Participations
Journal of Audience & Reception Studies

Volume 12, Issue 1
May 2015

Audience research and social media data: Opportunities and challenges

Rob Procter
University of Warwick, UK

Alex Voss and Ilia Lvov
University of St Andrews, UK

Abstract:
The widespread adoption of social media platforms and other information and communication technology innovations not only pose new challenges for audience researchers but also provide exciting opportunities to re-invigorate audience research as an academic topic as well as a practical pursuit. In this paper, we outline some general methodological issues that arise when seeking to exploit these opportunities, drawing on our experiences of using Twitter as a resource for measuring audience engagement with the BBC World Service (BBCWS) in the context of global media events, specifically the London Olympic and Paralympic Games in 2012 and the Sochi Winter Olympic Games in 2014. We conclude by arguing that social media are not simply a new source of data about audiences but a new forum for unprecedented interaction and collaboration with the audience and, in this respect, they are phenomena to be studied in their own right.

Keywords: audience research methods, social media, big social data, Twitter

Introduction. Social media and audience research

Numerous commentators have observed how the widespread adoption of first Web 2.0 and, latterly, social media platforms, have challenged established patterns and roles in media production and consumption and, in particular, the meaning of 'audience'. Gillmor (2004), for example, talked about the 'former audience' and Rosen (2006) referred to bloggers as the 'people formerly known as the audience' - both writing in the middle of the first decade of this century. Dutton (2009) and Newman, Dutton and Blank (2012) have

argued that social media platforms have contributed to the emergence of the 'Fifth Estate', thereby allowing audiences to share their views with worldwide publics and to engage directly with one another in this new public space and with international news media organisations.[1] News media organisations today are encountering audiences in ways that are quite different from the past. No longer are audiences the more or less passive recipients of news disseminated through a one-way channel but, thanks to modern information and communication technologies and, in particular, the global adoption of social media, they are connected to the newsrooms of the world and can interact with each other and also may take a more active role in setting news agendas.

This view raises interesting questions as to whether and how mainstream[2] media will succeed in reasserting their control over what counts as 'news' (O'Loughlin and Hoskins 2015). In part, this special issue documents how the relationship between media and audiences is changing and how mainstream media are responding to the challenges which this changing relationship represents to the established order.

Some audience members have become 'produsers' (Bruns 2008), that is consumers and active producers of content at the same time. This is reminiscent of the idea of 'prosumers' (Toffler 1980; Ritzer and Jurgenson 2010) who are consumers who critically engage and tinker with technical products, something taken to new levels by the recently emerged 'maker' movement[3] (Doctorow 2009). For news media organisations, these concepts suggest a potentially dramatic change in practices, where through the affordances (Gibson 1997) of modern technology, the actions of people formerly known as the audience (or the consumers) challenge the monopoly of news production, whether deliberately or otherwise.

Many journalists have also begun to embrace these developments and to develop stronger, multi-nodal relationships between media organisations and the 'former audience'. Increasingly, journalists, their editors and producers draw on the audience's enhanced capacity to 'talk back' through social media (Mackay and Tong 2011) for reactions to events or to programmes in order to inform their work. Mark Thompson, then Director General of the BBC, described the changed relationships between news media organisations and their 'active audiences' in these terms:

> They won't simply be audiences anymore, but also participants and partners – we need to know them as individuals and communities, let them configure our services in ways that work for them… this second digital revolution is going to enable the public to explore and investigate their world like never before. Programmes won't be shown once and then forgotten. They'll be there forever to be linked, clipped, rediscovered, built into bigger ideas… It's the active audience, the audience who doesn't want to just sit there but to take part, debate, create, communicate, share (Thompson 2006).

Social media platforms and other modern information and communication technologies are therefore relevant for audience researchers not only as a possible source of information about audiences but as a crucial enabling platform for the new, multi-nodal relationships between news media organisations and the rest of the world. As Hoskins (2013) points out, the traditional distinction between mainstream media and audience has been transformed. Now, the representation of events emerges from the *interaction* of traditional and emergent media rather than from each of them independently.

We would argue that uses of social media and the interaction with mainstream media are *phenomena to be studied in their own right.* For example, reports from the Reuters Institute on digital new consumption (Newman 2011 2014) highlight the rise in importance of social media platforms for news consumption. The 2014 report shows that 35% of survey respondents claim to have used Facebook for news consumption in the week prior to the survey. Twitter is less widely used for news consumption overall with 9% of respondents claiming use. However, it serves a more important role in relation to the consumption of breaking news and is also very actively used by journalists and included in the social media strategies of news media organisations (ibid.). This not only poses new challenges for audience researchers but also provides *exciting new opportunities to re-invigorate audience research* as an academic topic as well as a practical pursuit.

In the following sections, we provide a brief summary of audience research methods pioneered by the BBC and used traditionally in broadcast media and also discuss the performative role they play in shaping the relationships between media organisations and their audiences. We then argue that the rise of social media not only challenges these relationships but also, by virtue of being available at scale and in real time, social media present new challenges for audience research methods. We discuss in some detail the nature of data available from one particular source, the Twitter micro-blogging service, and the implications for its potential uses for audience research. Referring back to the requirements for audience research, we outline the methods used in the papers in this special issue. We conclude by critically reflecting on the methodological challenges posed by social media data and the opportunities they present to audience researchers.

Purposes, requirements and methods of audience research

In this section, we discuss the traditional purposes, information requirements and data collection methods of audience research using the example of work done at the BBCWS, which is well documented and which has been a focus of our ongoing research activities. Insofar as there is overlap between the interests of the BBC as a global broadcaster and other media organisations, the discussion is not BBC-specific but will lead to conclusions that are more generally applicable. The audience research activities at the BBC, in general, have had a significant impact on the development of audience research activities elsewhere (Gillespie, Mackay and Webb 2011; Mytton 2011), so we expect the purposes, requirements and methods to be similar in other organisations. This raises the question of our relationship

with the BBC, which Gillespie et al. (2011) describe thus: 'we were not engaged in research *for* the WS, but *with* WS researchers about WS audiences' (p.5). We have maintained throughout our research an intellectual independence from the day-to-day practices of audience research in this organisation and this was appreciated by our partners in the BBC.

Our aim is to identify how social media might be used as a source of data, given the purposes that audience research serves and the established methods used. Hence, firstly, we outline these before moving on to questions about big social data and social media specifically in the further sections.

Audience research has traditionally served a number of *purposes* within the BBC WS and in its relationship with its funders (Mytton 2011; Webb 2011; and Wilding 2011). Its primary purpose has been *accountability* (Wilding 2011; Gillespie, Mackay and Webb 2011) in order to justify funding of the service provided by the Foreign Office through Parliamentary Grant-in-Aid (that is, the BBCWS was funded by the Foreign Office but was formally independent, cf. Webb 2011). In April 2014, the BBCWS came under license fee funding, a move that has put increased pressure on it, leading to significant restructuring and closure of some language services (Wilding 2011; Gillespie, Mackay and Webb 2011). This change in funding is accompanied by financial pressures on the wider BBC, with cuts of a total £700 million projected by 2017. As an Editorial in the Observer newspaper remarks, 'the political strain of constant BBC cuts must hurt the services the UK licence-fee payer doesn't see or hear more than the ones they do' (Observer 2014).

