Using Social Media Data Aggregators to Do Social Research

by David Beer
University of York

Sociological Research Online, 17 (3) 10
<<http://www.socresonline.org.uk/17/3/10.html>>
10.5153/sro.2618

Received: 7 Dec 2011 Accepted: 5 Mar 2012 Published: 31 Aug 2012

Abstract

This article asks if it is possible to use commercial data analysis software and digital by-product data to do critical social science. In response this article introduces social media data aggregator software to a social science audience. The article explores how this particular software can be used to do social research. It uses some specific examples in order to elaborate upon the potential of the software and the type of insights it can be used to generate. The aim of the article is to show how digital by-product data can be used to see *the social* in alternative ways, it explores how this commercial software might enable us to find patterns amongst 'monumentally detailed data'. As such it responds to Andrew Abbott's as yet unresolved eleven year old reflections on the crucial challenges that face the social sciences in a data rich era.

Keywords: *The Future of the Social Sciences, By-Product Data, Digital Data, Social Media, Data Aggregators*

Introduction

1.1 There has been much talk recently of the challenges and potentials presented to the social sciences by new forms of digital data, yet so far we have made little progress in exploring exactly what we might be able to do with such data or what type of insights they might offer. In fact, as I write this article it is now eleven years since Andrew Abbott pointed out that:

'The single most important challenge facing empirical social sciences in the next 50 years is the problem of finding patterns in such monumental detailed data. And the blunt fact is that sociology is woefully unprepared to deal with this problem: We have neither the analytical tools nor the conceptual imagination necessary.' (Abbott, 2000: 298)

1.2 On reflection it would appear that we have not really moved much beyond this moment of realisation. Indeed, the challenge outlined by Abbott remains largely unchanged and unresolved some eleven years on. We know that there is potential in these accumulating masses of data but we are unsure about how to realise it. This article asks if it is possible to use commercial data analysis software and digital by-product data to do critical social science. In response it introduces a new 'inscription device' (Osborne et al, 2008) that allows us to tap into these new forms of digital by-product data and then illustrates how this inscription device might be used to identify social patterns and to create insights about specific topics. In short, this article looks to draw upon the analytical powers of the commercial sector to imagine the potential opportunities for rethinking the social sciences and for responding to the challenges that such commercial social analytics create for the jurisdiction and purpose of the academic social sciences (Savage and Burrows, 2007).

1.3 The software I introduce here can be described as Social Media Data Aggregators or Social Media Listeners. The aim is to introduce social media data aggregators to a social science audience and to explore what use this software might have for developing new possibilities in social research, for finding patterns in masses of data, for generating new practices and for creating new types of findings. To put this in the terminology of John Law's (2004 & 2009) recent influential work, this article reflects upon how this software might be used to 'enact' the social in alternative ways to that 'performed' by our existing methods. These data aggregator packages are already widely used in the commercial sector where they are used to follow aggregate level responses and 'sentiments' towards brands, to measure influential voices and 'buzz', and to measure temporal changes in how brands are perceived. Despite the apparent analytical power of data aggregators this software has yet to be adopted by academic social scientists.

1.4 The article begins with a brief description of the way in which by-product data is generated through everyday engagements with social media and explains how social media data aggregator software draws upon this data. The article then uses specific examples to explore the type of insights that this software is able to generate, it provides a series of illustrative insights that highlight the functionality of the software. The article concludes with some critical reflections on the use of social media data aggregators for doing social research. The conclusion also provides some closing thoughts on how we might proceed

further with this project.

Social media and by-product data

2.1 Let us begin with some very brief reflections upon the type of data that this software draws upon. Given that they are now so familiar it is not really necessary or appropriate to spend a significant space in this article describing the developments that are commonly referred to as social media or Web 2.0 (see Beer et al, 2007). These have been widely discussed and the chances are that the reader is already extremely familiar with these new media forms. In the last five or six years these participatory web cultures have become an established part of everyday life across various national contexts (Ritzer and Jurgenson, 2000). There are of course many who opt out of such practices, but these media are undoubtedly now part of the cultural mainstream and have become deeply embedded in everyday routine. The reader is also likely to be aware of the scale of the statistics about the population of Facebook users (which has been reported to be over 600,000 individual profiles) or the quantity of profiles and postings on Twitter (this is not to mention the many other popular applications like Youtube and the like). We need only watch a television news broadcast to see the frequent part that Twitter plays even in shaping and informing the broadcast news media – the significant international coverage of the use of Twitter to undermine privacy injunctions and super-injunctions in 2011 is one such example. In academia we find a similar integration of social media occurring, the British Sociological Association has a presence on Facebook, the 2011 BSA annual conference had a significant presence on Twitter, and many of our colleagues are forming profiles and networks on targeted niche social networking sites such as academia.edu. I could continue in some detail, rather I would like to assume a relatively well-formed knowledge of these developments even amongst those who do not participate but have probably heard the news stories and the like.

