Big Data and The Third Sector: Opportunities and Challenges

Arthur Harvey
August 2016

Abstract: The purpose of this short report is threefold: first, to define Big Data; to show how voluntary organisations in the third sector can use Big Data in decision-making processes to improve current services as well as to develop new services, initiatives and programmes; and to enumerate some of the challenges involved with using Big Data in decision-making processes in the third sector.

Keywords: Big Data, Data Analysis, Education, Healthcare, Third Sector

Key Points:

- The timeliness, accuracy and reliability of the information or insights extracted from Big Data may lead to more confident decision making and better decisions, which will allow the third sector to achieve its objectives.
- Big Data Analytics, based on a comprehensive understanding of people's behaviour, needs and preferences, can allow third sector organisations to predict new services, initiatives and programmes.
- Big Data can help third sector organisations improve existing services and operate more efficiently and effectively by providing tailored services.
- The main challenges posed by the Big Data revolution involve privacy and confidentiality; shortage of talent, resources and data analysts; and cultural challenges.
- The challenges posed by Big Data presuppose a culture of mutual understanding and trust that will enable public, voluntary and private sectors across the country to work co-operatively towards common objectives that will improve the quality of life of local communities through the use of data analytics.

Introduction

About a year ago, the opinion polls failed to predict the outcome of the general election which unexpectedly saw the first Tory majority in Government in more than 20 years.[1] A month ago, once again, almost all polls failed to forecast the outcome of the UK's EU referendum.[2] Such a failure can have a profound impact on a country's or an organisation's future and development. This is a story of the shortcomings of small data, or the traditional methods of statistical analysis and predictive models.

In an ever-changing and complex world, a comprehensive understanding of people's behaviours, needs and preferences will play an important part in deciding how reliable, timely and/or accurate are the insights or information on which we base decisions. A combination and analysis of different types of data – for example, geographic, demographic, psychographic or behavioural data – can play an important part in achieving higher standards of reliability, accuracy and timeliness when developing or revising initiatives.[3]

While charity and voluntary organisations will not be able to utilise Big Data as a solution to every problem, they should recognise the potential of using Big Data to improve current services and develop new services, initiatives and programmes.

What is Big Data?

In theory, Big Data is ''extremely large data sets that may be analysed computationally to reveal patterns, trends and associations, especially relating to human behaviour and interactions.''[4] Practically, Big Data or Big Data Analytics is the process of deriving or extracting insights from large data sets, analysing a variety of data ''to uncover hidden patterns, unknown correlations, market trends, [and] customer preferences …"[5]

There are two general approaches to decision-making and problem solving: the qualitative approach and the quantitative approach.[6] To analyse or solve a problem, an analyst or policy maker uses either or both of the two approaches. The qualitative approach to policy making is inductive, subjective and more flexible.[7] It gathers information that is not in numerical form, drawing on policy makers' experience and expertise to provide potential solutions to a problem.[8]


University of the Fraser Valley

The quantitative approach to policy making, as the term suggests, is concerned with the collection and analysis of data in numeric form. It is an approach that yields more comprehensive, objective and deductive results.[9] It uses mathematical reasoning and/or relies on scientific evidence or established facts to derive or deduce new knowledge and insights regarding a problem.[10]

Thus, the insights and information derived from Big Data, when accurately analysed and interpreted, can constitute the latest and most accurate form of scientific evidence because of the data's three distinct characteristics: 1) Volume, 2) Velocity and 3) Variety.[11]

1) **Volume** refers to the extremely large amount of data collected and analysed. In the past, storing and analysing such extremely large data was impossible, but new technologies (such as Hadoop software and R, the programming language) have made it possible.[12] About 90% of the world's data today was generated during the past two years, with 2.5 quintillion bytes of data added each day.[13]

2) The **velocity** of Big Data is evident in the fact that data streams in at an unprecedented speed and is analysed in a timely manner or in real-time.[14] For instance, when an individual does an Internet search, likes a post on Facebook, shops online or uses a loyalty card somewhere, he or she creates personal information. This data, combined with other types of data, is captured and then processed in real time to deliver personalised and targeted initiatives.