This means that while the importance of accountability is undiminished, the specific requirements may change. While it seems unlikely that the BBCWS as a whole will disappear anytime soon, the same cannot be said for all of its component parts. Language services have in the past been closed while others have been expanded (Wilding 2011; Gillespie, Mackay and Webb 2011) in response to political developments but also in the light of their perceived performance. Furthermore, the BBCWS is at risk of becoming accountable for transmitting the values of the British TV-license payers, rather than for appreciating the cultural differences and sensitivities of its broad international audience (Webb 2014). Finally, as Gillespie, Mackay and Webb (2011:13) point out, it is facing challenges from a much larger range of competitors now that global reach is available, not just to large organisations such as its state-sponsored competitors but also for much smaller organisations and, crucially, news organisations that have traditionally worked in the print media market rather than broadcast.

Advocacy is a related, broader purpose as the service needs to 'remind its stakeholders and wider publics of its value and of its continuing success' (Wilding 2011:179). The specific aims are not just to justify expenditure but to foster the relationships that keep the service alive, not least with current and potential audiences. As Gillespie, Mackay and Webb point out, in the past, 'the [BBCWS] could depend on the loyalty of audiences across generations' (2011:7). However, as nowadays the 'dynamics of how audiences consume (and, now, even produce) media are changing' (Napoli 2012:79; also Newman 2011; Newman and Levy 2014), this cannot be taken for granted in the future.

In addition to the above two purposes, audience research is also used in the formation of *strategy* (Mytton 2011) which defines, for example, what programmes to run, on what platforms and for which audience groups. Media organisations routinely use sets of key performance indicators to measure the performance of particular channels and outputs and to make decisions about funding and editorial strategy on this basis.

As social media data can potentially provide rapid feedback from audiences, arguably a *tactical purpose* might be added to the above list. The changes in audience relations described in the introduction mean that it is possible to draw on audience data to influence the *content* of individual programmes over short timescales - something that would not have been possible using the methods used for audience research in the past (see below) and which is more akin to the optimisation of web content on the basis of web metrics from clickstream information. For example, in the case of the BBCWS such metrics include the number of unique users (UU) to visit their website and page impressions (PIs), the number of times a specific webpage has been viewed (Gillespie, Mackay and Webb 2011; Mackay and Tong 2011).

Given the above purposes, what specific information do audience researchers seek to generate and what insights are they aiming to provide? In 1947, Ian Jacob (cited in Webb 2011:158) defined the following *requirements* for the emerging audience research group at the BBC's External Service (the predecessor of today's BBCWS):

a) Knowledge of what is going on in the country concerned, that is the political, social and cultural developments.
b) Knowledge of the distribution of receiving sets, their nature and type, as well as of listening habits of the different sections of people.
c) Assessment of size, type and distribution of their audience and of their programme preferences.
d) Knowledge of what is being read or heard by their listeners from other sources.
e) Collection of systematic and widespread reaction to their programmes.
f) Study of and reply to mail received from listeners.
g) Advice, based on all the information gathered, to those directing output and those directing publicity.

Another list from the 1930s, compiled by Robert Silvey (reproduced in Mytton 2011), calls for the following key measures:

h) The quantity of listening to each broadcast.
i) The degree of popularity of individual programmes among those who hear them as well as the causes of such popularity or lack of it.
j) The reaction of the public to BBC policies and changes of policy.
k) The state of public opinion on matters with which the BBC was concerned.

These requirements are still central to audience research practice although some of the questions may need to be asked in a slightly different way. For example, instead of counting radio sets, we would today inquire about the use of the BBC's iPlayer versus analogue radio and versus digital radio as well as the uses of different social media platforms by different target demographics. More fundamentally, though, an additional set of questions exists today:

l) Increasingly, [the BBCWS] is looking at measures beyond basic reach. We need to know whether significant numbers of people engage with our output, value it, use it, respond to it; whether people come to us by choice or by chance; whether they trust us, care about us, stick with us, recommend us. Within the wider BBC there is a move towards greater emphasis on the *quality* of the relationship with the audience (Wilding 2011:183, our emphasis).

These latter questions point to a more complex and sophisticated understanding of the media-audience relationship that goes beyond the simple broadcast model that can be seen in the list of requirements from the 1930s and 1940s.

Gillespie, Mackay and Webb (2011) describe how audience research has, for a long time, had a role to play within the BBCWS in shaping relations between audiences, the broadcaster, its management and its sponsor (the UK Foreign and Commonwealth Office). The main ***methods*** used in audience research in the BBCWS have themselves shifted over time in response to different demands and circumstances.

However, as a result of a focus on accountability as the main driving purpose of audience research at the BBCWS, the main output is the Global Audience Estimate (GAE), a careful estimate of the reach of the service expressed in units of millions people reached on a weekly basis (Wilding 2011). This measure for the global reach is the outcome of a careful process of combining sample survey data from a large number of countries (ibid).

Other methods used in audience research include the study of letters sent to the service, panel surveys and interviews. However, as we observe below, the expansion of media services onto digital platforms provides new ways of measuring audiences, such as via website clickstreams and mentions in social media, as well as new challenges such as the problems of tracking individual audience members across multiple devices (Ofcom 2014). Clickstream analyses can yield interesting insights into what drives traffic to media organisations' websites but the insights that can be generated are often limited (Gillespie, Mackay and Webb 2011).

It is not surprising that the methods used for the specific purposes outlined above play more than simply a passive role; by definition they affect the relationship between media organisations and audiences (cf. purposes). The 'social life of methods' approach (Gillespie et al. 2011; Savage 2013) emphasises this performative aspect of the role that research methods play, their contested nature and the ways in which their selection as well

as their use are shaped by outside factors. For example, Mytton (2011), while describing his work in audience research at BBCWS in the 1980s, remembers the issue of survey sampling when researching audiences in India. The perception of these audiences was dependent on the sampling methodology, that is, whether to count the whole of India as one state or to conduct surveys on independent samples from its various states.

We would argue that the novel ways to study audience-media relationships presented in this special issue are logical extensions of audience research methods as being deeply interwoven with understanding news media, their audiences and the news media market. They reflect this by considering the *dynamic interaction* between audiences and broadcasters during global media events and they show how social media respond to, transmit, translate and, crucially, *influence* mainstream media content. The relative novelty of these developments and their dynamism required us to adopt reflective approaches that are sensitive to the nature of the phenomena we are studying and to the nature of the data with which we are working.

In the following sections, we will describe the kinds of data publicly available (for free) through services provided by social media platforms such as Twitter (sometimes referred to as 'big social data'). We outline the analytical approaches available for new kinds of research on 'active audiences' before reflecting critically on the nature of social media data and drawing some conclusions.