2.2 The development of such media are something of a side issue in this particular piece, all we need to accept at this point is that these developments have become embedded in the everyday practices of vast numbers of people in a variety of locations (as I will demonstrate in a moment). The next thing to consider is the way in which these activities generate traces within digital data that capture these everyday practices and which contain details about the people engaging with and through them. The acts of prosumption (Ritzer and Jurgenson, 2010) come together with what Zygmunt Bauman (2007) has described as the cultures of a 'confessional society' in social media, the by-product data generated form into what might be thought of as 'archives of the everyday' (Featherstone, 2006; Beer et al, 2008). This is the *social* in social media preserved in the archived content. This vast archived content is available for social scientists to use if we so wish and if we find a means for doing so. This is a clear example of the widely discussed forms of social data that are emerging in a digital mediascape that have become part of recent debates on the future of the social sciences (Adkins and Lury, 2009). The question is whether we can see the social in this by-product data. Is this a new form of social data from which social scientists may engineer insights? The value of the digital traces left by interactions with social media has been acknowledged for some time now but has yet to be exploited. Social media of various types are now the sites of genuine social and cultural activity at a massive scale and it is important to open the net so as to consider the software that is emerging that may allow us to see the social in this by-product data.

Social media data aggregators? Capturing and using by-product data

3.1 In brief terms, social media data aggregators (SMDA) capture and facilitate the analysis of this accumulating social by-product data. This software is predominantly used commercially to track the 'buzz' around particular brands. These software packages, of which we have selected one to work with^[1], are themselves going through a period of rapid development. In a sense I will only be talking here of the current incarnation of this changing software, discussions with the software developers reveal that the current possibilities and functionality are only a snapshot of a mobile set of developments – which of course presents the classic problem of description in a time of social change. We are only at the beginning of the emergence of such software. This development is depicted by the developers in starkly 'posthumanist' terms (Hayles, 1999). These descriptions suggest that the type of 'machine learning' that underpins the technology is being honed, and that the software can also be 'taught' by those using it. This vision of autonomous machinic researchers might need to be tempered but still this suggests some important developments over the coming years that cannot be fully anticipated in this article. My hope is that the reader will see the potential based upon these initial descriptions but that we can also consider the likely future development of such devices for doing social research.

3.2 As things stand SMDA are already capable of some startling and revealing analytics. The package I have been working with, created by a company called Insightlytics, provides the user with a wide range of analytical options around chosen search terms. To give a sense of the commercial nature of these products, the company providing the software describe themselves as follows:

'Insightlytics develops cutting-edge software that helps organisations extract valuable business insights...Ever increasing use of social media and the availability of vast amounts of online conversations provide a never-ending stream of insights on brand perception, changing expectations and desires. We have developed software that can isolate important information from the vast amount of noise in social media, monitor, analyse and present actionable information.'

3.3 There are of course commercial imperatives that are shaping the formation of the software and its rhetorical presentation. We shouldn't be put off by this. We are entering here into the intersections between 'commercial' and academic sociology to which Burrows and Gane (2006) have referred. If we can see this as a form of 'commercial sociology' then it is possible to import these devices into academic sociology in order to explore new ways of seeing, this would be to reshape the purposes of the software to our ends. It is necessary of course to consider the divergent interests that inform commercial social research so as to explore how we might reshape the use of such software to suit the agenda of a more critical form of social research.

3.4 To return for the moment to a focus on what this software can do, this research device gathers data around selected search terms over time. The software provides a relatively longitudinal analysis of temporal variations in the data. In this instance the developers describe the coverage of data sources as follows:

'We cover a wide range of social and other online media sources. It includes blogs (including Blogger, Wordpress, Typad, and LiveJournal), microblogging sites (Twitter), social networks (public Facebook pages), video sharing sites (YouTube), discussion boards, review sites and online news sources (regional, national and international).'

3.5 This reveals a vast and inclusive coverage of many of the most popular and active cultural hubs. Once the chosen search term is entered then data is captured from these sources until the search term is removed. The commercial organisation stores the data and the software allows it to be accessed remotely for analysis (thus reducing the common barriers of storage and computing power). As the data accumulate the depth of insight created is enriched and trends and connections begin to emerge over time. These software devices find the specific data amongst the mass of general data and bring these together for analysis. The types of things that these devices make visible are the *geography* of the content containing the search term, the *association of words* with the chosen term, the *sentiment* of the references and the scale of the buzz over time, the *influential voices* who have made reference to the chosen term and, finally, they can be used in a number of ways to move *from aggregate quantitative analysis to locate and explore individual utterances*. In the following section we explore each of these core functions in turn.

