The Social life of real-time social media monitoring

Lawrence Ampofo, University of London, UK

Volume 8, Issue 1 (May 2011)

Real-time social media, social media that publish information as soon as it is available, has become a mainstay in contemporary society with the widespread adoption of status updates, tweets and blogging. In response to this growth of data, specific methodologies and software tools have been developed that aggregate and analyse the saliency of such content. However, despite the wealth of resources available, researchers face significant challenges in accurately and ethically conducting such an endeavour. This essay takes as its starting point the conception that real-time social media are artefacts of social and cultural interactions online. The analysis of such real-time information is therefore problematic as ethical and methodological issues suitable for such research are currently not well developed. Two case studies from BBC Mundo’s Your Say discussion boards contextualise the complexity inherent in the above issues within the framework of the BBC World Service’s mission to foster a global conversation. The analysis of discussions hosted by BBC Mundo highlights the intricate nature of correctly analysing such content and underscores the need for new methodological processes, in addition to heightened analytical sensitivity, in interpreting the results of real-time social media analyses.

Keywords: Social media; real-time internet research; social media monitoring; mobile device; public diplomacy; BBC World Service.

Introduction

The Internet is not only a text-based medium made up of communities, newsgroups and email lists. It is also a medium…where users can take control of the means of production, create their own cultural artifacts [sic] and intervene in the production of existing ones. The Internet can thus be perceived as a form of cultural production. (Bassett & O’Riordan, 2002 p. 1)
Growth in the use of social media in 2009 was without precedent in the history of the internet\textsuperscript{1} and World Wide Web (Web). The ubiquity of social media has been enabled by the widespread availability of Web-enabled devices such as the growing array of mobile devices and personal computers. This in turn has enabled the proliferation of user-generated content, exemplified by the recent publication that 22 percent of all time spent online in 2009 was with Web 2.0 social media sites (Nielsenwire, 2010).

Since its inception in 1979 with the creation of Internet Relay Chat, bulletin boards and Usenets, social media has increasingly provided its users the ability to consume, create and make their content available to a global audience in real-time. Whether via status updates, news curation or instant message services, social media now records contributions from its users in real-time to the extent that the Web is: ‘…the most comprehensive electronic archive of written material representing our world and peoples’ opinions, concerns, and desires’ (Eysenbach & Till, 2001, p.2).

The production of user generated content raises significant ontological problems over the conduct of real-time social media monitoring (RTSMM) programmes. For example, RTSMM can be construed as the analysis of data from websites hosted on global communications networks called the internet and the World Wide Web. RTSMM, however, is more complex and should be conceived of as the analysis of artefacts from a social and cultural space as outlined above by Bassett & O’Riordan (2002, p.1).

This essay takes as its starting point the understanding that RTSMM is fundamentally an analysis of social and cultural spaces. If RTSMM is taken as the analysis of the production of a social and cultural milieu, then it is important to consider whether it is possible to conduct accurately due to the fact that the observer of social and cultural data has to be sensitive to the various social and cultural complexities inherent within such data. As such it is questionable whether RTSMM can achieve its ambitious target of accurately monitoring and analysing social media content. For this reason, this essay will examine some of the difficulties associated with RTSMM such as the ethical privacy issues. The essay also evaluates some of the methodological approaches used to conduct RTSMM and evaluates the advantages and disadvantages associated with their use.

The essay will also consider how the British Broadcasting Corporation, as a social and cultural space, can utilise real-time social media to foster a global conversation. This issue will be framed in the context of two case studies from the BBC Mundo website in which audiences posted responses to articles detailing the problem of narcotics in Mexico and Raul Castro’s re-shuffling of the Cuban Government in 2008. An analysis of discussions hosted by BBC Mundo highlights the intricacies involved when correctly analysing such content while

\textsuperscript{1}Footnote: This refers to the use of social media in 2009 being without precedent in the history of the internet and World Wide Web.
demonstrating the need for new methodological processes, in addition to heightened analytical sensitivity, in interpreting the results of real-time social media analyses.

**What is Real-time Social Media Monitoring?**

For the first time in history...the media make possible a mass participation in a productive process at once social and socialized, a participation whose practical means are in the hands of the masses themselves. (Baudrillard, 1985, p.2)

Social media are an intrinsically important element of contemporary society. The late French philosopher Jean Baudrillard was correct in his argument that modern media is now simultaneously in the hands of society and participatory in nature. It is used as much for creating and maintaining personal relationships as for commerce and news generation.

In order to understand RTSMM, it is critical to first establish a clear definition of social media so the nature of the content being monitored in real time is understood clearly. Any definition of social media should include a description of Web 2.0 and user generated content. Web 2.0 is a term that was coined by the Web analyst Tim O’Reilly and was used to describe the more collaborative use of web technologies (O’Reilly, 2007, p.19). Web 2.0 sites allow its users to create, collaborate, share and publish their own content such as video, text and audio files. Websites such as Dictionary.com and MSN are spaces where the user simply consumed content without contributing to its creation, and are therefore technologically and culturally different to Web 2.0 sites. These sites are ostensibly different to other websites which actively seek user participation in order to create new content such as Wikipedia and the various blog platforms such as Wordpress and Blogger.

