Contents lists available at ScienceDirect

Social Science Research

journal homepage: www.elsevier.com/locate/ssresearch

The role of administrative data in the big data revolution in social science research

Roxanne Connelly [a], [*], Christopher J. Playford [b], Vernon Gayle [c], Chris Dibben [d]

[a] Department of Sociology, University of Warwick, Social Sciences Building, The University of Warwick, Coventry, CV4 7AL, UK
[b] Administrative Data Research Centre – Scotland, University of Edinburgh, 9 Edinburgh Bioquarter, Little France Road, Edinburgh, EH16 4UX, UK
[c] School of Social and Political Science, University of Edinburgh, 18 Buccleuch Place, Edinburgh, EH8 9LN, UK
[d] School of Geosciences, University of Edinburgh, Geography Building, Drummond Street, Edinburgh, EH8 9XP, UK

ARTICLE INFO

Article history:
Received 8 July 2015
Received in revised form 5 April 2016
Accepted 13 April 2016
Available online xxx

Keywords:
Big data
Administrative data
Data management
Data quality
Data access

ABSTRACT

The term big data is currently a buzzword in social science, however its precise meaning is ambiguous. In this paper we focus on administrative data which is a distinctive form of big data. Exciting new opportunities for social science research will be afforded by new administrative data resources, but these are currently under appreciated by the research community. The central aim of this paper is to discuss the challenges associated with administrative data. We emphasise that it is critical for researchers to carefully consider how administrative data has been produced. We conclude that administrative datasets have the potential to contribute to the development of high-quality and impactful social science research, and should not be overlooked in the emerging field of big data.

© 2016 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

1. Introduction

Big data is heralded as a powerful new resource for social science research. The excitement around big data emerges from the recognition of the opportunities it may offer to advance our understanding of human behaviour and social phenomenon in a way that has never been possible before (see for example Burrows and Savage, 2014; Kitchin, 2014a,b; Manovich, 2011; Schroeder, 2014). The concept of big data is vague however and has never been clearly defined (Harford, 2014a,b). We contend that this is highly problematic and leads to unnecessary confusion. Multiple definitions of big data are available and many of these seem to unwittingly focus on one specific type of data (e.g. social media data or business data) without appreciating the differences between the various types of data which could also reasonably be described as big data. We argue that there are multiple types of big data and that each of these offer new opportunities in specific areas of social investigation. These different types of big data will often require different analytical approaches and therefore a clearer understanding of the specific nature of the data is vital for undertaking appropriate analyses.

We highlight that whilst there may be a 'big data revolution' underway, it is not the size or quantity of these data that is revolutionary. The revolution centres on the increased availability of new types of data which have not previously been available for social science research. By treating big data as a single unified entity social scientists might fail to adequately

[*] Corresponding author.
E-mail addresses: R.Connelly@warwick.ac.uk (R. Connelly), chris.playford@ed.ac.uk (C.J. Playford), Vernon.Gayle@ed.ac.uk (V. Gayle), chris.dibben@ed.ac.uk (C. Dibben).

http://dx.doi.org/10.1016/j.ssresearch.2016.04.015

0049-089X/© 2016 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

appreciate the attributes and potential research value of these new data resources. We argue that careful consideration of these different types of data is required to avert the risk that researchers will miss valuable data resources in the rush to exploit data with the highest profile.

In this paper we aim to provide a thorough treatment of administrative data which is one particular type of big data. Administrative data can be generally described as data which are derived from the operation of administrative systems (e.g. data collected by government agencies for the purposes of registration, transaction and record keeping) (Elias, 2014). We emphasise this form of big data for two reasons. First, we observe that administrative data has been largely neglected from many of the mainstream discussions of big data. Second, because administrative data are particularly valuable and may provide the means to address fundamental questions in the social sciences and contribute directly to the evidence base (e.g. answering questions relating to social inequality). This paper begins with a review of available definitions of big data, and we emphasise why administrative data should be characterised as a form of big data. We then consider how administrative data compares with the traditional types of data used in the social sciences (e.g. social survey data). Finally, we discuss the opportunities and challenges offered by the use of administrative data resources in social science research.