Using social media data aggregators: seeing sociology in the mess of by-product data

4.1 Topics can be followed using these data aggregators for an indefinite period of time. Due to its likely general interest to the readership of this journal, we followed the search term 'sociology' from 10 May 2011 to the 1 November 2011. Here we use the data generated by this exercise as a case study that may provide some concrete illustrations from which we might explore the analytical capabilities of this social media data aggregator software. To give some sense of scale of the volume of data generated figure 1 compares volumes of content containing 'sociology' with that which contained 'celebrity' during the period 13 June to 9 July 2011.

Figure 1. Graph showing total Twitter posts comparing the topics Sociology and Celebrity

4.2 Despite the overwhelming presence of celebrity as compared to sociology, it is perhaps surprising just how much coverage sociology is actually getting, on the 14 June 2011 there were 3,006 posts about sociology on Twitter and there were only 9,031 using the term celebrity. The relatively high score for sociology suggests instantly just how embedded social media is in academic life. The norm though would seem to be somewhere around the region of those recorded on the 3 July 2011, where sociology received 707 mentions and celebrity 13,885. We will return to the software's analysis of sentiment in a moment, but what is also interesting is that the gap between sociology and celebrity narrows somewhat if we perform the same comparative exercise but we include only Twitter posts which are deemed to have a negative sentiment, that is posts judged by the algorithms to be saying something negative in association with sociology or celebrity (see figure 2).

Figure 2. Comparing the volume of posts on Sociology and celebrity that are judged to have a negative sentiment.

4.3 Unfortunately, here we see that where we focus on negative sentiment the difference between sociology and celebrity is narrowed a little. On the 20 June 2011 the variation between sociology and celebrity in terms of negative statements is limited to only 994 posts (1,450 for sociology and 2,444 for celebrity). On this day a relatively high number of people had something negative to say about sociology. We can return to this issue of sentiment and temporality in a moment, let us first focus upon the issue of scale a little further by understanding how social media content might be mapped to generate a geography of a topic.

Geography

4.4 The software can be used to generate a cumulative map showing the geography of the chosen search term. This enables the user to see the concentration of content on a global scale and to appreciate where it is that these discussions are occurring (figure 3 is a map of sociology extracted on the 13 July 2011). We might imagine that there are some highly nationalised topics, in this instance sociology appears to have some predictable areas of concentration in North America and the UK – exacerbated of course by the use of the English language term for sociology in the search process. Despite national variations the map of sociology in social media data shows a global spread with most populated regions containing at least some sociological content.

Figure 3. A map of the world showing the location of data generated about sociology as of 13 July 2011. The dots show the location of content creation on this topic.

4.5 It is a little hard though to get an exact sense of the national variations in terms of volume from this map alone, but the content can be further broken down into national categories in order to see where the most discussion is occurring in a bit more detail. Figure 4 shows the national 'geo coverage' of sociology.

Figure 4. Pie chart showing the breakdown of references to sociology by the national location of the person posting it as of 13 July 2011.

4.6 The live version of this chart is much easier to read as the pointer can be used to isolate the segments and to read the number of postings and the percentage of the overall content that this represents. Let me isolate a few notable instances that we can extract from the 'geo coverage' extracted on the 13 July 2011. The USA is the highest with 42.28% of the content, the UK is second highest with 33.26%, Canada is third with 7.52%, and fourth is Brazil with 2.86%. What this map and pie-chart also reveals is the genuinely global spread of sociology with small quantities of content captured in Peru, Belgium, Australia, India, Indonesia, Burkina Faso, and Singapore. There is a prominent concentration in the places we would expect but then a large number of places with a relatively small amount of content. So despite the limiting factor of language variations in our search – with the Anglophone term sociology used in this project – we still find a genuinely global discussion occurring about sociology in even some quite surprising places (it also of course tells us something of the global spread of social media and of the way in which social media are interfacing with academic subjects). More generally this software generates a geography of interaction for specified topics producing, as it were, a global perspective.