User generated content, which is produced using Web 2.0 as a technological platform, can be described as the creation of online content by the users of particular social media platforms. The Organisation for Economic Cooperation and Development (OECD) defined user generated content as needing to be placed on a website, to have demonstrated a degree of creativity and, finally, to not have been professionally created (OECD, 2007).

The combination of Web 2.0 technologies, and the resulting emergence of user generated content, gives rise to a group of Internet-based applications that build on the ideological and technological foundations of Web 2.0, and that allow the creation and exchange of User Generated Content’ (Kaplan & Haenlein, 2010, p.61). This definition of social media emphasises the notion that the internet is a social and cultural milieu in addition to its technological foundations. This definition supports the notion that RTSMM is the aggregation and analysis of social media content, as it is published, from social media portals.
There exists a range of processes through which to monitor and analyse real-time social media. However, the conception of the Web as a social space means that RTSMM programmes need to account for the social and cultural dynamics of the content they analyse. The next section of this essay outlines the evolution of RTSMM from 1979 to the present day.

**The Evolution of Real-Time Social Media**

The first iterations of real-time social media came in 1979 with the implementation of User Networks (Usenet), bulletin boards and Internet Relay Chat (IRC) portals, precursors to the more contemporary discussion forums and Instant Message clients.

UseNets, bulletin boards and IRC clients were described by their users as real-time media because users could comment on items of discussion as they appeared. Discussions were hosted on distributed servers and anyone in the world could access them and contribute. Indeed, Usenets and bulletin boards have been described as the first internet peer-to-peer technology and global repositories of useful knowledge as ‘so many questions are asked and answered there. It is also particularly useful when looking for information about late-breaking or non-mainstream subjects likely to be part of the popular conversation.’ Usenets and bulletin boards were the first incarnation of what would later become discussion forums in which users could participate in ‘live’ online conversations.

IRC portals were the first incarnation of modern-day instant message clients and chatrooms. Developed in 1988 by the Finnish graduate student Jarkko Oikarinen, IRC is a multi-user chat program that is still widely used today. IRC clients became widely used in 1990 as people relied on the technology to gather real-time information about events occurring in the Gulf War. Contemporary discussion forums and instant messaging services subsequently displaced Usenets, bulletin boards and IRCs as widely used social media platforms. Sizeable communities coalesced around topics of discussion on bulletin boards and Usenets, in the same way as today’s discussion forums.

The premise and technology behind discussion forums is today incorporated in virtually all online social media such as Web logs, video-sharing Websites and social networks. Conversations hosted on discussion forums operate in real-time, similar to that on Usenets and bulletin boards. Contemporary instant messaging services, such as Microsoft Network Messenger and ICQ, permit their members to converse with each other in real-time whilst sharing a range of other content such as images, documents and video.

In addition, massively multiplayer online role playing games (MMORPGs) are an important component of real-time social media. MMORPGs, best represented by World of Warcraft, are
online gaming portals that in addition to permitting its users to play video games, support live, in-game discussions on a range of topics.

Social networking sites and blogs have become deeply influential elements of social media, attracting a great number of users. Over 130 million blogs are now tracked by the influential blog search engine Technorati (Arrington, 2010), and one in four minutes spent by users online is dedicated to social networking sites (Nielsenwire, 2010). Sites such as Facebook, Friendster, LinkedIn, Hi5 and MySpace enable their users to create personal profiles that contain a range of content. In addition, contemporary social networking sites include a range of technologies that enable an array of real-time service, such as instant messaging, discussion forum and status update facilities, allowing users to update to their profile, and their larger network, details of what they are doing at any particular moment.

As well as the development of real-time social media services, the ubiquity of these sites has been further enhanced by the continued development of mobile devices. Mobile devices now allow social network users to update their profiles wherever they are, without the use of a desktop computer in a fixed space.

Real-time social and cultural interactions conducted on social media have resulted in an explosion of the volume of data available. While this is of immense value to researchers, it is simultaneously apparent that effectively monitoring, aggregating and analysing social media feeds is becoming increasingly difficult. This is due to the fact that RTSMM simultaneously monitors the value and impact of social interactions as well as the content of conversations and volume of website traffic. The following section examines the various tools and processes involved in the conduct of RTSMM.