Audience research and social media data

Social media platforms often provide 'built-in' access to data available at the level of individual interactions with a single piece of text, image or video posted by a user (see below for an explanation of the kinds of data which Twitter provides). Such naturally occurring data at a population scale in near real-time allows researchers to shift back-and-forth between an individual-level view and an aggregate view of a target population. A consequence is that analytic approaches no longer fit neatly into the traditional categories of qualitative versus quantitative. They also transcend what have become known as 'mixed methods' where, for example, qualitative case studies might be used to add depth to quantitative statistical analyses (Tashakkori and Teddlie 2003; Mason 2006; Johnson et al. 2007) as the same data can be analysed both qualitatively and quantitatively (cf. Mackay and Tong 2011). Consequently, they pose a challenge for the development of research methods that can extract the value that lies in combining these two perspectives.

Not only can we innovate in the area of research methods, we can also begin to address new research *questions*. Social media data offers a complementary view not only on audience reactions to programme content but also on how these reactions form part of wider debates within society. That is, we can ask whether news media organisations are playing a role in a kind of 'global conversation' (Gillespie and O'Loughlin, this issue).

At the same time, increasingly complex news consumption patterns – partially a product of the uptake of social media platforms – raise new methodological challenges in measuring audience reactions:

> The complexity of news consumption means that the challenges of accurately measuring the audience are greater than in other, more straightforward areas of digital content. The systems that have been developed for these more straightforward content types are leading the way in online audience metrics – but the complexity of news consumption means that these metrics do not yet meet the needs of scholars interested in better understanding how news consumption takes place, how it flows in society, and the influences of news on opinion (Ofcom 2014:4).

Another problem is that social media data, like other 'big data', is often privately owned or controlled so its provenance is often not clearly identifiable. Savage and Burrows (Savage and Burrows 2007; Burrows and Savage 2014) have formulated this as 'The Coming Crisis of Empirical Sociology', resulting in a call for a renewed vigour in developing appropriate research methods to study social phenomena using 'born digital' data. The argument they make is not limited to sociology but we would argue extends to all of the social sciences including audience research. One of our aims in this paper is to formulate a response to this challenge for audience research by summarising the methodological innovations represented in the papers of this special issue.

According to Savage and Burrows (2007), the increased availability of data and a drive by private sector organisations to analyse and utilise it poses a challenge to academic social science. Similarly, we have seen in our own work that the availability of social media analysis tools such as SYSOMOS, Radian6 or Socialbakers poses challenges to audience researchers. These tools draw on commercially available big data streams to provide analyses of social media data, often from more than one social media platform. One may think that the availability of such tools provides news media organisations with an immediate way to study their audiences but they miss important aspects of the new world of digital journalism, which are the performativity of methods and the consequent need to engage with social media data in a iterative and reflexive way. The static arrangement of tools in these systems does not allow the kind of interchange between analytical perspectives that bring the quantitative and qualitative views into dialogue with each other.

These systems are also 'black-boxed', allowing little flexibility to conduct forms of analyses that are not already available in the system. More importantly for academic research, they provide little insight into how they analyse data, thus making verifiability and reproducibility difficult if not impossible (Halfpenny and Procter 2015). Their price tag means that even in large, well-funded organisations their uptake is often limited to a few individuals, which is sufficient perhaps for more traditional forms of audience research as a way to inform editorial and management decisions but not to drive open journalism. Our

work aims to provide complementary methods for analysing social media data that are *publicly* available using methods that are *open* and *can be inspected* rather than being proprietary and black-boxed.

Endpoint	Selection	Criteria	Limits
Filter	follow accounts	list of account identifiers	a maximum of 5000 account ids; maximum of 1% of all tweets
	track	keywords locations language	a maximum of 400 keywords; maximum of 1% of all tweets
	locations	longitude/latitude pairs language	25 0.1-360 degree location boxes; maximum of 1% of all tweets
Sample	sample of all tweets	N/A	1% of the total stream of tweets

Table 1: Overview of Twitter Streaming Services

Social media as a source of data

The studies presented in this special issue used data from Twitter's Streaming service, so we will focus on this source of data. Twitter, rapidly adopted since its inception in 2006, is a micro-blogging service that allows its users to post short (usually public) messages known as 'tweets' of a maximum length of 140 characters. Examples of Twitter-based research include mining public opinions (Thelwall et al. 2011), tension (Williams et al. 2013), patterns of communication (Bruns and Stieglitz 2012) and reactions to events (Procter et al. 2013). Out of all the different social media platforms, Twitter is of a particular interest for researchers as it provides them with arguably the most open access to its data (Housley et al. 2014) in that it provides a real-time stream of tweets, either as a 1% sample or as a dataset matching criteria that are specified by the user (see below). Other companies such as Google or Facebook (to name but two) do not provide similar access to their data.

The discussion of Twitter's service could serve as a general example of the principles of how users might access social media data but, unfortunately, there is a lack of standardisation or even commonality between different platforms. Each social media provider has their own terms and conditions so the extent to which data are made publicly available varies widely as do the formats this data comes in and the ways in which it can be obtained. We will therefore discuss the specifics of Twitter's service not as a generalizable example but as a specific case.

Twitter Streaming services

Twitter makes data available in near real-time through its set of Streaming services. Of interest in the context of the paper are the public streams that carry public data flowing through Twitter (Twitter Streaming API Documentation 2014). They comprise services that filter the stream of data by a set of criteria, as well as a service that provides a random sample of all data. See **Table 1** for an overview of the data streams and selection criteria that are available and limitations of these public services. The amount of data made available at no cost is currently limited to a maximum of 1% of all public tweets. This service is called the 'spritzer' and other service levels provide elevated access to the service: the 'garden hose' at 10% and the 'fire hose' at 100% of all public tweets. The latter services are available only to organisations with negotiated contracts with Twitter or through a number of official resellers (Gnip or Datasift), who provide ways to filter content from the fire hose not only in real-time but through querying an archive of historic data.


```
{
   "created_at": "Tue Sep 23 10:21:22 +0000 2014",
   "id_str": "514358858379440128",
   "text": "RT @guardian: Use our interactive map to see
     how climate change affects different areas
     disproportionately http://t.co/rPUBFxWj04 http://t.…",
   "user":  {...},
   "geo": null,
   "coordinates": null,
   "place": null,
   "retweeted_status":  { ... },
   "retweet_count": 187,
   "favorite_count": 0,
   "entities":  { ... },
   "lang": "en"
```


Figure 1: Data for a single tweet (abridged)

The data from these Streaming services can be accessed using tools such as the COSMOS desktop tool (www.cosmosproject.net), which allows the user to specify a data collection and stores the resulting data for subsequent analysis, providing some aggregate statistics in real-time while the collection is running. This allows the user to ensure that the data they are collecting are indeed what they intended to collect. This is especially important when using the filtering service with sets of keywords as incoming data may suggest additional keywords that should be used or may indicate that there are false positives which are tweets that a keyword matches but that are not relevant to the study.