Word associations

4.7 We can quite easily get a sense from this of the scale of the data and its geographical spread. This tells us little though about the exact nature of the content. One of the mechanisms for seeing the detail of what is being said are what are described as 'word maps'. Figure 5 shows two word maps created for sociology, the left-hand word map was created on the 13 July 2011 and the right-hand word map was created on the 1 November 2011. This reveals the words that are most commonly used in association with the search term sociology over time, showing how such techniques might be used to track particular topics over time – in figure 5 we can compare the word maps to see that there are some words that are consistently associated with sociology and others that are associated with particular times of year (such as exams). The lines linking the words show where these associated words have commonly been used together. We can see that the words 'long', 'good' and 'year' have been used together in reference to sociology. So have 'time' and 'exam', 'give' and 'read', and 'people' and 'years'. These would suggest that a high number of those creating the content are likely to be discussing their degree courses, this seems to be particularly true in the 13 July instance. Perhaps the most surprising word contained in the left-hand word map is 'green'. Following this particularly surprising presence is helpful in seeing how we can access the actual content to assess the presence of this trend. Selecting the key word 'green' generates a list of posts that use sociology in conjunction with green. Amongst the predictable references to the associations between sociology and environmentalism, we also discover that at least two people studying sociology enjoy drinking green tea and at least three were planning to see the newly released film *The Green Lantern* at the cinema (the relatively short time span of the project has made the visuals a little more volatile and susceptible to what might be momentary increases in the volume of particular types of content). This shows how these word maps might be used to reveal particular trends in content and then to explore their origins in the content itself, it is also suggestive of ways in which this type of software might be used to analyse cultural tastes and movements.

Figure 5. Two word maps showing the words most commonly associated with sociology. The left-hand word map was produced on the 13 July 2011 the right-hand word map was produced on the 1st of November. The lines between words indicate the most commonly connected associations.

4.8 What should be noted here is that the density of information in these word maps improves over time as data accumulate. The greater the quantity of data the more detailed the word map. The connections between commonly associated words also show more clearly as the volume of data increases. Sociology's word map gives a sense of the content of the posts, but the relatively limited data for this search term restricts the nuance of the map. Longer scale projects or more commonly used terms generate much more detail.

Buzz and Sentiment

4.9 The notion of measuring sentiment at an aggregate level will be something of a controversial issue for social scientists. In this case the software automatically analyses the content of the data and categorises it as positive, negative or neutral – this is obviously a quite crude categorisation of sentiment but the nuance is only likely to be integrated into such systems over time. The software uses other words in the posting to make these judgements over sentiment, the algorithms decide what combination of contextual discourse indicates what type of sentiment. The creators of the software describe this process as follows:

'Our automatic sentiment analysis algorithms examine the grammatical structure of sentences to measure the tone of the sentiment related to a chosen topic. We use leading-edge machine learning and natural language processing algorithms to measure sentiment related to your brands. You can view raw scores (positive, negative and neutral) as well as our qualitative judgement on the tone of the sentiment. If our customers do not agree with our judgment, they can change it by clicking a button and our training algorithms will automatically learn on individual preference.'

4.10 Social scientists have a number of issues to consider here, not least of which would be to accept that sentiment can be judged in this way in the first instance. That is to say that we might actually be able to look at sentiment on an aggregate level and that we can at least partially trust an automated algorithmic set of judgments to differentiate the underlying sentiment of particular utterances. Clearly this is going to be an inexact art that will find things like irony and insider discourse difficult to cope with. The use of these three categories is also quite cumbersome, there are of course wide ranging sentiments that might not fit easily into such a framework, the sociology of emotions and affect would tell us as much.

4.11 Another issue here is the removal of human agency and judgment from the initial research process, here we as social researchers are inviting the power of 'software sorting' (Graham, 2005) further into research processes. We are enabling algorithms to do some of the analysis on our behalf, as such we leave ourselves, our practices and ultimately our findings open to the 'power' of algorithms (Lash, 2007; Beer, 2009). We may have issues with this. There are though some important benefits to consider. We are dealing here with what can easily become an unfathomable quantity of data within which these algorithms allow us to find patterns, they can therefore help us with the very problem that Abbott outlined at the opening of this article. This software enables us to see things in the data that we might not otherwise be able to see. Perhaps we should proceed with caution and take a quite critical approach to what it is that we might obtain from this algorithmic analysis. This aggregate level analysis of sentiment is a starting point, something that reveals patterns that need to be explored.

4.12 If we return to our focus upon sociology, it is possible to observe how the sentiment of the content varies over time and across different types of media platforms. These graphs can be created for different media and different time periods, the figures below display the sentiment of content created during one

month between the 13 June 2011 and the 13 July 2011 across Twitter (Figure 6), Blogs (Figure 7) and Facebook (Figure 8). One thing that is obviously visible from looking across these three visualisations is that there was a good deal more content created about sociology in late June than in July. The most obvious explanation is that this was in response to university assessments. The live versions of these graphs enable the user to select particular dates and to look at the actual postings for that date. Selecting the peaks in content from late June reveals a high volume of students communicating concerns about revision, exams and the like and then sharing reflections on how the exams had gone. This pattern is most pronounced on Facebook where the content almost tailed off completely after June, suggesting that the sociology content on that particular application is predominantly generated by students. The Facebook peaks occur across all sentiments reflecting the variety of responses to the exams, selecting the peak in negative comments from the 19th of June reveals posts about how 'hard' sociology is, about how 'boring' sociology revision is and about the fear of failure in these sociology exams. The Twitter and blog content has shown a much less marked reduction in content during the period. The blogs have maintained a more consistent quantity of sociological content, particularly of a positive nature, which suggest that this might well be the place in which professional sociologists are creating content or where their findings are being disseminated. Selecting the 6 June 2011 for blog sentiment (figure 7) supports this notion with posts detailing job vacancies and others reporting on research findings by those with a general interest in sociological issues. Sociology then is being presented in a particular way across these different applications. Perhaps the main general observation we can note here is the way that the quantity of sociology related content posted often follows the same pattern across positive, negative and neutral sentiment^[2].