Litterley CE Primary School

3) **Variety** of Big Data is discernible from the fact that data is collected from a variety of sources, including business transactions, and social media. Such data may also come in all types of formats – from structured, numeric data email, video, audio, stock ticker data and financial transactions.[15]

With the advent of cost-effective tools to analyse this data, though the private sector and public sector are leading Big Data application development, organisations in the third sector can experience great advantages using Big Data or Big Data Analytics in decision-making processes to improve current services and develop new services and initiatives.[16]

New Services, Initiatives and Programmes

Governments and public authorities everywhere acknowledge the profound potential influence of Big Data on future developments.[17] David Willetts MP, the former UK Minister for Universities and Science, said in 2014: "Data is a huge priority for government as it has the potential to transform public and private sector organisations, drive research and development, increase productivity and innovation and enable market-changing products and services.''[18] Research conducted recently by Nesta shows that data-driven firms and organisations in the UK are 40% more likely to report launching new products and services ahead of their competitors that are not data-driven firms.[19] So how can keen insights extracted from Big Data help the third sector develop new services, initiatives and programmes?[20] How can examining the applications of Big Data by leading organisations, public and private, reveal how the third sector can predict new services and programmes and thus achieve similar successes?[21]

A thorough understanding of people (including service users) will allow voluntary organisations to identify services that are needed.[22] There are numerous examples of how Big Data Analytics based on a comprehensive understanding of people's behaviour, needs, preferences can predict new services and programmes.

For example, mobile phone calling patterns have helped researchers: 1) detect flu outbreaks;[23] 2) identify the socioeconomic status of the phone users and their access to housing, education, healthcare and services such as water and electricity;[24] and 3) determine the movement of displaced populations after the 2010 Haïti earthquake, aiding in the distribution of resources.[25] Another example is how the United States Postal Service collects vehicle data from sensors on each of its many delivery trucks and uses predictive algorithms to delegate preventive maintenance, resulting in lower maintenance costs and ensuring timely delivery.[26]

Perhaps more inspiring examples of Big Data applications in the development of new services and initiatives are in healthcare and public health. Big Data has led to "significant discoveries, the development of new therapies and a remarkable improvement in health care and public health."[27] For example, Banner - University Medical Centre, a leading hospital in the USA, used Big Data Analytics to identify which patients were more likely to be readmitted to hospital in order to reduce operational costs.[28] Big Data Analytics helped Banner identify new potential services by identifying the need to equip the patients with telehealth monitoring technology at home, and a mobile intensive care team was equipped to respond to wireless alerts from the sensors, allowing medical staff to visit patients' homes and intervene at the first sign of trouble.

Alternately, educational institutions are using Big Data in decision-making to identify better teaching strategies, highlight areas where students may not be learning efficiently and transform the delivery of

education.[29] For example, Big Data Analytics helped analysts determine which students in a course at the University of British Columbia were likely to fail a class so that instructors may intervene in time to assist them.[30]

Thus a comprehensive understanding of service users through data can help voluntary organisations develop new initiatives suited to their changing needs.

Improvement of Existing Services

Big Data can help third sector organisations operate more efficiently and effectively by providing services tailored to people's needs and preferences. For example, MacMillan Cancer Support has analysed NHS data sets, including those of cancer registration and mortality, and used predictive analytics to predict how many people will be diagnosed with cancer, what types they will have and what their needs will be in one to three years' time.[31] Big Data has helped the charity rapidly gain a comprehensive understanding of people's needs and allowed them to plan services for various areas' populations. Mike Hobday, the former director of policy and research at MacMillan, commented: "Charities can be entrepreneurial in terms of saying where significant interventions can be made using big and open data because we are not caught up in the daily hurly-burly of the NHS."[32] In addition, Cancer Research UK use Big Data analytics with DNA sequences to develop personalised cancer treatment.[33] Meanwhile, Columbia University Medical Centre is analysing physiological data to offer targeted and personalised treatment that is proactive rather than reactive.[34]

In education, Knewton, one of the world's leading adaptive learning companies, analyses the progress of millions of students to create better test questions and personalised course goals. In addition, Keyfund, a Newcastle-based charity, used demographics and other types of data related to young people that use its programme each year to improve and personalise its services. The charity found that, among other things, young people from deprived backgrounds benefited more from leisure projects than from other schemes, and that girls from an Asian background progressed better when supported by females.[35]

Thus a comprehensive understanding of people's needs can help voluntary organisations improve their current services by developing services more responsive and tailored to individual users.