Rate limiting explained

The public streams are by default limited to a maximum of 1% of all tweets. This is trivial in the case of the sample endpoint, which is 1% of all tweets by definition. In the case of the filter endpoint, the criteria often select more than 1% of the data stream. What happens in this case is that Twitter deletes tweets from the data stream to ensure that the overall

volume remains capped to 1%. Twitter informs the user that this has taken place by sending control messages that provide the total number of tweets deleted since the data stream was initiated. In our experience, these messages are frequent enough to provide an accurate estimation of the overall volume of data matched by the selection criteria before the rate-limit was applied but they may affect measures calculated on the basis of the data received. The mechanisms by which rate-limiting is applied thus complicate the analysis of Twitter data and we discuss issues of bias below.

Data structure

The data streamed contains all the essential information about a tweet (i.e. its metadata). This includes profile information for this user at the time the tweet is sent. Entities embedded in a tweet (links, hashtags, mentions) capture relational properties of tweets (see **Table 2**) and are identified by Twitter and represented as structured data items. The language of a tweet is also detected. **Figure 1** shows some of the interesting elements in a tweet. A complete description is available in the Twitter Developer Documentation (2014). In addition to the simple attributes like the date of posting, the tweet id, the text, etc. the data also contains a number of nested elements. In the case of the tweet in **Figure 1**, these are the profile of the user who posted the tweet, the original tweet that was retweeted, as well as any embedded entities the tweet contains, in this case a mention (@guardian), a URL and an embedded media element. **Figure 2** shows the embedded profile information for the tweeter.

One feature that is of potential interest is the identification of the place from where the tweet was posted. This is available when users turn on the geo-location services of their devices and allow Twitter to use the resulting coordinates. Where possible, Twitter will map the coordinates to a set of known locations and make these available as part of the place attribute. Our experience shows that only a small number of tweets contain such directly useful place information but it is possible to get some indication of location from the user profile data as well (see discussion of profile information below).

The retweeted_status element in this tweet indicates that it is a retweet of another tweet (one from the *Guardian*) and it contains a copy of this tweet's data. The number given in the retweet_count attribute shows how often that tweet has been retweeted. The entities element contains information about embedded user mentions (@guardian) and links embedded in the tweet. In this case, the link in the tweet itself is http://t.co/rPUBFxWj04, a link to Twitter's own link shortening service, which expands to a link to the Guardian's own link shortening service (http://gu.com/p/4xn2c/tw), which Twitter provides in the data in the entities section. Twitter does not fully resolve links but uses only the data in its own link shortening service. When a user posts a link that is already shortened, this will not be unpacked by Twitter (as illustrated by the *Guardian* link in the example), so any analyses of links need to take this into account. Also, some attributes are optional, so they may be set to null or simply be missing. This is the case, for example,

where the user has not enabled geo-services or has not allowed their Twitter client to make use of them.

Property name	Property meaning	Property implementation	Example tweet
Mention	The tweet's author mentions another Twitter user (or themself) in their tweet.	The name of the mentioned user is put in the body of the tweet -with an "@" sign at the beginning.	BBCBreaking: "Join **@BBCSport** coverage of #Sochi2014 #Paralympics #OpeningCeremony from 1600GMT http://bbc.in/1geTYjV"
Reply	The tweet's author replies to another Twitter user (or to themself) in their tweet.	The replying user uses a "reply" button of the Twitter interface while viewing the other user's tweet and starts the tweet with "@" and the other user's name.	greenlin45: "**@BBCBreaking** @BBCSport no chance"
Retweet	A Twitter user retransmits the tweet of another user.	The retransmitting user uses a "retweet" button of the Twitter interface while viewing the other user's tweet.	J B Smith **retweeted** BBCBreaking: "Join @BBCSport coverage of #Sochi2014 #Paralympics #OpeningCeremony from 1600GMT http://bbc.in/1geTYjV"
Hashtag	A Twitter user relates their tweet to a topic formulated by a keyword.	The keyword is put in the body of the tweet with the "#" sign at the beginning.	BBCBreaking: "Join @BBCSport coverage of **#Sochi2014 #Paralympics #OpeningCeremony** from 1600GMT http://bbc.in/1geTYjV"

Table 2: Relational properties of tweets (Twitter Help Center 2014a; Twitter Help Center 2014b; Twitter Help Center 2014c)

Figure 2 shows the profile information provided with each tweet by the streaming service. It contains both a human-readable identifier for the account, the screen_name and technical identifiers that Twitter uses internally and in their externally accessible services. In addition, a number of user-provided fields such as location, description, url or time_zone are included. The information in these attributes can be trusted only as far as users provide accurate information and, as it consists of free text, coding will be required before it can be utilised. It is possible, to some extent, to automate this process using machine learning tools however the results will not be accurate enough to trust an individual coding, just to trust that they are good enough as an average. Finally, attributes such as friends_count and followers_count provide information about how well connected an account is. This is shown by how many accounts it is following (the 'friends') and how many accounts are following it (the 'followers'). The statuses_count shows how many tweets the account has posted.

```
"user": {
 "id_str": "70930287",
 "name": "Alex Voss",
 "screen_name": "alexvossuk",
 "location": "St Andrews",
 "description": "Lecturer in Software Engineering, University of
St Andrews",
 "url": "http://t.co/Cadp6NS89O",
 "entities": {...},
 "protected": false,
 "followers_count": 118,
 "friends_count": 39,
 "listed_count": 9,
 "created_at": "Wed Sep 02 10:12:06 +0000 2009",
 "favourites_count": 240,
 "utc_offset": 3600,
 "time_zone": "Edinburgh",
 "geo_enabled": false,
 "verified": false,
 "statuses_count": 527,
 "lang": "en",
```


Figure 2: Profile information contained in tweet data

Other Twitter services

In addition to the Streaming service, Twitter provides another set of services that allow the user to obtain information about specific elements from Twitter, for example, the data for a tweet given by its identifier (illustrated in **Figures 1** and **2**). Twitter clients (such as Tweetdeck) use these services to interact with Twitter but their use as a data collection mechanism is even more restricted than the Streaming services, although they have the advantage that they can retrieve information post-hoc. In the studies presented in this issue, they were not normally used for data collection but were used, for example, to convert account screen names to Twitter account IDs. We did use the REST API to collect some data retrospectively in one case (cf. Dennis, this issue). Generally, the use of the Streaming API is preferable as the REST API is subject to more restrictive rate-limiting mechanisms and only provides access to a limited number of tweets for each account. The REST API also does not provide information about replies and retweets. There is also an API for the Twitter Search functionality but as Twitter does not make any guarantees about the results returned it is not suitable for research purposes.

Analysis methods

A variety of analytical methods can be applied to social media data to fulfil the requirements of audience research. We now discuss those that are applied to the data collected from Twitter for this special issue. Detailed discussions of each method can either be found in each paper where they are employed or, for the 2012 London Olympic Games studies, in

Dennis (this issue). Here we provide a short overview of these analytical modes, so the reader can understand how we approach audience research, based on the Twitter data. We will also refer back to the information requirements for audience research outlined earlier in this paper and numbered (a) to (l).