Figure 6. Sentiment on Twitter 13th June to 13th July 2011

Figure 7. Sentiment on Blogs 13th June to 13th July 2011

Figure 8. Sentiment on public areas of Facebook 13th June to 13th July 2011.

4.13 I have only really touched the surface of what it is possible to extract in terms of sentiment on the topic of sociology. It is only possible here to give a glimpse, but the topic chosen can be analysed in some depth to understand trends and to then open up the content in order to explore why peaks and troughs are occurring or to understand the responses across different media. It is possible to trace topics over time to see variations in response (and potentially how these are shaped by broader social events, news stories and the like). Let me leave the possibilities outlined as suggestive and move on to reflect on how we might differentiate the content further to understand which are the most important voices behind some of the observable trends.

Influence

4.14 If sentiment reveals something about a general sense of feeling towards a topic, influence allows us to see where we might locate the most powerful actors that shape these senses of feeling. There are of course many ways to understand influence and how it operates. One of the advantages of the type of

digital traces and by-product data I have described here is that it makes it possible to measure influence in a variety of ways. Given the centrality of influence in social science we might instantly see some value in being able to measure it. In terms of prominent social scientific themes around social networks (Bottero and Crossley, 2011) and Bourdieusian interests in social and cultural capital (Bennett et al, 2009), we can imagine that tracking and measuring influence might be a useful tool. To take these two popular themes in sociology, it becomes possible to imagine using these aggregator software packages to find the powerful actors in networks, or to attempt to unpick the significant actors in the formation of cultural capital. So what is this vision of influence, how is it measured and what might we begin to see with it?

4.15 According to the software developers:

'The software can identify influential bloggers and twitter accounts for a chosen topic based on a number of measurements. It displays all measurements so that you will see how they derive their authority. To fine-tune the process, you will be able to adjust the weights.'

4.16 The software measures 'influence' in a variety of ways and these measures can be weighted to create either a list of the most influential bloggers or Twitter profiles of those who have posted on that chosen topic. The measures vary from straightforward measures such as levels of activity and number of comments posted in response, to things like number of mentions by other people and the number of followers, and then on to more complex measures of Klout, amplification, and net influence. It is possible to create a list of those with most influence using one of these measures or a combination.

4.17 Rather than discuss all of the possibilities in detail let me focus on Klout as a particular example and because this is calculated using a freely available web application so is easily accessible to the reader (available at klout.com). The Klout score, as with the other measures of influence can be used alone or can be combined with other measures to calculate the level of influence that an individual has. The Klout score uses 'over 35 variables' to measure three things: 'true reach', 'amplification probability' and 'network score'. Klout describe these in the following terms:

'True Reach is the size of your engaged audience and is based on those of your followers and friends who actively listen and react to your messages. Amplification Score is the likelihood that your messages will generate actions (retweets, @messages, likes and comments) and is on a scale of 1 to 100. Network score indicates how influential your engaged audience is and is also on a scale from 1 to 100.'

4.18 Here we can see that even this one form of measurement is actually quite sophisticated. Combined with the other measures of influence available through the aggregator software we can get a sense not just of how active individuals are or how much content they generate, we can also gain measures of how engaged the audience of an individual is, of how much activity they stimulate in other people, and in how influential this audience of followers are in their own right. This last factor is important because you may only have a small number of followers but these followers may themselves all be influential.

4.19 Instantly we can see here the possibilities for the ways in which this might be used to understand how cultural capital and 'subcultural capital' (Thornton, 1995) are shaped within a hierarchical frame of influence. This could create new insights into the way in which cultural divisions work in practice and how powerful voices influence the positioning and parameters of such divisions. The data captures influence and the software allows influence to be measured, this is a prime resource for further understanding how cultural and social capital are formed and shaped by influential actors within networks – at this point the reader may also be imagining how this type of measurement of influence might become part of the evaluation of the 'impact' of academics (academic Klout scores and the like could be part of the metric evaluation of our worth!). Clearly though a much more detailed engagement is needed with the technical details of the calculations of influence and how exactly they might be used, but to give some illustrative examples from this study of sociology we can find out what some of the most influential voices are saying in reference to our discipline.