Challenges

While Big Data may be a valuable tool for non-profit organisations, there are a number of challenges they need to overcome to take full advantage of the Big Data revolution. For the purpose of brevity this report will focus on three:

1) Privacy and confidentiality;
2) Shortage of talent, resources and data analysts; and
3) Cultural challenges

1 – *Privacy and Confidentiality*

One of the biggest concerns with the copious amounts and types of data available is confidentiality. With the tremendous amount of data being collected and analysed constantly by statistical agencies and organisations, privacy has become a key concern. This is particularly the case when an individual's personal data is inappropriately included and predictive analysis is conducted without their knowledge or express

consent. For example, in 2012 a well-publicised *New York Times* article revealed that the retail chain Target had used data mining techniques to predict which female customers were pregnant, even if they had not yet announced it publicly.[36] This activity resulted in the unauthorised disclosure of personal information to marketers.[37]


Russell Watkins

Since social organisations are more likely to deal with data on vulnerable population groups and sensitive topics such as child welfare, domestic violence and drug and alcohol misuse, the issues of privacy and confidentiality are of primary consideration to the third sector. A number of statistical methods which address privacy and confidentiality concerns, such as suppression, aggregation and adding random noise, can mitigate the privacy issues.

2 – *Shortage of Talent, Resources and Data Analysts*

Perhaps a greater challenge associated with the implementation of Big Data techniques is the shortage of talent and resources available to access the necessary data and to utilise it, in all sectors but particularly in the third sector both globally[38] and in the UK[39]. This ranges from structural constraints such as computing power and software tools required to integrate, analyse and visualise Big Data, through to a need for data scientist with specialised skills in integrating datasets in a meaningful and valuable way. The UK government predicts an increase in demand for big data staff of between 13% and 23% per annum between now and 2017.[40]


Marco Antonio Torres

New data management systems aim to meet the challenges of Big Data; for example, Hadoop, an open-source platform, is the most widely applied technology for managing storage and access, overhead associated with large datasets and high-speed parallel challenge for many businesses and voluntary organisations, especially small- and mid-size ones, as such applications require expertise and experience not widely available and may thus need outsourced help.[42]

Finding the right talent to analyse Big Data is perhaps the greatest challenge for voluntary organizations, as the required skills are neither simple nor solely technology-oriented. Searching for and finding data scientists – competent in datamining, visualisation, analysis, manipulation and discovery – is difficult and expensive for most organisations, and for some it is simply unfeasible.

3 – *Experience Vs. Data Science*

Whilst experts tend to agree that the use of Big Data analytics in decision-making is important, in practice, in the context of the third sector, this can be challenging. A transition to practices more informed by data (or

Big Data) may mean that the charitable objectives of an organisation in the third sector have to be driven by data science, and not by experience and personal insights only.[43] This paradigm shift can be especially difficult to be imposed on an organisation with years of implementing a particular process.[44]

Another aspect of this is the willingness to take different actions if the data suggests that this would be the right thing to do. This can be a very challenging position to take, as it might result in an admission that what an organisation has been doing is not the most effective use of resources.

Conclusion

There is no doubt that Big Data has the potential to achieve worthwhile social goals and bring about lasting social changes. Big Data Analytics can particularly benefit the areas of health care, education, poverty, unemployment and social exclusion – these areas are where traditional analytic techniques fail most to deliver results, or where the most challenges are faced by governments and local authorities. Big Data should not be seen as a solution to every problem faced by voluntary organisations or a substitute to the use of small data in decision-making processes, given the challenges inherent in using Big Data Analytics, but as a tool integrated into an organisation's best practices to achieve a greater efficiency and innovation.[45]

Recommendations

1. Work with and support voluntary organisations and the statutory sector to explore and trial the use of Big Data and / or more data-driven approaches to inform policy making and service design.
2. Consider how such a change in their approach can operate within the organisations' context.
3. Assess the role Big Data or more data-driven approach may play in developing and improving initiatives and programmes taking into account the organisations operational and strategic priorities and constraints.
4. Promote and foster a culture of mutual understanding, trust and confidence which will enable public, voluntary and private sectors to find effective ways of working co-operatively towards the achievement of common objectives and outcomes through the use of Big Data that will improve the quality of life of local communities.
5. Adopt a good practice code to help organisations adhere to the highest ethical standards in its use of data analytics.
6. Urge strategic and operational cooperation and partnership between the third and public sectors.
7. Ensure that organisations have access to the necessary data to inform the design and delivery of services.
8. Provide technical assistance programmes and support for third sector organisations contemplating implementing data science in decision-making processes.

Acknowledgements
Special thanks to Gaia Croston (People Know How), Jenni Inglis (Managing Director, VIE – Value, Involve & Evolve), Glenn Liddall (Head and Founder, People Know How) and Stef Milenkovic (Development Worker, Edinburgh Voluntary Organisations Council).