**Real-Time Social Media Monitoring and Analysis**

The monitoring and evaluation of social media services has become increasingly difficult to conduct for a host of reasons. This is in direct contrast to the overwhelming volume of content available to the researcher, as mentioned. In the nascent stages of the development of social media platforms, access to the conversations hosted on discussion fora, IRCs and bulletin boards was widely permitted. Social media sites simply required potential users to create a profile, and access to all of the current and historical data was granted to them. Contemporary social media services however restrict widespread access to their data as part of their terms of service, restricting effective monitoring and analysis of social media content. The creator of the World Wide Web, Tim Berners-Lee, echoed this assertion by arguing that large social networks were a threat to universality on the Web by effectively ‘walling off information posted by their users from the rest of the Web’ (Berners-Lee, 2010).
In spite of the restrictions in openly accessing social media content, a range of software tools specially developed for social media monitoring have been created. Social media monitoring tools fall into two specific categories; Web analytics programs that quantitatively analyse the performance of specific sites based on the behaviour of its users, and other content aggregators that accumulate content from social media sites.

The term ‘Web analytics’ refers to the process of measuring, collecting, analysis and reporting of internet data for the purpose of understanding and optimising Web usage (Web Analytics Association, 2010). There are two categories of Web analytics which are on-site and off-site.

Off-site Web analytics involve the measurement of general Web activity and includes the measurement of a Website’s potential audience, share of voice and the overall volume of commentary in relation to general internet activity. They are macro-level tools which allow users to evaluate the performance of a website in relation to others.

On-site Web analytics involve the measurement of traffic to a particular Website. Such an analysis could include an examination of the landing pages that encourage people to make a purchase (Clifton, 2010).

Other software tools aggregate content from social media sites, such as the news filtering and market analysis tool CyberWatcher, the online print news database Lexis Nexis, and the news aggregation tool Moreover. More recently, the social media monitoring industry has seen the introduction of service providers such as Techrigy, Sysomos and Radian6 that incorporate a hybrid of content aggregation and Web analytic offerings. As a result, Radian6 and Techrigy both offer users the opportunity to witness the development of conversations and content while simultaneously analysing various aspects of the content such as sentiment and share of voice.

Recently, natural language processing (NLP) technology has been applied to social media monitoring. It automatically analyses text-based digital content, speeding up the process of hitherto time-intensive analysis techniques such as content and discourse analyses. NLP provides the opportunity to examine a vast range of content in a number of languages and nuances.

The use of both qualitative and quantitative social media monitoring tools provides a powerful mechanism through with which to analyse social media content in real-time. However, the ability of these tools to fully analyse the wealth of information available is somewhat limited by the access restrictions that companies place on their data. As part of their privacy policies, individual organisations, such as Facebook and Google stipulate that all content posted to
their services remains their property. However, other services, such as Twitter, have less restrictive rules on their data, and users can pay to access all the tweets generated or access a sample of content through Twitter’s application programming interface (unless users specifically request that their tweets be restricted to people in their network). In addition to restrictions in accessing data, there are other issues associated with quantitative measures of RTSMM.

Content posted on social media sites by users depicts their innermost feelings and daily activities, which bestows upon analysts a unique opportunity to gain insight from people of the kind that has traditionally been available through polling and surveys. This type of social interaction is complex to capture and is difficult to acquire and interpret using the aforementioned techniques. However, crisis mapping, polling public opinion political elections and news curation are techniques that go some way to interpreting this complexity.

The relatively new technique of crisis mapping is increasingly utilised by governments and humanitarian agencies as an early warning system to provide them with more information to help populations following complex emergencies. Crisis mapping is conducted using a ‘mashup’ of a wide range of information from the digital realm including blog posts, SMS messages, images, and audio files, which are then superimposed onto virtual maps in real-time (Meier & Leaning, 2009). The International Network of Crisis Mappers declared that the technique should be described as ‘the process of Leveraging mobile platforms, computational and statistical models, geospatial technologies, and visual analytics to power effective early warning for rapid response to complex humanitarian emergencies.’

There is now a growing number of organisations who are experts in the field of crisis mapping such as Ushahidi, and the Harvard Humanitarian Initiative. The Kenya-based non-governmental organisation Ushahidi, established following the post-election violence in Kenya in 2008 when the national media did not provide regular updates on the events, ‘...this communication gap was filled by some 800 bloggers who communicated the unfolding violence via blog posts, pictures, YouTube videos and Twitter updates. Many of these bloggers took to the streets to observe the developments with their own eyes’ (Meier & Leaning, 2009, p.6).

Ushahidi, as a result, created a platform where people could send their digital content and provide information for others. This data ‘mashup’ would then be visualised on an interactive mapping platform such as Google Earth or Microsoft’s Virtual Earth.
There are signs however that crisis mapping is evolving. Automated crisis mapping, which involves the automated aggregation of digital content using natural language processing technologies, is being used to monitor the development of humanitarian disasters.

Finally, news curation refers to users who determine the importance of a news article in relation to other articles. This function is being increasingly performed by online users who vote on the popularity of a particular article on sites such as Reddit and Digg.

Crisis mapping and news curation demonstrate the close interaction between real-time social media and the social and cultural processes driving the production of that content. While it remains enticing for the researcher to aggregate and analyse such content, it is also deeply important to understand the ethical issues that surround the practice of RTSMM. The next section of this essay discusses some of the most salient issues in RTSMM, illustrating how the social and cultural elements of social media that attract such stringent cause for robust ethical guidelines.