4.20 Creating a list of influential Twitter profiles that have referred to sociology during May to June 2011, we find at the top a Chicago based branding strategist who reveals that she is actually having an interaction with a sociologist about online branding. The second, based in New Orleans, makes a distinction between sociology and biology, based upon parentage or environment as determinant. The fourth on the list, an actor and civil rights activist based in Hollywood, says that it is inevitable that they watch the reality TV show Big Brother as they majored in sociology. Clearly we are returned here to the mess and chaos in such data, but there is value in exploring the comments by influential actors and how these present sociology to a large and 'influential' audience – the link between being a sociologist and feeling compelled to watch reality television in this last post is notable. These are the comments, to use the jargon of the software, that are most likely to get 'amplified' and thus shape perceptions of our discipline^[3].

From the aggregate level to qualitative utterances

4.21 Before concluding let me just briefly reflect upon one of the crucial functions of this type of aggregator software for conducting social research. This relates to the ease with which the researcher can switch from quantitative aggregate level analyses of the data to small scale qualitative levels of analysis of particular utterances. This makes it possible to observe trends and then to access the data that constitutes that trend in order to generate a more nuanced understanding of the headline observations. The meta level analysis allows the content to be organised so that more small scale forms of analysis can be targeted and specific. I have already demonstrated some of the ways in which this might be used, the checking of the content behind the word map or the analysis of the peaks in content volume and sentiment are two examples of this. There are a number of ways in which the aggregate data can be mined down to the individual level. It is also possible, once the content has been selected, to open up a profile that provides some specific details about the individual who posted it. These profiles contain information about names, locations, the various influence scores that the profile has accumulated and even a photo and some description of the profiler and what they use their social media for (if available). We have three levels of analysis here, aggregate level, content level, and individual level.

4.22 To illustrate this, let me very briefly summarise the pathways that allow the researcher to move between trends and specific content. First, it is possible to create samples of content. Lists of posts can

be created through filtering criteria including the date of the posting, the sentiment (positive or negative), by most recent or on the sentiment score, and by media type. Perhaps more useful though are the links from the visualisations into the samples of content. Clicking through by selecting particular dates from the graphs (see figures 6, 7 and 8 for example) is particularly useful in understanding the variations in types or volume of content in more detail. The option to click through to particular content from the word map is also helpful in understanding the clustering of discourse and the exact context in which these commonly used terms arise. Finally, the option to explore the content of the most influential voices means that the researcher is able to see exactly what these most influential voices are saying about any chosen topic and to explore their profiles in more detail (the analysis can also be set so as to include related terms or exclude those that are not desired).

4.23 Perhaps one of the most significant aspects of this software is that it highlights apparent trends and then makes it possible for the social researcher to explore these in detail by trawling through the content and the individuals creating it. It is very easy to move here from quantitative to qualitative modes of analysis, which, of course, is not always easy to do. Here we have statistics and individual utterances that can be easily worked together.

Conclusion

5.1 It is hoped that this article is only the start of something and that sociologists will engage with this type of social media data aggregator to do social research. There are limitations but there is also potential for us to productively expand our analytical repertoire. In this software we have the possibility of finding patterns in the monumentally detailed data to which Abbott (2000) referred over a decade ago. The mass of accumulating by-product data can be thought of as a new form of social data, this software allows us to do something with it, to find things out and to generate insights. It pulls out content relating to our topic and allows this to be analysed in macro and microscopic detail. This article begins to show a little of what this software can do now, it provides only a snapshot of one example amongst a set of developing analytical technologies (as I revise this article it is seen that new developments in sampling have been added to the software that allow for more comparative and multi-dimensional forms of analysis). It would be a shame for social scientists to overlook such developments and to miss this opportunity for refreshing what it is that we do and how we do it. We obviously should not abandon our established methods and run head long into the promises of these new analytical devices, but there is some obvious potential here that suggests a pressing need to look at how we might 'enact' the world in different ways (Law, 2004), to see it differently through by-product data. This is an opportunity to do just that, it is an opportunity, as Savage has put it, 'to broaden our repertoire and recognize the changing stakes involved in the circuits of 'knowing capitalism'' (Savage, 2000: 249). On top of this concern, and on a more positive note, the data and the analysis we can perform are really interesting and revealing, these types of social media aggregators can be used to find things out about virtually any topic that social researchers might be interested in. Social media are now so embedded in the routines of everyday life that much of the detail of those lives are being communicated through them, this means that social and cultural issues of virtually all types are in abundance in this data. This software makes it possible to see this at an aggregate level and to use this to explore the individual details of the content and those making it.