Firstly, the collected tweets are analysed dynamically. This includes building **timeline graphs**, that is, the graphs for the distributions of tweets over time. While a simple measure, it can, nevertheless, be revealing. A number of papers make use of timeline graphs. A striking example appears in the paper of Hutchings et al. (this issue) showing tweets on a number of political issues prominent in the run-up to the Sochi Winter Olympic Games trail off soon after the start of the competition and the topic of Ukraine rising sharply after 17 February 2014 during the climax of the Euromaidan protests (a, d, k). This analysis makes use of the created_at field and information about the hashtags embedded in a tweet as provided by the entities property.

Willis, Fisher and Lvov (this issue) use the attributes retweeted_status and in_reply_to_status_id of Twitter data to produce a **social network graph** that represents the interactions of Twitter users in English-language discussions in relation to the London Olympic Games. Such a network graph provides information on the size of an audience (c) and other sources of information valued by the audience (d). Subsequent in-depth analysis of this graph allows the authors to identify specific subgroups of the audience defined by their connection to specific BBC accounts that address their interests (b). The second part of the paper uses the same relational parameters for Twitter data (see **Table 2**), as well as the timestamps of the tweets (created_at), to estimate the **ageing factor**, which shows the durability of the audiences' engagement with the BBC's tweets (l). Finally, a comparison of retweets and replies shows how willing the audience is to engage with the BBC compared to simply re-transmitting BBC content (l).

Hutchings et al. (this issue) perform **hashtag co-occurrence** analysis to identify the topics that were particularly prominent during the Sochi Winter Olympic Games (a). They subsequently use the 'retweet_count' attributes of the collected tweets to identify the most retweeted tweets for each topic. **In-depth reading** of such tweets informs the authors on the public sentiment of the social media audiences (k). Finally, the most retweeted accounts for two of the topics examined are also identified (d). The analysis of the Twitter data complements the paper's main analytic strand, which focuses on the analysis of broadcast content. Pussy Riot, a Russian feminist punk rock protest group, whose members were beaten and arrested by security forces during the 2014 Sochi Winter Olympic Games is taken as an example. The combined analysis shows the interaction between the activism on the ground, the 'hijacking' of the global media event of the Olympic Games by dissemination of pictures and videos to social media and news media as well as the public reaction to these unfolding events (l) (cf. also Burchill, this issue).

Selected tweets or information flows can be analysed in-depth. For this purpose, **coding** and **in-depth reading** are employed. These techniques, as stand-alone analytical tools, have a natural limitation of being applicable only to small subsets of tweets. However,

in this special issue, it is effectively used either in interaction with other methods or as a tool for analysing corpora of tweets selected using strict criteria (cf. Dennis et al. 2015). A number of papers on the London Olympic Games provide examples. For example, Aslanyan (this issue) shows that the Russian audience used Twitter to favourably compare the coverage of the London Olympic Games Closing Ceremony by BBC Russian to the coverage by the national Russian channels (e, i).

Our approach to focus on media events has the advantage that it is relatively easy to bring together social media data associated with unfolding events, observations of how media organisations are themselves influencing their reflection in the social media sphere and how they are in turn reflecting social media in their own programming schedules. The papers by Aslan, Dennis and O'Loughlin as well as by Hutchings et al. provide strong examples.

Limitations of social media analysis

So far we have argued that the analysis of social media data fulfils the requirements of audience research and, in a way, we have distinguished these data from data collected using traditional methods. We see the analysis of social media data as complementing traditional research methods rather than the definitive solution for audience researchers.

Social media data, as well as the analytical modes applied to them, have limitations of their own. However, inasmuch as the social media data are very different in their nature and properties from the traditional audience research data, so they are *different in their limitations*. Drawing on the research presented in this special issue, what these limitations mean for researchers in practice will be illustrated.

The key distinctive properties of social media data are that they are naturally occurring, nearly in real-time and at population scale. Such properties allow researchers to avoid some of the limitations they would face with traditional research data, for example, risks of leading interview questions (Foddy 1993) or aforementioned biases of survey samples (Mytton 2011). However, the very same properties imply that a researcher has limited control over the quality of the data and the degree to which it is tailored to the research question and to the researched population (Purdam and Elliot 2015). The huge scale of social media data amplifies these issues and makes them more difficult to analyse (Murphy 2015). This results in a range of limitations for the use of social media data in audience research.

Firstly, social media data are generated by a self-selected sample of people who choose to engage on social media. Hence, these data may give a biased presentation of the wider audience. Social media users are not generally representative of wider populations and their observable behaviour on social media is not necessarily a reflection of their behaviour offline (e.g., Bruns and Stieglitz 2014, Ruths and Pfeffer 2014).

Most of the research presented in this special issue has not been affected by this limitation, as we mostly examine the behaviour of the audience on social media as a

research subject in itself. However, when we are seeking to derive conclusions that go above and beyond the level of social media, we analyse data from additional sources. For example, Aslan, Dennis and O'Loughlin (this issue), who examine the media controversies around the Chinese swimmer Ye Shiwen's performance in the 2012 London Olympic Games, enhance the social media analysis with the analysis of media coverage of the event.

Some limitations are associated with the quality of social media data. For example, different attributes of the data can be of varying completeness. In the case of Twitter, some tweets contain geo-coordinates but most do not. Evidence suggests that the number of tweets that carry geo-location information is less than one percent (Graham et al. 2014). This special issue does not report any geo-analysis and hence is not affected by this limitation.

Another limitation is that some attributes of social media data may be less reliable than others. For example, profile information provided by users is not easily verified. Users may record a location that does not reflect their normal residence or they may prefer to provide the location of their workplace, as in the example in **Figure 2**. Also, the aforementioned data on a tweet's language comes from the language detection algorithms used by Twitter. These kinds of algorithms are less than perfect, especially for short texts such as tweets. Graham et al. (2014) measure the rate of agreement between several automatic language-detection algorithms and manual language-detection by human coders on a sample of 4000 tweets from various locations. They find that the most accurate of the tested algorithms agree with human-coding in only 76.4% of cases.

Several papers in the special issue do rely on language-specific datasets. With one exception, these datasets were obtained through selecting the tweets by or addressed to language-specific Twitter accounts. Hence, there is no reliance on the language-detection mechanisms. Willis, Fisher and Lvov (this issue), by contrast, do rely on tweets that have been detected as English by Twitter. However, their analysis is primarily concerned with identification of the most influential actors in the Twitter interaction examined (for example, the most retweeted accounts). As the key actors identified tweet in English, the analysis proves to be not affected by mistakes in language detection.

The natural occurrence of social media data, compounded by their scale, results in another limitation of social media analysis: its outcomes largely depend on the data a researcher selects to analyse. For instance, in a case of filtering Twitter data by using relevant hashtags, González-Bailón et al. (2014) show that employing an incomplete list of hashtags can significantly shape descriptive statistics of the derived sample of tweets. At the same time, on occasion, a single hashtag is employed by Twitter users in tweets both related and unrelated to the topic of a study. For example, Rogers (2014) examined the trends of Twitter discussions around various football teams during the first half of the 2014 World Cup. He looked at the frequency of use of team-specific three-letter hashtags (for example, #BRA for Brazil) and pointed out that the hashtag #USA was used far more often than any of the others. While it may be the case that this reflects the popularity of the

United States football team, we would argue that a more probable explanation is that there are other uses of the "#USA" hashtag that have added to the volume of traffic.