Ethics, Privacy and Trust in Online Social Media Monitoring

Privacy and trust in online social media monitoring is an extremely contentious issue, involving a wide range of actors. Online users are today constantly reminded of the need to protect their information from criminal actors by prominent media campaigns lobbying for greater public awareness. Without doubt, the information contained in social media portals
can be aggregated and analysed but, before RTSMM can commence, the question as to whether rigorous ethical standards can be maintained should be tackled.

Before conducting RTSMM, one must consider whether the study conforms to the general ethical standards in Human Subject research. A variety of ethical guidelines have been written for numerous scientific disciplines based on the principle of Human Subject research. These guidelines refer to scientific experiments which include human beings as active participants, including digital research programmes (Association of Internet Researchers, 2002). The Nuremberg Code is one of the earliest set of ethical guidelines for scientific research in general, and was established following the Second World War to protect participants from questionable scientific studies. The first of its ten directives for human experimentation states:

The voluntary consent of the human subject is absolutely essential. This means that the person involved should have legal capacity to give consent; should be so situated as to be able to exercise free power of choice, without the intervention of any element of force, fraud, deceit, duress, over-reaching, or other ulterior form of constraint or coercion; and should have sufficient knowledge and comprehension of the elements of the subject matter involved as to enable him to make an understanding and enlightened decision…The duty and responsibility for ascertaining the quality of the consent rests upon each individual who initiates, directs or engages in the experiment. It is a personal duty and responsibility which may not be delegated to another with impunity. ¹²

The principles outlined in the Human Subject model and subsequent Nuremberg Code guide the focus of ethical guidelines for other social scientific organisations where human behaviour is the focus of the research process (British Sociological Association, 2003, American Psychological Association, 2008, American Anthropological Association, 2010).

The human-centric nature of social media monitoring requires that any analysis project follows the principles of the Nuremberg Code and other social science research ethics. However, the computer-mediated nature of social media content mandates that a specially crafted set of guidelines outside the Human Subjects model is necessary for ethical digital research. Bassett and O’Riordan (2002) underscored this assertion, claiming that a new ethical framework for digital research should not be based on the current guidelines around the Human Subjects model because it could potentially limit the depth of analysis researchers are able to derive from digital content:

To maintain a research model akin to the human subjects [sic] model would be to risk impeding and potentially eliminating promising research. It is not always possible, for
example, to gain the consent of a large number of participants who may have changed their email address or ceased posting to a Web site on which the material under research is located. (Bassett and O'Riordan, 2002, p.1)

The issue of participant consent is central to any research programme that places the Human Subject model within the guidelines of other social scientific disciplines. The British Sociological Association guidelines for example state that:

As far as possible participation in sociological research should be based on the freely given informed consent of those studied. This implies a responsibility on the sociologist to explain in appropriate detail, and in terms meaningful to participants, what the research is about, who is undertaking and financing it, why it is being undertaken, and how it is to be disseminated and used (Bassett and O'Riordan, 2002).

Bassett and O'Riordan's argument (2002) is particularly compelling when conducting RTSMM. Online users are, in general, hostile to the notion that researchers might use their content for analysis purposes, risking the possibility that the researcher might be unable to use the content. In addition, soliciting permission for the use of social media content is often unfeasible due to the fact that the content creators might not be available. As a result, RTSMM programmes could become difficult to implement as any unauthorised use of social media content could go against the wishes of the original content creator. Commenting on the challenges of collecting data for internet ethnographic projects, the researcher Malin Sveningsson claimed that when looking at online discussion forums,

…users would probably classify us as spammers, get annoyed and treat us the way spammers are generally treated, i.e. filter us out or harass us to leave. As a last resort, they might leave the chatroom themselves. Complicated studies often require that researchers give participants additional information, beyond the informed-consent statement. In practice, many studies are so complex that it is impossible to give participants a full explanation of the research before they participate, without running the risk of skewing the results (Sveningsson, 2003, p.4).

The difficulties of soliciting permission for participant consent can be alleviated. Scholars have argued that conducting research on the understanding that information placed in a public place is considered available for public consumption. Johns, Chen, & Hall (2004) are such scholars who have supported the proposition that publicly available social media content should be used for scientific research:
We view public discourse on Computer Mediated Communication as just that: public-analysis of such content, where individuals’, institutions’ and lists’ identities are shielded, is not subject to human subject restraints. Such study is more akin to the study of tombstone epitaphs, graffiti, or letters to the editor. Personal? Yes. Private? No… (Sudweeks & Rafaeli, 1995, p.1).

Finally, data ownership and the extent to which researchers can aggregate social media content for their own purposes is another issue that should be considered before undertaking RTSMM. Within the user agreements of most social media portals is the condition that user-generated content is owned by the company and third parties cannot gather their content for their own purposes without authorisation.