References

[1] P. Whiteley, "Why did the polls get it wrong in the 2015 General Election? Evaluating the inquiry into pre-election polls," *The Political Quarterly*, 20 July 2016, <http://onlinelibrary.wiley.com/doi/10.1111/1467-923X.12274/full>.
[2] *Ibid.*
[3] *Ibid.*
[4] O. Tene and Jules Polonetsky, "Big Data for all: Privacy and user control in the age of analytics," *Northwestern Journal of Technology & Intellectual Property*, 11:5, 2013, pp.239-40 [Defining Big Data to include personal data generated from a variety of sources].
[5] M. Galetto, "What is big data analytics?" *NGDATA*, 5 July 2016, <http://www.ngdata.com/what-is-big-data-analytics/>.
[6] T. Håkonsson and T. Carroll, "Is there a dark side of Big Data – point, counterpoint," *Journal of Organization Design*, 5:5 (2016); S. Mack, "Roles played by the qualitative & quantitative approaches to managerial decision making," *Houston Chronicle Small Business*, <http://smallbusiness.chron.com/roles-played-qualitative-quantitative-approaches-managerial-decision-making-35440.html>.
[7] Family Health International, "Module 1: Qualitative Research Methods Overview," *Qualitative Research Methods: A Data Collector's Field Guide*, Northeastern University, <http://www.ccs.neu.edu/course/is4800sp12/resources/qualmethods.pdf>.
[8] S. Lewis, "Qualitative inquiry and research design: Choosing among five approaches," *Health Promotion Practice*, (2015): 1524839915580941.
[9] A. Takhar-Lail and A. Ghorbani (ed.), *Market Research Methodologies: Multi-Method and Qualitative Approaches*, 2014.
[10] D. Brooks, "Op-Ed.; The philosophy of data," *New York Times*, 5 February 2013, at p.A23 [highlighting Big Data's potential]; *Strata 2013 Is a Wrap*, Strata Conference: Making Data Work, 26-28 February 2013, <http://strataconf.com/strata2013>, *archived at* <http://perma.cc/8KYZ-FPGQ>; Håkonsson & Carroll, 2016.
[11] Tene & Polonetsky, 2013; D. Boyd and K. Crawford, "Critical questions for Big Data: Provocations for a cultural, technological, and scholarly phenomenon," *Information, Communication & Society*, 15:5 (2012), pp.662-3; Håkonsson & Carroll, 2016.
[12] "Big data: What it is and why it matters," SAS, <http://www.sas.com/en_th/insights/big-data/what-is-big-data.html>.
[13] IBM, *Smarter Cities Challenge: Syracuse*, 2011; <http://smartercitieschallenge.org/city_syracuse_ny.html>.
[14] Y.A. de Montjoye et al., "Unique in the Crowd: The privacy bounds of human mobility," *Scientific Reports*, 2013, p.1, <http://www.nature.com/srep/2013/130325/srep01376/pdf/srep01376.pdf>, *archived at* <http://perma.cc/BUS5-26VS>
[15] Tene & Polonetsky, 2013; de Montjoye et al., 2013, p.240.
[16] P.C. Broekema et al., "Towards the ASTRON and IBM Center for ExaScale Technology," In *Proceedings of the 2012 Workshop on High-Performance Computing for Astronomy Data*, 2012, pp.1-4.
[17] Office of Science and Technology Policy, "Fact Sheet: Big Data Across the Federal Government," 29 March 2012, <http://www.whitehouse.gov/sites/default/files/microsites/ostp/big-data-fact-sheet-final-1.pdf>; Her Majesty's Government, *Seizing the Data Opportunity; A Strategy for UK Data Capability, 2013*, <https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/254136/bis-13-1250-strategy-for-uk-data-capability-v4.pdf>.

