled viral marketing campaigns. Diverse connections between like-minded individuals can lead to inspiration, learning, and growth. Businesses may make real-time data-informed recommendations which include similar products/information, or to encourage the take-up of new and potentially attractive products/information.

7. Furthermore, mapping real-time social media data, particularly across different platforms, can help map how ideas travel across social networks and how to identify the roles that individuals play across those networks (leading to an increasing emphasis on the role of social media stars within networks in real time).

8. Data concerning individuals may also be geographical. Users disclose their location on a number of networked services. People movements, information searches, buying patterns, and related behavioural data can be combined to deliver optimal retail solutions and promotions driven by current activity. Linking searches to geographical space can provide solutions linked to spending opportunities, occasions, and preferred timing for transactions.

9. The nature and scale of social media lends itself to the national and international scale. Movements, behaviours and options of tourist visitors are tracked. Tourists are valuable customers who are likely to spend heavily and contribute to local economies. Opinions – positive and negative – of their experiences can guide service providers and contribute to UK brand positioning internationally.

10. Social media may service the interests of public organisations and governments directly. This could be through the placement of messages within media and communications, including TV and radio as well as online, with extensive value in health, welfare and education. Instantaneous monitoring and response will have particular impact in times of political turmoil. During the Arab Spring, real-time data from social media and search engines was used by international agencies to plan effective communications and response – as well as playing a significant role in the coordination of local political activity.

11. Secondary analysis of social media data is also of importance to commercial, governmental and academic organisations. Current academic research includes monitoring daily urban mobility patterns, consumer behaviours, and decision-making as revealed by social media and related big data sources. This can inform decision-making in relation to emergency and
disaster planning and to monitoring crime and disorder. New data sources can inform retail
organisations about stocking and pricing policies, facilitate moves towards more sustainable
packaging and purchasing behaviours, and provide information on obesity and diabetes in
relation to lifestyle, attitudes and shopping habits. The UK has a sophisticated retail and
market analytics industry. Making real-time social media available will build analytical skills
in both the commercial and academic sectors, which can develop this leading and competitive
role on the world stage.

What should the Government be doing to maximise these benefits?

12. Government data should continue to be made available as a means of generating public trust,
confidence and engagement, in addition to broadening the resources available to commercial
organisations.

13. Government strategies for the funding of educational programmes and partnerships between
universities, business and public sector organisations should continue to encourage the
sharing of commercial data and its exploitation in academic research and teaching. Co-
operative developments with other states (e.g. the EU, US, and China) are of crucial
importance to maintain the UK at the forefront of a globalised information economy. A
combination of reassurance, regulation and legislation is needed to maintain public trust in the
appropriate and secure use of individual data.

What are the barriers to implementing real time data analysis?

14. The major societal barriers to implementation are public education, management
understanding, and the availability of trained individuals who can bridge the gap from deep
data understanding to context-specific action and impact. The major technological barriers to
implementation are curation, analysis of visual language, and linkage of data sets.

15. Public understanding of big data, social media and analytics is at best patchy and at worst
hostile. Management understanding is equally patchy, and there is a dearth of trained
individuals in this area. Those working in data analysis (“data scientists”) tend to have a solid
foundation in computer science and applications, modelling, statistics, analytics and
mathematics. Beyond this, strong business acumen, coupled with the ability to effectively
communicate findings to businesses, is essential. Good data scientists do not just address
business problems; they identify problems that have the most value to the organization.

16. Strategies and policies are required to facilitate curation throughout the lifetime of the data
lifetime. This includes the type of information available, the retention and protection of data,
the handling of collection and linking additional information to subjects, and the implications
of evolving processing abilities.

17. Big data analytics are needed to analyse visual language to avoid excluding a large part of
social media. Much of what is shared online is pictures and videos. These spread faster than
pure text in social media. Some of the larger players, such as Pinterest or Instagram, are visual
platforms.

18. Linking across data sets must be robust to promote real-time applications. Finding
relationships across data sets does not necessarily equate to explanation, and the enormous
number of potential correlations in very large datasets increases the risk of finding spurious,
as well as valuable, correlations. Data are evidence which drives business and marketing
decisions, but such actions need to be grounded on sound theoretical and conceptual bases.
Academic thinking has an important role to play in this process.
19. The Government data capability strategy is robust. However, to take full advantage of the data opportunity more attention should be paid to the interdisciplinary nature of the task at hand, and associated education needs. Data scientists have been described as “part analyst, part artist.” Good examples of education exist in both the UK and US. The University of Bedfordshire offers a BSc in Data Science with collaboration from computer science, social sciences and the business school. This builds a critical understanding of the phenomenon of “big data” coupled with much-needed business skills. The Illinois Institute of Technology Master of Data Science programme is designed for students with backgrounds in mathematics, computation, life and physical sciences, engineering, and business. It combines high-level mathematics, statistics and computer science theory with the business acumen to explore data sets, gather insights, visualize results and communicate meaningful findings.