In addition to the difficulty of gaining consent from content creators, internet researchers face the added problem of being unable to accurately verify the accuracy of the information contained in social media content. Online users actively protect their identities using a range of methods, one of which is alternate identities or noms de guerre:

One of the main subversive ways that users try to protect their social privacy is the use of an alias. Pablo, an [sic] newspaper editor, told me his boyfriend used “Awesome Andrew” as his Facebook name…The goal of this is to make it difficult for people to find them via search, or to attribute their Facebook activities to their “real” identities (Raynes-Goldie, 2010, p.1).

The use of alternative identities has the effect of shielding a person from possible reprisals or any unwanted responses, while allowing them to express themselves more freely than with their personal identities. The ability of researchers to trust social media content rests therefore on a willingness of the original content creator to truthfully disclose their personal data.

The anonymity afforded by the internet highlights one of the main differences between RTSMM and traditional research methods. Participants are generally required to devolve their real identifying information as part of the research (Kozinets, 2010). Social media users however are not bound by this and this, according to Beckmann & Langer, makes the job of RTSMM more difficult: ‘Cyberspace appears to be a dark hallway filled with fugitive egos seeking to entrap the vulnerable neophyte’ (Beckmann & Langer, 2005, p.3).

The potential insight to be gained from the analysis of social media content, verified or not, enshrined within a workable ethical policy however is at once enticing and incalculable. One of the benefits of RTSMM is that the researcher has the opportunity to conduct unobtrusive research without disturbing the research environment, something that is not possible with
other research methods (Kozinets, 2010). Johns et al (2004) underscored this point, claiming that a ‘benefit of virtual research is the extent to which it provides one with the ability to conduct research with virtually no “observer effects.”’...Thus, virtual settings may provide the opportunity for “naturalistic research” in the extreme’ (Johns, Chen & Hall, 2004, p.5).

One of the main advantages of an undisturbed research environment is that the participants are free to express themselves in a more ‘naturalistic’ way; ‘...in which features such as one’s age, gender, ethnicity, and aesthetic appearance do not dominate social interaction...In such a setting, people are more likely to respond to the content of other’s interaction rather than their appearance or personality’ (Johns, Chen & Hall, 2004, p.6).

The issues presented above relating to ethics, privacy and consent in social media monitoring are some of the most prominent that the contemporary RTSMM have to consider. However, while it is possible within research programmes that utilise traditional methodologies to request the consent of participants, it is impractical and at times impossible in the case of RTSMM.

It appears that the increasing volume of user generated content requires that RTSMM develops a wholly new set of guidelines based on the principles of the Human Subjects model and other ideas that can accommodate the complexity of the social and cultural content being analysed. Furthermore, these guidelines have to be able to protect the privacy of those people who produce the content in a way that does not compromise the efficacy of the research.

Real-time Social Media Monitoring Methodologies
The methodologies used to conduct RTSMM are intrinsically important to the completion of any successful RTSMM research process. This section will discuss the uses of methodologies used to monitor social media content in real-time. It will also discuss methodologies used to reflexively analyse social media content that occurs in real-time.

Netnography, or ethnographic studies mediated by the internet, is a technique used to analyse the behaviour and composition of online communities. Developed in 1995 by Professor Robert Kozinets (2002), studies are conducted of online communities and cultures utilising publicly available information to identify the needs and desires of a particular online community. Netnography is distinct as a research approach as it ‘is capable of being conducted in a manner that is entirely unobtrusive. Compared to focus groups and personal interviews, ‘Netnography’ is far less obtrusive, conducted using observations of consumers in a context that is not fabricated by the marketing researcher’ (Kozinets, 2002, p.3). The methodology, described by Kozinets (2010) as ‘...a new qualitative research methodology that
adapts ethnographic research techniques to the study of cultures and communities emerging through computer-mediated communications’, analyses online content and discourse as real examples of social interaction, an embedded expression of meaning and cultural artefact (Kozinets, 2010, p.6).

Netnography is widely used in online community research from consumer insight analysis in the commercial market research industry, to academic research. According to Kozinets (2010), Netnography’s popularity amongst researchers is apparent because it is ‘...naturalistic, following social expression to its online appearances. It is immersive, drawing the researcher into an engaged, deeper understanding. It is descriptive, seeking to convey the rich reality of contemporary consumers’ lives, with all of their hidden cultural meanings as well as their colorful [sic] graphics, drawings, symbols, sounds, photos, and videos. It is multi-method, combining well with other methods, both online and off, such as interviews and videography’ (Kozinets, 2010, p.7 [author’s emphases]).

Kozinets’s description of Netnography’s scope to deeply interrogate the various social and cultural interactions from social media or computer mediated conversations represents one way in which RTSMM can be conducted to accurately analyse the underlying complexities of social media. Kozinets also underscores how an ethnographic approach will uncover the behaviour of online communities. This approach therefore implies that it is important to ascertain not only what people say online but why they might say it.