Research presented in this special issue is only marginally affected by the limitations discussed above. As outlined by Dennis (this issue), in the case of the London Olympic Games we collected data on interactions around particular Twitter accounts that were selected following consultation with the BBC.com audience research team. In the case of the Sochi Winter Olympic Games we adopted a different approach, making the initial collection criteria as broad as possible and then performing exploratory analysis of the collected corpus to inform the filtering parameters. Please refer to Hutchings et al. (this issue) for details.

The last limitation is related to the issue of rate-limiting (cf. section 'Rate Limiting Explained'). The amount of data provided through the Streaming API is limited to 1% of the total Twitter stream over time. A collection of a small set of tweets may not be affected by the rate limiting but a collection selecting a large number of tweets will return only a sample plus a message indicating how many tweets have been removed from the data stream.

If the pace at which those tweets are produced is uneven in time, so will be the limiting rate, that is, while part of the tweets from the peak-times will be cut-off, the tweets from the valley-times will be captured in full. For this reason, the sample of tweets provided by the Streaming API under broad collection criteria may be biased towards the periods of time with lower Twitter activity. Morstatter et al. (2013) show that this can significantly distort the descriptive statistics of a collected corpus of tweets and can influence many of the social media analysis methods mentioned in this paper.

In general, the messages that Twitter sends if the rate limit is reached (cf. Rate Limiting Explained above) can be used as a mechanism for tracking the rate limiting. However, while it is certainly possible to accurately track the overall volume of tweets over time using these messages, so far no general mechanism for *repairing* the bias created by the rate limit has been suggested that would mitigate the impact on other research methods.

The research on the London Olympic Games presented in this special issue is not affected by rate limiting, since the selected tweets never amounted to more than 1% of the total Twitter stream. We suggest that this would be the case with most collections based on accounts rather than keywords or hashtags. As of April 2015, Twitter has 288 million active users per month,[4] while the collection criteria may include only up to 5000 accounts.

By contrast, while collecting data during the 2014 Sochi Olympic Games, we selected tweets by hashtags. Hence, our collection was rate-limited during the Opening Ceremony day, which was the day of peaking Twitter activity around the Olympic Games. However, this does not invalidate the outcomes of our analysis. Indeed, the message of the timeline graph (cf. analysis methods) could have been only stronger if we had obtained more tweets during the day of the Opening Ceremony. The rest of the analysis was done on a day-by-day basis; hence, under-representation of tweets from the Opening Ceremony day could not influence the results.

Essentially, all models are wrong, but some are useful (Box and Draper 1987:424). In the context of the limits of social media analysis, the famous statistical aphorism above also applies to data and methods. Models help to make sense of the real processes by capturing a subset of their properties while inevitably losing the information on the rest of them. Similarly, selected properties of real phenomena are captured by data, while selected properties of the data are captured by methods. From this perspective, all methods are wrong, all data are bad, but some are useful. In the same manner in which complicated social processes may require a multitude of models to be adequately described (cf. a whole range of industrial relationship models in economics), complicated social objects like audiences may require a variety of data and methods to be properly understood. Hence, social media analysis, while arguably a great help for audience researchers, does not make the use of the traditional audience research methods unnecessary. Rather it is their interaction that can bring the greatest insight.

To this end, this paper has discussed how social media data can aid audience researchers in fulfilling the requirements of audience research. These requirements are derived from the three purposes of audience research, which are accountability, advocacy and strategy. However, in the next section we will show that social media and their analysis may satisfy a fourth purpose of audience research: tactics.

Tactical uses of social media data

Social media are not simply a new source of data for traditional audience research. They also offer broadcast media organisations new opportunities for interaction with their audience, providing the capability to use social media data tactically in making decisions about programme content and thus transforming how they relate to their audience. Hence, for broadcast media organisations, social media need to be understood as phenomena to be studied in their own right, which requires taking seriously the 'big data' nature of social media.

There are at least two interrelated ways to define 'big data' and these have an influence upon potential tactics that news organisations can develop to engage their audiences. The first considers the *attributes* of the data itself. Doug Laney originally defined big data in terms of Volume, Velocity and Variety (Laney 2001), that is, data is large in comparison to the capacity of a single computer, it arrives at a high rate in real-time or it consists of a range of different types of data (for example, text and video data). Others have since identified additional attributes that help to characterise big data and big data applications. Kitchin (2013) mentions its exhaustive nature. It seeks to capture entire populations or systems rather than being a sample. He also points out that it is often fine-grained, reflecting some frequently-occurring real-world events such as orders in a shopping system or page-views on a website.

An alternative definition might focus on the *uses* of big data rather than its *attributes*. For example, Davenport, Barth and Bean (2012) suggest an alternative definition

of big data that focuses on its business applications. Traditionally, organisations dealt with routine transactional data processing (e.g., updating users' bank accounts) and infrequent analytical data processing (e.g., analysing sales figures for managerial reports) separately. However, in the age of big data analytical processing of large datasets can also be performed routinely and hence be embedded into the core of the business processes. For example, social media services such as Twitter provide timeline information for a user, including posts from a range of people they are connected to and they *provide analyses* of what topics are 'trending' within the service in real time.

Picking up on the novel uses that data are put to, Boyd and Crawford (2012:3) define big data as 'a cultural technological and scholarly phenomenon that rests on the interaction of: technology … analysis … [and] mythology…' Defined in this way, big data is as much about what people do with data, the context in which this doing takes place and the implications, as it is about the attributes of the data itself as in the more technical definitions mentioned earlier. 'Big Data is less about data that is big than it is about a capacity to search, aggregate, and cross-reference large data sets (*ibid.*:*2*). Such a definition allows Boyd and Crawford to strip down 'both utopian and dystopian rhetoric' (ibid:3) around big data and examine the socio-technological changes – and challenges – provoked by it (including the ethics of 'big data' research, cf. Anderson and Jirotka 2015).

The change in uses of data is fundamentally enabled by the changing attributes of the data available as well as by technical and organisational innovations associated with the big data trend. While the technical challenges in some application areas are immense (Jagadish et al. 2014), many big data applications have been developed in the recent past. Commoditized technologies such as Hadoop, Spark or Storm (all three available as open source from the Apache Foundation) offer the potential to bring big data applications into the mainstream. Platforms such as COSMOS make functionality for social media analysis readily available to end-users.

News media organisations are devising strategies to utilise data from social media, from their websites' click-streams and other sources to orient their programming to the interests and activities of the active audience, as well as to integrate user-generated content (for example, Hermida 2013; Newman 2011). An example from the 2012 London Olympic Games would be the Guardian Second Screen App (Richards 2012), which incorporated Twitter volume data into a central timeline navigation element, allowing readers to navigate the day's events not only by time but also by the level of Twitter traffic. The spikes visible for key events allowed for effective navigation to points of interest such as Mo Farah's gold medals.