[18] University of Glasgow, "New urban research centre to break down Big Data," [press release] 6 February 2014, <http://www.gla.ac.uk/news/archiveofnews/2014/april/headline_306335_en.html>.
[19] H. Bakhshi and J. Mateos-Garcia, *Rise of the Datavores: How UK Businesses Can Benefit from Their Data*, Nesta, 28 November 2012, <http://www.nesta.org.uk/publications/rise-datavores-how-uk-businesses-can-benefit-their-data>.
[20] R. Abbas, "The social implications of location-based services: An observational study of users," *Journal of Location-Based Services*, 5:3-4, 2011, pp. 156-181.
[21] G.H. Kim, S. Trimi and J.H. Chung, "Big-data applications in the government sectors," *Communications of the ACM*, 57:3, 2014, pp.78-85, <dl.acm.org>.
[22] M. Bienkowski, M. Feng and B. Means, *Enhancing Teaching and Learning Through Educational Data Mining and Learning Analytics*, U.S. Department of Education Office of Educational Technology, 2012, <https://tech.ed.gov/wp-content/uploads/2014/03/edm-la-brief.pdf>.
[23] J. Giles, "Cellphones reveal emerging disease outbreaks," *New Scientist*, 13 October 2010, <https://www.newscientist.com/article/mg20827824-800-cellphones-reveal-emerging-disease-outbreaks/>.
[24] V. Frias-Martinez, "Data for Development," September 2012, <http://www.vanessafriasmartinez.org/Data_For_Development.html>.
[25] L. Bengtsson et al., "Improved response to disasters and outbreaks by tracking population movements with mobile phone network data: a post-earthquake geospatial study in Haiti," *PLOS Medicine*, 2011; <http://journals.plos.org/plosmedicine/article?id=10.1371/journal.pmed.1001083>.
[26] G. Satell, "Yes, Big Data can solve real world problems," *Forbes*, 3 December 2013, <http://www.forbes.com/sites/gregsatell/2013/12/03/yes-big-data-can-solve-real-world-problems/>, *archived at* http://perma.cc/VSW4-QVG3.
[27] P.M. Schwartz, "Information Privacy in the Cloud," University of Pennsylvania Law Review 161, no. 1623, (2013), p.1631.
[28] New Scientisti, "Big data, better health," Living Health, *New Scientist*, 4 November 2015, <https://www.newscientist.com/article/dn28340-big-data-better-health/>.
[29] Bienkowski, Feng & Means, 2012.
[30] L. Macfadyen and S. Dawson, "Mining LMS data to develop an 'early warning system' for educators," *Computers & Education*, 54:2, 2010, pp. 588-599.
[31] E. Pannaman, "Big data for social good in London," *Innovation Enterprise*, 19 July 2016, <https://channels.theinnovationenterprise.com/articles/big-data-for-social-good-in-london>; Third Sector, "Analysis: Crunching numbers for charity," *Third Sector*, 11 March 2014, <http://www.thirdsector.co.uk/analysis-crunching-numbers-charity/finance/article/1284284>.
[32] *Ibid* [see both articles].
[33] S.A. Mathieson, "Genomics England exploits big data analytics to personalise cancer treatment," *Computer Weekly*, October 2014, <http://www.computerweekly.com/feature/Genomics-England-exploits-big-data-analytics-to-personalise-cancer-treatment>.
[34] C.A. Weaver, M.J. Ball, G.R. Kim and J.M. Kiel (eds.), *Healthcare Information Management Systems: Cases, Strategies, and Solutions (Health Informatics)*, Springer International Publishing: Cham, 2015.
[35] Third Sector, 2014.
[36] C. Duhigg, "Psst, you in aisle 5," *The New York Times Magazine*, 19 February 2012, p.30.
[37] *Ibid.*
[38] J. Manyika et al., *Big Data: The Next Frontier for Innovation, Competition and Productivity*, McKinsey Global Institute, 2011, <http://www.mckinsey.com/business-functions/business-technology/our-insights/big-data-the-next-frontier-for-innovation>.
[39] *Deloitte, Market Assessment of Public Sector Information*, commissioned by the Department for Business, Innovation and Skills, May 2013;
E-skills UK, "Big data analytics: An assessment of demand for labour and skills, 2012-2017," on behalf of SAS UK, 2013.
[40] e-skills UK, 2013.
[41] Definition: Parallel processing or parallel computing is a type of computation in which many calculations are carried out simultaneously, or the execution of processes are carried out simultaneously.

[42] P.C. Zikopoulos et al. (eds.), *Understanding Big Data: Analytics for Enterprise Class Hadoop and Streaming Data*. (McGraw-Hill: New York, 2012).
[43] C.J. Sheperis, J.S. Young and M.H. Daniels, *Counseling research: Quantitative, qualitative, and mixed methods*. Pearson: London, 2010.
[44] Nominet Trust, <http://www.nominettrust.org.uk/>.
[45] K. Bodine-Forrester, "The Danger of Big Data," Service Design Network, <https://www.youtube.com/watch?v=ZZm-QI8nvrk>.

Edited by Gaia Croston

Contact **People Know How**
Website: peopleknowhow.org – Tel: 07714 586971 – Email: contactus@peopleknowhow.org
Community Space, Ocean Terminal, Ocean Dr, Edinburgh EH6 6JJ
Scottish registered charity number: SC043871