The internet scholar Annette Markham also discusses the use of other qualitative methods to conduct RTSMM such as email interviews, instant message interviews and interviewing via video conferencing. However, these approaches are not without their problems. Email interviews and instant messaging interviews allow the participant to retain a degree of anonymity, although, as Markham comments, this “may be inadequate for certain participants or research questions”. (Markham, 2011 p. 115) She adds that some participants while preferring to conduct an interview via video conferencing might provide more information if they were able to email or instant message the researcher during the process of the interview. Content analysis is another approach that is used for RTSMM. This quantitative approach is different to the qualitative method described previously. It uses a range of procedures for the systematic quantitative analysis of text, pictorial verbal and symbolic communications data (Krippendorff, 1980) and is used in communications research, particularly for ‘character portrayals in TV commercials, films and novels, the computer-driven investigation of word usage…and so much more’ (Neuendorf, 2002, p.1).

A wide range of definitions of content analysis exist, posited by numerous experts. The influential US behavioural scientist, Bernhard Berelson (1952) described content analysis as a
research technique for the objective, systematic and quantitative description of the manifest content of communication. Riffe, Lacy, & Fico (1998) define content analysis as ‘...the systemic and replicable examination of symbols of communication, which have been assigned numeric values using statistical methods, in order to describe the communication, draw inferences about its meaning, or infer from the communication to its context, both of production and consumption’ (Riffe, Lacy & Fico 1998, p.33). Content analysis is used frequently for RTSMM in order to capture the frequency of certain words, phrases and images. In this age of burgeoning online data, content analysis is a useful process to gather a quick approximation of the saliency of social media content.

In addition to the above techniques, specialist software programs automate the content aggregation and some elements of the data reporting processes, such as content analysis, which can be used in tandem with human researchers.

Real-time social media has evolved from a range of different internet technologies. However, the term ‘real-time’ in the context of social media is fluid, and open to misinterpretation. Social media services provide their users with the tools to create, publish, and consume information with ever-increasing speed and ease. Since their inception in the late 1970s, the time taken to create, publish and analyse information online has progressively shrunk. This has led to the technology being described as ‘real-time’, or as online interactions that ‘...[happen] Without Waiting’ (Winer, 2002). As a result, real-time social media is:

...based on pushing information to users as soon as it’s available - instead of requiring that they or their software check a source periodically for updates. It can be enabled in many different ways and can require a different technical architecture. It’s being implemented in social networking, search, news and elsewhere - making those experiences more like Instant Messaging and facilitating unpredictable innovations. (Kirkpatrick, 2010).

Online research methods that have been specifically modified for mobile devices represent a new frontier. Mobile device is an umbrella term used to describe a range of web-enabled mobile devices such as tablet computers, mobile phones and laptops. Mobile device penetration has now superseded that of personal computers, televisions and even the Web13. With experts predicting that people will increasingly connect to the Internet via their mobile devices, the opportunity to conduct research either using mobiles or social media optimised for such devices is significant. Indeed, people in developing countries often have greater access to mobile phones than they do to landline telephones or televisions (Warah, 2009).
There is a range of mobile research methods applicable to the monitoring of social media in real-time, one of which is location-based research. The accumulation of real-time location data by such techniques as GPS tracking has been used extensively by social media organisations to promote services such as crisis mapping (Coyle & Meier, 2010) and social networking. However, researchers must also be mindful of unwittingly encroaching on the users’ privacy when accessing a person’s specific location:

… privacy is an essential issue...and the subject is often addressed in terms of how sensitive information is kept secured in the application...Identity has several aspects to it and we consider a person’s position to be a specific attribute of identity, like full name and social security number. The major difference between location and most other attributes is that location changes continually and is mostly relevant to mobile computing (Barkus & Dey, 2003, p.3).

The market research industry also uses the combination of mobile research methods and RTSMM to analyse consumer behaviour. An example of this is seen in instances when participants provide information revealing their consumption behaviour to research teams in real-time using their mobile device. It is claimed that information delivered in this way can be used to instantaneously alter communications strategies (Think Tank Research, 2009).

In addition, the increasing use of mobile devices by the global population has led to a greater use of mobile phone surveys in studies designed to be more representative of the general population, and supplement or even to replace traditional telephone surveys. There are, however, a number of methodological, ethical and technical problems associated with the use of mobile telephones as a means of conducting survey research over traditional fixed line telephones.

At present, 77 percent of the adult UK population uses mobile devices, representing a figure of 38.3 million people (UK Office for National Statistics, 2010). On the surface, this figure suggests that mobile devices are an ideal medium through which to conduct survey research. However, on closer inspection, mobile devices seem to be more suited to national surveys as the absence of a telephone directory ‘...forces mobile phone number samples to be created by randomly generating numbers, which implies the risk of generating many numbers not attributed’ (Vicente, Reis & Santos, 2009, p.3). In addition, it is impossible to indicate the geographic location of a user from their mobile telephone number, a problem not associated with fixed line telephones.