This example shows how going beyond the now common use of individual tweets in programming, social media can play additional roles such as, as in this case, providing navigation to content of interest. On the whole, however, media organisations have barely begun to explore the full range of opportunities for programming inherent in social media. As an example, we have been collaborating with Wire Free Productions[5] an independent media production company to deliver the 'BBC Radio 5 live Hit List', a 'Top 40' of the most

popular UK stories on social media broadcast weekly on the channel.[6] This format is an example of how social media can be brought into journalistic practice as 'big data', as opposed to as individual messages.

Conclusions: Iteration and reflection on methods

The challenges social media data poses for existing analytical frameworks and opportunities for methodological innovation it provides highlight the importance of reflecting carefully on research design for studies that seek to exploit the potential of social media for explicating social phenomena. Social media not only offers a window onto the phenomena of interest but also is inextricably implicated, through the agency of individuals and organisations (large and small) alike, in shaping these phenomena.

The media landscape is changing week by week, with new apps, software and platforms being used by audiences and broadcasters alike, presenting new challenges for commercial and academic audience researchers. In this context, it is impossible to apply a static research design over any period longer than a few months, as the services that comprise 'media' will have expanded and together with them the units for analysis. What is required is a methodological approach open to adding new layers and sources of data as well as patching together comparisons in an experimental way. Paradoxically, an approach that is agile and fluid gives us the surest grip.

Hence, the research presented in this special issue is a composite of different methods and data sources employed at different times. We argue that these fragments give us as good an impression of the whole as is realistic in a period of methodological flux (Halfpenny and Procter 2015). They give us slices of the event, chosen because they are relevant to our research questions but which defy 'systematicity' or 'comprehensiveness'. Such standards are not currently possible in audience research because data sources and media keep changing. Testing the methods was as important as using them to answer research questions. Methodological developments linked to the emergence of big data are happening continually and we cannot yet be certain what impact such data will have on audience research or more generally on social science research.

Acknowledgements:

We would like to thank the reviewers for their constructive comments, which have helped shape this paper.

Biographical Notes:

Rob Procter is Professor of Social Informatics in the Department of Computer Science, University of Warwick. Rob led a multidisciplinary team working with the Guardian/LSE on the 'Reading the Riots' project, analysing tweets sent during the August 2011 riots. Current projects in social media include studies of rumours and 'digital wildfires'. He is co-editor with Peter Halfpenny of 'Innovations in Digital Research Methods' published by Sage.

Alex Voss is Lecturer in Software Engineering in the School of Computer Science, University of St Andrews. Alex has a long-standing interest in Social Informatics and has been studying uses of social media data in social science research as well as in journalism. He has collaborated with the Guardian as well as Wire Free Productions Ltd. on forms of open journalism drawing on social media data.

Ilia Lvov is a PhD student in the School of Computer Science at the University of St Andrews working under supervision of Dr Voss. He studies the methodological challenges that are posed by new forms of data in social research. Ilia worked as a consultant for the 'Cultural Value' research project and participated in the analysis of Twitter data on the 2012 London Olympic Games documented in this special issue. He also was a member of an interdisciplinary research team that studied the 2014 Sochi Winter Olympic Games as a global media event. This research is also presented in this special issue.

References:

Anderson, R.J. and Jirotka, M, 'Ethical Praxis in Digital Social Research', in Peter Halfpenny and Rob Procter (eds.), *Innovations in Digital Research Methods*, London: Sage, 2015.

Boyd, Danah and Kate Crawford, 'Critical questions for Big Data: provocations for a cultural, technological, and scholarly phenomenon', *Information, Communication & Society,* 15(5), 2012, pp. 662–679.

Box, G. E. P. and N. R. Draper, *Empirical Model Building and Response Surfaces*, New York, NY: John Wiley & Sons, 1987.

Bruns, Axel, *Blogs, Wikipedia, Second Life, and Beyond: From Production to Produsage*, New York: Peter Lang, 2008.

Bruns, Axel and Stefan Stieglitz, 'Quantitative approaches to comparing communication patterns on Twitter', *Journal of Technology in Human Services,* 30(3-4), 2012, pp. 160–185.

Bruns, Axel and Stefan Stieglitz, 'Twitter data: what do they represent?', *it - Information Technology,* 56(5), 2014.

Burrows, R. and M. Savage, 'After the Crisis? Big data and the methodological challenges of empirical Sociology', *Big Data & Society,* 1(1), 2014, pp. 885-899.

Davenport, Thomas H., Paul Barth and Randy Bean, 'How "Big Data" is different', *MIT Sloan Management Review,* 2012, pp. 22–24.

Doctorow, Cory, *Makers*, London: Harper Voyager, 2009.

Dutton, William H., 'The Fifth Estate emerging through the Network of Networks', *Prometheus*, 27(1), 2009, pp. 1-15.

Foddy, William H., *Constructing Questions for Interviews and Questionnaires: Theory and Practice in Social Research*, Cambridge, UK; New York, NY, USA: Cambridge University Press, 1993.

Gibson, James J., 'The theory of affordances', in Robert Shaw and John D. Bransford (ed.), *Perceiving, Acting, and Knowing. Towards an Ecological Psychology*, Hillsdale NJ: Lawrence Erlbaum Associates, 1977.

Gillespie, Marie, Hugh Mackay and Alban Webb, 'Designs & devices: towards a genealogy of audience research methods at the BBCWS, 1932-2011', *Participations: Journal of Audience & Reception Studies,* 8(1), 2011, pp. 1-20.

Gillespie, Marie, Simon Bell, Colin Wilding, Alban Webb, Ali Fisher, Alex Voss, Andrew W. M. Smith, Jess Macfarlane; Nat Martin, Tot Foster and Ilia Lvov, *Understanding the Changing Cultural Value of the BBC World Service and the British Council*, Arts and Humanities Research Council, Swindon, 2014.

Gillmor, Dan. *We the Media: Grassroots Journalism by the People, for the People*. 1st ed. Beijing; Sebastopol, CA: O'Reilly, 2004.

González-Bailón, Sandra et al., 'Assessing the bias in samples of large online networks', *Social Networks,* 38, 2014, pp. 16–27.

Graham, Mark, D. J. Hale and Devin Gaffney, 'Where in the world are you? Geolocation and language identification in Twitter', *The Professional Geographer,* 66(4), 2014, pp. 568–578.

Halfpenny, P, and R. Procter, 'Introduction and Overview' in Peter Halfpenny and Rob Procter (eds.), *Innovations in Digital Research Methods*, London: Sage, 2015.

Hermida, Alfred, '#JOURNALISM: Reconfiguring journalism research about Twitter, one tweet at a time', *Digital Journalism* 1(3), 2013, pp. 295–313.

Hoskins, A, 'Death of a single medium', *Media, War & Conflict,* 6(1), 2013, pp. 3–6.

Housley, W. et al., 'Big and broad social data and the sociological imagination: a collaborative response', *Big Data & Society,* 1(2), 2014, pp. 1-15.

Jagadish, H. V. et al., 'Big Data and Its Technical Challenges', *Communications of the ACM,* 57(7), 2014, pp. 86–94.

Johnson, R. B., A. J. Onwuegbuzie and L. A. Turner, 'Toward a definition of mixed methods research', *Journal of Mixed Methods Research,* 1(2), 2007, pp. 112–133.