20. Too many UK-based BSc and MSc programmes in data science currently pay insufficient attention to business or innovation. This is not a criticism of any institution, as many programmes are world class, and most institutions in the UK and globally take the same approach. However, analytics alone are just half of the story – turning the analytics into action, organisation change and business innovation is the other half. UK institutions must systematically address curriculum needs, blending technology, and business in a range of courses at different levels. No longer is “IT” the issue. It is about using analytics to drive action and innovation in enterprises of all kinds.

21. Ethical concerns are widespread regarding the collection and use of personal data. Major issues include informed consent, security, ownership and benefits in the use of data.

22. Informed consent is a standard requirement when an organisation seeks to make use of a person or their belongings in most contemporary settings. Overriding or ignoring the requirement for informed consent involves harm amounting to stealing a person’s data, demonstrates a lack of respect, and requires justification.

23. Individual’s personal data are conveyed in social media and big data. The data are shared by a particular person for a particular reason and with a particular understanding of the context and implications. If these data are then pooled and used for purposes other than those to which the participant has agreed, the data are being used without informed (or possibly any) consent.

24. A potential objection is that, in some cases, consent has been given through the signing of a contract (e.g. to sign up for a shop loyalty card). However, these forms are rarely read in detail or understood. Even if signing a form is considered an act of consent, it is not necessarily informed consent. Caveat emptor becomes exploitation if a large number are signing up for the service based on a limited understanding of the future use of their information.

25. A second objection appeals to ownership: a company owns the data and can do with it as the company pleases. Owning something does not give the owner carte blanche to use it as they choose, though. Ownership of medical records does not permit a GP to sell those records to insurance companies. Possible harm must be taken into account, both on the part of those to whom the data pertains and of those who may be affected by the data. Furthermore, it is disputable that companies do own the data they receive from individuals. If I share an original idea with you, it is still “my” idea. This is widely recognized in professional ethics, academic practice, and intellectual property and copyright law. While you now have access
to my idea, you do not own it and are not justified in using it at will without reference to me or gaining my permission.

26. A third objection to seeking informed consent is that of public benefit (e.g. service improvement benefits of sharing medical records). However, if these benefits rely on unethically-sourced data then they are problematic. A more ethical approach would be to seek permission from patients in order to build a database to which those contributing have consented. This may take longer to compile, but it will meet the ethical concerns raised above.

27. Data security is a constant problem. The US business Target was hacked in December 2013 leading to the loss of data pertaining to 40m credit cards, and up to a further 70m phone numbers and zip codes of customers. High end retailer Neiman Marcus, a number of well-known hotel chains, and software company Adobe were also hacked in 2013, leading to the loss of millions more credit card details.

28. Data can be made secure, but true security is difficult to achieve. The businesses named above had made a considerable investment in security, and yet still fell victim to hackers. In the case of Target it seems likely that the breach was down to human error and warnings were ignored. Security is therefore only ever as good as the people who use it. This is a truth long recognized in intelligence, where Top Secret documents have been stolen, left on trains, or leaked to the press.

29. Once collected, data may be abused. People are wont to use databases for personal ends. It has recently transpired that several hundred US intelligence analysts have inappropriately accessed national security databases to review information on current or former love interests. Similar revelations occur regarding the UK police on a not-infrequent basis. Large data sets can also be used on “fishing trips” to locate particular information. Hacking voicemail accounts is one (illegal) example. Statistical correlations have been demonstrated between freely available information and “private” information (e.g. homosexuality). This suggests that fishing trips could uncover legitimate but embarrassing information which could harm an individual or group. Finally, businesses may employ large data sets for economic advantage such as targeting advertising, sales, and offers to certain societal groups. This may lead to the neglect of other societal groups, typically those at the lower end of the socio-economic scale. This “social sorting” may be unjust if it exacerbates divisions in society.

30. One approach to security is anonymisation of the data. This is a means of protecting individuals when their data is made public, or in the event of a security breach. Genuine anonymisation is difficult to achieve, though. Data may be re-identified through a combination of sources to reveal individual characteristics which were intentionally concealed. Anonymisation is therefore helpful but does not provide total security.

31. Any retention of big data must be acknowledged to carry with it associated risks to those whose data is held. This does not mean that such retention cannot be justified, but the justification involves weighing risks and benefits. It would be wrong to mislead the public with suggestions that the holding of data can be made perfectly secure.

32. The harms of data being released include privacy violations, chilling effects, increased vulnerability, and decreased trust. While individuals may publish much private information in social media, this may be restricted or published for a specific purpose. To use this for any other purpose is a violation of consent, and to re-publish the data in another forum would be a violation of privacy. The knowledge that data may inadvertently become public may “chill” people from engaging in legitimate or even socially necessary activities such as visiting a doctor. If personal data is made available then individuals may become vulnerable to blackmail, identity theft, or a sense that their neighbours know all of their secrets. Finally,
trust in society may diminish if personal data is discovered to have been used for purposes for which consent was not sought.

33. A last concern is that of the interests being served in the use of big data. Typically these will be the interests of business and bureaucracy, and not necessarily that of the citizen. There is therefore an issue of exploitation if the possession of one is used to the benefit of the other. It is important to establish the extent to which a particular use serves “the public interest”, and what the public interest is in each relevant context.

March 2014