One of the principal ethical considerations that must be considered for RTSMM using mobile devices is that mobile device owners are not always adults, in sharp contrast to the owners of
fixed line telephones. Finding eligible participants therefore is problematic for mobile-only surveys as opposed to dual-frame interviews.

This essay will now apply some of the issues highlighted thus far in the context of two case studies from the BBC’s ‘Have Your Say’ discussion forums. It will consider the ways in which the content is aggregated as well as the ethical and privacy issues included in the analysis of this content. The discussion threads will then consider the extent to which the BBC can host a global conversation on the site and discuss the public diplomacy benefits of doing so.

**Case Studies**

The following case studies focus on discussions from articles on the BBC Mundo website referencing Cuban President, Raul Castro’s reshuffle of Government in 2008\(^{14}\), and the widespread social problems faced by Mexico from the proliferation of narcotics\(^{15}\). These articles were selected as they are two of the most active topics on the BBC’s ‘Have Your Say’ page from 3–9 March 2009.

An online analysis of these discussion threads was conducted to demonstrate how ethical and privacy issues contribute greatly to RTSMM. In addition, this essay highlights the ways in which the BBC World Service can better host global conversations using real-time social media, as well as analyse the public diplomacy benefits of such conversations. Global conversations, for the purpose of this essay, are dialogues between people from different geographic locations and cultural backgrounds on a specific topic.

It has been mentioned already that analysis of content hosted on discussion forums is fraught with ethical issues that do not occur in traditional research methods. When attempting to seek permission from users on the use of the content of their posts for research, it was impossible to do so, particularly from those who had commented in 2008. This was due to the fact that commentators had created and published their post and subsequently left the discussion without leaving any contact details. It is therefore apparent that:

> The status of real-time chatrooms is ambiguous. On the one hand, one can argue that they are like a public square. It is considered ethical to record activities in a public place without consent...On the other hand, one can argue that chatroom conversations are normally ephemeral... (Ess, Association of Internet Researchers, 2002).

The inability to first solicit permission from users is problematic if RTSMM of this kind is to adhere to the dictates set out in the Human Subjects model. Such a situation might be avoided however if the user posted their comments on the understanding that it was going to be made available for public consumption.
A total of 100 posts were analysed from the thread on the Cuban government article, and 100 posts for discussions on Mexico and narcotics. An analysis of the sentiment shows that the vast majority (88%) commented negatively, in comparison to the nine percent of posters who commented positively.

The sentiment was measured in order to gauge the polarity of the content emanating from online discussion. The measurement of online sentiment is in many ways problematic because of its inherent subjectivity. What one researcher considers strongly negative could be interpreted as slightly negative or even neutral to another. As a result of this, a five point sentiment measurement scale was created that succinctly reflected the polarity of feelings online. The scale was comprised of strongly negative, slightly negative, neutral, slightly positive and strongly positive measures. In order to limit the propensity for bias, each post was analysed in individual context units. Each context unit, usually a paragraph, was isolated and analysed individually before being assigned a sentiment rating. The totality of the sentiment ratings for a particular post was then tallied up at the end of the analysis to provide a final sentiment score.

Figure 1: Sentiment Analysis of BBC Mundo Content

Verification of the identity of the forum participants was facilitated by the fact that only two participants failed to provide any identifying information about themselves, indicating that participants desired their comments to be attributed to them personally. Indeed, a content
analysis of the participants’ motivations to comment shows that their main intention was to declare their points of view, rather than to engage in deliberative discussion with other users.

Figure 2: Content Analysis of User Motivation for Commentary on BBC Mundo

![Volume of Comments](image)

Source: Lawrence Ampofo

Figure 2 demonstrates that online users were more declarative in their posts than conversational\(^6\) when referencing the reshuffle of the Cuban Government. The need to express one’s self in online discussion forums is a trait regularly witnessed in digital conversations:

In addition to revealing details about themselves, participants may also use online forums as advertising platforms for content they have created apart from the forum itself (i.e. other than standard text posts), be it visual, aural or textual (Freelon, 2010).

It was however clear that the initial articles on the issues of narcotics in Mexico and the reorganisation of the Cuban government attracted an active, multinational audience. Like-minded participants also engaged in community-type behaviour, rounding on others who would express particularly disparaging views. In such instances, participants who expressed viewpoints that were not in keeping with the general sentiment of the rest of the thread were subject to flaming and asked to leave the discussion.
Accusations were however rife amongst participants in both threads that the BBC was not acting sufficiently to foster a constructive discussion amongst the participants. Rather it was accused of creating and publishing inflammatory articles that incited unfocused discussion and ill-feeling. Comments from the participants suggested that if the BBC continued to write inflammatory articles, it would attract the wrong type of users to its web pages. The BBC would therefore be recommended to take a more active approach in nurturing online discussions as this could result in greater participation by more people wishing to engage in a global conversation in a safe environment.

The BBC could do more to promote a global conversation by expanding online discussions beyond its current linguistic borders. While the issues of the reorganisation of the Cuban government and narcotics in Mexico have global relevance, the comments on the articles were made by a Spanish-speaking audience. More could be done to integrate other foreign language comments using translation software so they can feel as though they are participating in a truly global discussion.