5.2 As I have said, we are at a moment in which the function of the software I have described is developing rapidly – since this project was completed new functionality has been added to the software that allows the data to be re-organised and analysed in various types of subgroups. Despite this nascent moment in the lineage of its development it is already possible to begin to see how social researchers might be able to use such software. We could use this software to research individual people, places, events, topics, objects, genres, tastes, social issues, and the like so as to reveal their geography, their common discourse, the sentiment of the content, the influential voices and to explore the aggregate level and small-scale utterances on topics, amongst other things that are already possible or which will soon be possible. These types of data and software could actually be used to analyse many of the topics and issues that are central to contemporary sociology and social science more broadly. Over time the density of the data will only escalate as will the analytical power of these new software 'inscription devices' (Osborne et al, 2008). There has been a wide-ranging discussion in recent years of sociological methods outside of the academy and how powerful these might be in terms of doing social research (Savage & Burrows, 2007; Adkins & Lury, 2009 & 2011; Webber, 2008 amongst others). Here we have an opportunity to actually import one of these commercial inscription devices and to see what academic social researchers might be able to do with it. Where we feel the analytics of the software are too prescriptive we always have the option of extracting the accumulated data sets and employing alternative analytical techniques.

5.3 There are of course reasons to show at least a little caution. The practices I have outlined here depict a future form of social research that has become a part of what Kitchin and Dodge (2001) call 'code/space'. This is to technologize our research practices in a way that was not previously possible, in so doing it moves some more of the analytical processes of social research into the hands of machines. We should have concerns about fostering what might be thought of as a *posthumanist social science* and we should of course find ways to retain our own critical and analytical faculties. If we are to venture further we need to make a priority of looking under the bonnet of these devices, particular as there is a necessity for us to understand how the data is being sorted for us. That is to say that the *a priori analytics* occurring in this software need to be fully appreciated where possible. This will take further consultation with software developers and some ongoing engagement with these developers in order for their functions to be shaped to our research agendas and approaches. We are faced with the presence of *the power of algorithms in our method*, these algorithms will come to shape the outcomes of social research methods as they sort-out our data and present it to us in a manageable form (Beer, 2009). The power of algorithms in this instance is to enable us to find patterns in unfathomable data. As social media sink further into the everyday, so this data will escalate even further beyond our analytical comprehension. The decision this leaves us with is whether to accept such software into our practices, to further bring sociology into 'code/space', or to not use the data in any sustained research programme. We might be able to build some small scale case studies without the software, but the quantitative insights we might generate through this type of software would be beyond us – and so then would the opportunities for qualitative exploration that it makes possible. I would suggest that we open-up sociology to such possibilities, the challenge then will be to find ways of doing *critical* social science with such inscription devices – devices that are not initially designed for such a purpose. We have an opportunity here for developing a new strand of social science, a social science that begins to take advantage of the masses of social data accumulating around us, a social science that speaks directly to the changing social context and which is

keen to expand its methodological repertoire (and to interface and mix old and new methods). This article has outlined one example in a rapidly changing field, but we are left now to ask how we might draw upon the commercial sector to find solutions to the digital problem. The types of solutions offered by the digital entrepreneurs often address similar problems around data inundation to those experienced by social scientists. Perhaps the way forward is to work closely with some of these software developers so that we understand the software and its automated sorting processes, and so that we might use such a dialogue in order to morph versions of this software to suit our own critical and comparative interests.

Notes

¹ There are a number of pieces of software on the market that are designed to enable companies to measure public responses to chosen topics (such as brands and the like). The choice of the piece of software described here is a somewhat arbitrary in that it was motivated largely by the geography of the project and the relative ease with which the researcher could meet face-to-face with the software developers. It is hoped that the general functionality of this software is representative of the types of developments that are occurring and that by discussing a specific example the reader will get a general sense of these developments.

² The thing that sociologists might want to reflect upon is that in most instances sociology is spoken of in neutral terms. We might also be pleased to hear that where a strong sentiment is to be found it is usually positive. If sociology were a brand, which in some ways it is, and if we were commercially orientated, which is some ways we are, then we might be concerned that our brand receives such an overwhelmingly neutral response. Brands are not usually intended to be neutral. This treatment of a discipline as a brand does of course raise some issues in its own right. As one of the reviewers of this article has asked me, what are the political and social implications of treating a discipline as a brand? I'm not sure that I can answer that question, indeed the reviewer generously did not expect an answer, but it seems to me that we have already been witnessing the branding of disciplines for some time, particularly as they compete for funding, fees and public attention.