Laney, Doug, '3D data management: controlling data volume, velocity, and variety', *Application Delivery Strategies*, 6 Feb. 2001, pp. 1-5

Mackay, Hugh, and Jingrong Tong, 'Interactivity, the global conversation and World Service research: digital China', *Participations: Journal of Audience & Reception Studies,* 8(1), 2011.

Mason, Jennifer, '*Six Strategies for Mixing Methods and Linking Data in Social Science Research*', NCRM Working Paper Series, Manchester: ESRC National Centre for Research Methods, 2006.

Morstatter, Fred et al., 'Is the sample good enough? Comparing data from Twitter's Streaming API with Twitter's Firehose', *International Conference on Weblogs and Social Media*, AAAI, 2013, pp. 400–408.

Murphy, Joe, 'Survey methods: challenges and opportunities', in Peter Halfpenny, and Rob Procter (eds.), *Innovations in Digital Research Methods*, London: Sage, 2015.

Mytton, Graham, 'Audience research at the BBC World Service 1932-2010', *Participations: Journal of Audience & Reception Studies,* 8(1), 2011, pp. 75–103.

Napoli, Philip M., 'Audience evolution and the future of audience research', *International Journal on Media Management,* 14(2), 2012, pp. 79–97.

Newman, Nic (2011) 'Mainstream Media and the Distribution of News In the Age of Social Discovery. Reuters Institute for the Study of Journalism' [WWW document] URL http://ora.ox.ac.uk/objects/uuid:94164da6-9150-4938-8996-badfdef6b507

Newman, N., W. H. Dutton and G. Blank, 'Social media in the changing ecology of news: The fourth and fifth estates in Britain', *International Journal of Internet Science*, 7(1), 2012, pp. 6-22.

Newman, Nic and David A. L. Levy, '*Reuters Institute Digital News Report 2014'* Reuters Institute for the Study of Journalism, 2014.

Observer editorial (16th February 2014) 'Should UK licence-fee payers still fund the World Service?' [WWW document] URL http://gu.com/p/3mnda/stw [visited 24/04/2015]

Ofcom, *Measuring Online News Consumption and Supply*, July 2014.

O'Loughlin, Ben and Andrew Hoskins (January 14th, 2015) 'Arrested War: After Diffused War', Global Policy Journal [WWW document] URL http://www.globalpolicyjournal.com/blog/14/01/2015/arrested-war-after-diffused-war [Visited 24.05.2015]

Procter, Rob, Farida Vis and Alex Voss, 'Reading the riots on Twitter: methodological innovation for the analysis of big data', *International Journal of Social Research Methodology,* 16(3), 2013, pp. 197–214.

Purdam, K. and M. Elliot, 'Exploiting New Sources of Data' in Peter Halfpenny and Rob Procter (eds), Innovations in Digital Research Methods, London: Sage, 2015.

Richards, Jonathan (15 August 2012) 'Reimagining the timeline: behind our Second Screen web app' on Guardian Developer Blog [WWW document] URL http://gu.com/p/39jh6/stw [Visited 24/04/2015]

Ritzer, G. and N. Jurgenson, 'Production, consumption, prosumption: the nature of capitalism in the age of the digital 'Prosumer', *Journal of Consumer Culture,* 10(1), 2010, pp. 13–36.

Rogers, Simon (27th June 2014) 'The Twitter #WorldCup group stage recap' [WWW document] URL https://blog.twitter.com/2014/the-twitter-worldcup-group-stage-recap

Rosen, Jay (27 June 2006) 'The People Formerly Known as the Audience' [WWW document] URL http://archive.pressthink.org/2006/06/27/ppl_frmr.html [Visited 20/04/2015]. Also in Michael Mandiberg (ed.), *The Social Media Reader*, New York: New York University Press, 2012.

Ruths, D. and J. Pfeffe, 'Social media for large studies of behavior', *Science,* 346(6213), 2014, pp. 1063–1064.

Savage, M. and R. Burrows, 'The coming crisis of empirical sociology', *Sociology* 41(5), 2007, pp. 885–899.

Savage, M, 'The 'Social Life of Methods' in a critical introduction', *Theory, Culture & Society,* 30(4), 2013, pp. 3–21.

Tashakkori, Abbas and Charles Teddlie (eds.), *Handbook of Mixed Methods in Social & Behavioral Research*, Thousand Oaks, Calif: SAGE Publications, 2003.

Thelwall, Mike, Kevan Buckley and Georgios Paltoglou, 'Sentiment in Twitter events', *Journal of the American Society for Information Science and Technology,* 62(2), 2011, pp. 406–418.

Thompson, Mark (25 April 2006) 'BBC Creative Future: Mark Thompson's Speech in Full', *Guardian* [WWW document] URL http://www.theguardian.com/media/2006/apr/25/bbc.broadcasting [Visited 22.09.2014]

Toffler, Alvin, *The Third Wave*, 1st ed., New York: Morrow, 1980.

Twitter Streaming API Documentation, [WWW document] URL https://dev.twitter.com/streaming/overview [accessed 21.09.2014]

Twitter Help Center, 'FAQs about Retweets (RT)', Web. 18 Dec. 2014. [WWW document] URL https://support.twitter.com/articles/77606-faqs-about-retweets-rt [accessed 05.05.2015]

---, 'Types of Tweets and Where They Appear', Web. 18 Dec. 2014. [WWW Document] URL https://support.twitter.com/articles/119138-types-of-tweets-and-where-they-appear [accessed 05.05.2015]

---, 'Using Hashtags on Twitter', Web. 18 Dec. 2014. [WWW document] URL https://support.twitter.com/articles/49309-using-hashtags-on-twitter [accessed 05.05.2015]

Webb, Alban, 'A leap of imagination: BBC Audience Research over the Iron Curtain', *Participations: Journal of Audience & Reception Studies,* 8(1), 2011, pp. 154–172.

Webb, Alban, 'An uncertain future for the BBC World Service', *The Political Quarterly* 85(3), 2014, pp. 378–383.

Williams, Matthew L. et al., 'Policing cyber-neighbourhoods: tension monitoring and social media networks', *Policing and Society*, 23(4), 2013, pp. 461–481.

Notes:

[1] We consider mainly broadcast media but the traditional distinction between broadcast and other forms of news media organizations is becoming blurred through online delivery platforms and changes in publication rhythms that these entail.

[2] There is no simple way to scope what 'mainstream media' entails. In this day and age this arguably includes organisations that started out as challengers to more traditional print and broadcast media. The Huffington Post is arguably an example of an organisation that has become part of the 'mainstream'.

[3] 'The maker movement is a trend in which individuals or groups of individuals create and market products that are recreated and assembled using unused, discarded or broken electronic, plastic, silicon or virtually any raw material and/or product from a computer-related device.' (http://www.techopedia.com/definition/28408/maker-movement) It has arguably re-invented 'DIY' with the help of new technologies such as 3D printers.

[4] https://about.twitter.com/company

[5] http://www.wirefreeproductions.com/

[6] http://www.bbc.co.uk/programmes/b04p59vr