In addition, it is unclear from an examination of the posts whether the commentators used either mobile devices or personal computers to make their comments. The inclusion of indicators demonstrating the origin of online commentary enables the BBC to engage greater numbers of people by identifying the media types in various countries, while helping researchers derive deeper insights from the differences between the two types of media platform.

There are extensive public diplomacy benefits of facilitating a truly global conversation that traverses language and geographic borders. As seen in the analyses above, forum posters invest considerable emotion and energy in their contributions. Discussions of this kind send a powerful message to online users that the BBC encourages free speech and tackles controversy. Encouraging online users to participate in discussions of this kind helps the BBC develop strong long-term relationships with its target audiences in key non-English speaking countries and, in particular, states where internet content is controlled. The opportunity to participate in a discussion forum helps the BBC cultivate a deeper one-to-one conversation with people in its network, rather than the traditional one-to-many model, often associated with other traditional broadcasters. This echoes remarks made by a staff member for the BBC World Service at the Wilton Park conference on Public Diplomacy, that the organisation is attempting to change from a “London-knows-best” model to an “end user–knows-best” paradigm... The international debates held on the BBC World Service website underscore the ability of the Corporation to host such discussion in a way that Facebook does not ... [they] demonstrate the need for a protected space where people can debate with each other. The audience creates the rules of engagement in the forums, meaning that people who break the
forum rules are not in breach of the guidelines set by the BBC but those set by the community’ (Ampofo, 2010, p.11).

Conclusion
Social media have undergone profound change since they were first developed in the 1970s. Social media organisations now incorporate a range of real-time features as standard offerings to their users. In tandem with the abundance of real-time social media services, the number of social media monitoring tools and analysis methodologies has also grown, allowing researchers to interrogate saliency of the wealth of user-generated content available. The increasing popularity of real-time social media in contemporary society has simultaneously fostered a manifold increase in the volume of user generated content. As a result, this essay has conceptualised real-time social media as artefacts of social and cultural interactions between its users, in opposition to the notion that it is purely a technological construct. This conceptualisation presents researchers with a dilemma of the way in which such content should be effectively and ethically analysed; whether the focus for effective guidelines on internet research should come from traditional research methods, or if entirely new guidelines should be adopted.

The application of ethical guidelines to RTSMM that are consistently employed in traditional research methods is problematic. In addition, employing research methods suitable for the task of accurately analysing such complex social and cultural interaction is equally uncertain. An examination of two discussions on BBC Mundo underscored the profound opportunity to forge genuine, effective global discussions. However, it also illustrated the difficulty in analysing elements such as sentiment and gaining informed consent from participants within the research environment. In order to effectively accurately assess the impact of such conversations, it is proposed that a new set of internet research guidelines be created that fuse elements from both traditional and non-traditional research methods. In addition, RTSMM would benefit greatly from the application of research methods that account for and accurately demonstrate the saliency and complexity of the wealth of social interactions that take place on real-time social media.

Biographical Note
Lawrence Ampofo is Director, Semantica Research, and Doctoral Research Student in the International Relations Communication Unit, Royal Holloway, University of London.
References


Bruckman, A, (2002). ‘Ethical Guidelines for Research Online’ Georgia Institute of Technology 
College of Computing.

<http://www.bbc.co.uk/blogs/technology/2009/08/mapping_the_afghan_elections.html> 
[Accessed 18 July 2010].


Indianapolis, Indiana.


Coyle, D., & Meier, P., (2009). The Role of Information and Social Networks. UN Foundation 
and Vodafone Foundation Partnership.

Crisis Mappers Net, (Date Unknown). The Crisis Mappers Network. [online]. 

Research Department, Federal Reserve Bank of Philadelphia.

Directives for Human Experimentation, (1949). The Nuremburg Code, Office of Human 
13 August 2010].

Ess, C.M., Association of Internet Researchers, (2002). ‘Ethical Decision-making and Internet 
Research. Recommendations from the AOIR Ethics Working Committee’, AOIR Ethics 

Communities’, BMJ; 323: 1103-1105.

Freelon, D., (2010). ‘Analyzing online political discussion using three models of democratic 


1 'Internet' is often spelled with a capital ‘I’ In keeping with other works in contemporary internet studies and for the purpose of this article, the author will spell with the lower case ‘i.’ Capitalising suggests that the ‘internet’ is a proper noun and implying that either it is a being, or a place. Both metaphors suggest that the internet has agency and power that are better granted to its users and developers.


<http://newsforums.bbc.co.uk/ws/es/thread.jspa?sortBy=1&forumID=7197&start=0&tstart=0#aginator> [Accessed 04 August 2010]

16 For the purpose of this essay, the term 'Responding' refers to participants who engaged in conversation with others on the topic in question. Declarative posts refer to the type of statement in which participants chose to state their position on a given topic and not participate in the discussion with other users.