2. What is big data?

There is no single clear definition of big data. de Goes (2013) has gone as far as to suggest that the term big data is too vague and wide-ranging to be meaningful. In this section we summarise some of the definitions of big data in an attempt to bring more clarity to what big data constitutes.

Taylor et al. (2014) conducted a series of interviews with high profile economists working in this field in an attempt to better understand big data and its uses. These researchers identified the size and complexity of datasets as a key component of big data. Centrally, they emphasised that the increased number of observations and variables available in datasets were the result of a shift in the sources of data which were available to them (especially from the internet and social media). Generally the size and coverage of datasets are a central element of the definition of big data. Einav and Levin (2013) also emphasise that data is now available faster, and has a far greater coverage than the data resources which were previously available to social researchers.

Much of the literature discussing big data focuses on data which results from online activities and the use of social media (see for example Tinati et al., 2014). This type of data may be produced through online searches, internet viewing histories, blogs, social media such as Twitter and Facebook posts, and the sharing of videos and pictures.[1] The growth of the internet and electronic social networking has resulted in the unprecedented collection of vast amounts of data. The use of internet and social media data have resulted in numerous research studies investigating a wide range of topics such as individual's moods (e.g. Dodds et al., 2011), politician's impression management (e.g. Jackson and Lilleker, 2011) and collective political action (e.g. Segerberg and Bennett, 2011).

Big data should not be considered as synonymous with data collected through the internet. This is because big data can also originate from sources such as commercial transactions, for example purchases in-store from supermarkets or from bank transactions (see Felgate and Fearne, 2015). Big data can originate from sensors, for example satellite and GPS tracking data from mobile phones (see Eagle et al., 2009). Genome data is a source of big data and programs such as the '100,000 Genomes Project' in the UK and the 'Precision Medicine Initiative' in the US have resulted in the collection of massive amounts of data for the purpose of genome sequencing (see Eisenstein, 2015). Administrative data, for example education records, medical records, and tax records, are also sources of big data (see Chetty et al., 2011a,b).

Perhaps the most well-known definition of big data is provided by Laney (2001), who describes big data in terms of volume (i.e. the amount of data), variety (i.e. the range of data formats available such as text, pictures, video, financial or social transactions), and velocity (i.e. the speed of data generation). Tinati et al. (2014) highlight the all-encompassing nature of big data (i.e. the data captures all of the information from a particular platform such as twitter). Tinati et al. (2014) also consider the real-time nature of big data as one of its key features (i.e. big data may be captured on events or interactions as they happen).

Schroeder and Cowls (2014) emphasise that the concept of big data is strongly associated with a step-change in the types of data resources which are becoming available to researchers. Similarly, Harford (2014a,b) highlights that the 'found' nature of big data is one of its fundamental features. In the era of big data we are increasingly dealing with data resources that have been discovered by researchers as potential sources of valuable research data, but which have been collected for different (i.e. non-research) purposes. Traditional sources of data in the social sciences are 'made' by researchers. Even large scale social survey data resources which are used by many researchers, who are often working in different fields, to answer different questions are designed specifically for research purposes. By contrast big data are data resources which were collected for purposes other than research and researchers do not have any input into the design of these data or its content. A central characteristic that could be added to a definition of big data is that these data are not collected for research purposes but can be suitably re-purposed by social science researchers.

[1] This type of internet data is distinct from the exhaust data generated as trails of information created as a by-product resulting from internet or online activities (e.g. log files, cookies, temporary files). Exhaust data is of value to marketers and businesses (see Ohlhorst, 2012), however it tends to have less relevance for social science research.

Please cite this article in press as: Connelly, R., et al., The role of administrative data in the big data revolution in social science research, Social Science Research (2016), http://dx.doi.org/10.1016/j.ssresearch.2016.04.015

Download English Version:

https://daneshyari.com/en/article/7338982

Download Persian Version:

https://daneshyari.com/article/7338982

Daneshyari.com