³ It is important to be aware that as things stand the measurement of influence is not based specifically upon the influence of the individual in the field of the chosen topic. Currently the software provides a measurement of the things that those that have general influence say about a chosen topic. Discussions with the software designers reveal that they are also developing ways of measuring who is influential in the area of the specific chosen topic. In this instance this new measure would reveal who is influential in sociology rather than what it is that influential people who mention sociology are saying about it. This would be an important development. To give an example, it might be that a student blogger who is influential in the gaming scene might say something about a sociology class. This is important because this voice will be heard and will shape perceptions. However, this doesn't mean that this voice is influential in discussions about sociology. As I understand it the next step is to make this differentiation so that we might then see how influence varies across topics.

References

- ABBOTT, A. (2000) 'Reflections on the Future of Sociology', *Contemporary Sociology*, vol. 29, no. 2, pp. 296-300. [doi:10.2307/2654383]
- ADKINS, L. and Lury, C. (2009) 'What is the Empirical?', *European Journal of Social Theory*, vol. 12, no. 1, pp. 5-20. [doi:10.1177/1368431008099641]
- ADKINS, L. and Lury, C. (eds) (2011) *Measure and Value*, Sociological Review Monograph, Wiley-Blackwell, Oxford.
- BEER, D. (2009) 'Power through the algorithm? Participatory web cultures and the *technological unconscious*', *New Media & Society*, vol. 11, no. 6, pp. 985-1002. [doi:10.1177/1461444809336551]
- BEER, D. & Burrows, R. (2007) 'Sociology and, of and in Web 2.0: some initial considerations', *Sociological Research Online*, vol. 12, no. 5, <<http://www.socresonline.org.uk/12/5/17.html>>. [doi:10.5153/sro.1560]
- BAUMAN, Z. (2007) *Consuming Life*, Cambridge, Polity.
- BENNETT, T., Savage, M., Silva, E., Warde, A., Gayo-Cal, M. & Wright, D. (2009) *Culture, Class, Distinction*, London, Routledge.
- BOTTERO, W. & Crossley, N. (2011) 'Worlds, Fields and Networks: Becker, Bourdieu and the structures of social relations', *Cultural Sociology*, vol. 5, no. 1, pp. 99-119. [doi:10.1177/1749975510389726]
- BURROWS, R. and Gane, N. (2006) 'Geodemographics, Software and Class', *Sociology*, vol. 40, no. 5, pp. 793-812. [doi:10.1177/0038038506067507]
- FEATHERSTONE, M. (2006) 'Archive', *Theory, Culture & Society*, vol. 23, no. 2-3, pp. 591-596. [doi:10.1177/0263276406023002106]
- GRAHAM, S. (2005) 'Software-sorted geographies', *Progress in Human Geography*, vol. 29, no. 5, pp. 1-19. [doi:10.1191/0309132505ph568oa]
- HAYLES, K.D. (1999) *How we became posthuman: Virtual bodies in cybernetics, literature and informatics*, Chicago University Press, Chicago.
- KITCHIN, R and Dodge, M. (2011) *Code/Space: Software and everyday life*, MIT press, Cambridge, Massachusetts.
- LASH, S. (2007) 'Power after Hegemony: Cultural Studies in Mutation', *Theory, Culture & Society*, vol. 24,

no. 3, pp. 55-78. [doi:10.1177/0263276407075956]

LAW, J. (2004) *After Method: mess in social science research*, Routledge, London.

LAW, J. (2009) 'Seeing like a survey', *Cultural Sociology*, vol. 3, no. 2, pp. 239-256. [doi:10.1177/1749975509105533]

OSBORNE, T., Rose, N. and Savage, M. (2008) 'Reinscribing British sociology: some critical reflections', *The Sociological Review*, vol. 56, no. 4, pp. 519-534. [doi:10.1111/j.1467-954X.2008.00803.x]

RITZER, G. & Jurgenson, N. (2010) 'Production, consumption, presumption: The nature of capitalism in the age of the digital 'prosumer'', *Journal of Consumer Culture*, vol. 10, no. 1, pp. 13-36. [doi:10.1177/1469540509354673]

SAVAGE, M. (2010) *Identities and Social Change in Britain since 1940: The Politics of Method*, Oxford University Press, Oxford. [doi:10.1093/acprof:oso/9780199587650.001.0001]

SAVAGE, M. and Burrows, R. (2007) 'The Coming Crisis of Empirical Sociology', *Sociology*, vol. 41, no. 6, pp. 885-899. [doi:10.1177/0038038507080443]

THORNTON, S. (1995) *Club Cultures: Music, Media and Subcultural Capital*, Polity, Cambridge.

WEBBER, R. (2009) 'Response to 'The Coming Crisis of Empirical Sociology': An Outline of the Research Potential of administrative and transactional data', *Sociology*, vol. 43, no. 1, pp. 169-178. [doi:10.1177/0038038508099104]