SOCIAL MEDIA IN QUALITATIVE RESEARCH: CHALLENGES AND RECOMMENDATIONS

Abstract

The emergence of social media on the Internet provides an opportunity for information systems researchers to examine new phenomena in new ways. However, for various reasons qualitative researchers in IS have not fully embraced this opportunity. This paper looks at the potential use of social media in qualitative research in information systems. It discusses some of the challenges of using social media and suggests how qualitative IS researchers can design their studies to capitalize on social media data. After discussing an illustrative qualitative study, the paper makes recommendations for the use of social media in qualitative research in IS.

1. Introduction

The emergence of social media on the Internet provides qualitative researchers with a new window into people's outer and inner worlds, their experiences and their interpretation of these. There is literally a flood of qualitative data pouring into the Internet every day on Twitter, Facebook, LinkedIn, blogs, wikis and so forth, all of which can be downloaded, interpreted, and analysed by the qualitative researcher. At the moment our quantitative colleagues are making good use of this flood of data, for

example, by using big data analytics to analyse such things as the statistical relationships between the users of Twitter and their information sharing behaviour (Shi, Rui, & Whinston, 2014). By contrast, qualitative studies in IS using social media data are few and far between (Müller, Junglas, vom Brocke, & Debortoli, 2016). This suggests to us that we as qualitative researchers in IS have a tremendous opportunity to use social media in order to provide additional insights to those provided by our quantitative colleagues. This is especially so given that 90% of all digital content on the Internet is estimated to be unstructured data (Vijayan, 2015), with most of this of a qualitative nature.

1.1 Motivation

We began this research project with a hunch that quantitative researchers in information systems are making good use of social media data, but qualitative researchers are not. Therefore, to confirm this hunch, we conducted a literature search for social media articles within ten highly ranked information systems journals from 2009 to 2015. These journals were from the AIS Senior Scholars basket of eight top journals namely, *European Journal of Information Systems, Information Systems Journal, Information Systems Research, Journal of AIS, Journal of Information Technology, Journal of MIS, Journal of Strategic Information Systems* and *MIS Quarterly*. To this list, we added two journals where qualitative research is particularly welcome i.e. *Information & Organization* and *Information Technology & People.*

One of the first things we discovered is that authors use many alternative terminologies for social media. For example, Vaast, Davidson, and Mattson (2013) use the term "new media of the Internet", while Ameripour et al. (2010) use the term "Internet social networks." Therefore, we searched using a variety of terms such as social media, microblog, wiki, enterprise 2.0, online social network, online community, web 2.0, and blog (Wang, Min, & Liu, 2014).

Our hunch about the use of social media data by IS researchers was confirmed. We found that the vast majority of articles use quantitative research methods, with only a small number using a qualitative methodology of some kind (sometimes as part of a mixed methods study). We performed the same search queries mentioned above in journals from other business disciplines, and found the same pattern: in disciplines such as marketing, management, human resources, and international business, most studies using social media data are quantitative, not qualitative.

We also discovered a difference in the type of data used by quantitative and qualitative IS researchers. Most quantitative researchers used data directly extracted from social media platforms. The types of quantitative data varied but included data items such as message counts, messages downloaded, friend counts, number of posts, or level of participation. Only a few quantitative papers used data from online surveys. By contrast, most qualitative papers on social media did not use qualitative data extracted directly from social media platforms. The most common data collection method was interviews of social media users. Only a small number of papers used qualitative data directly gathered from social media platforms (e.g. Ameripour, Nicholson, & Newman, 2010;

Germonprez & Hovorka, 2013; Payton & Kvasny, 2012; Vaast et al., 2013; Vaast & Levina, 2015).

Our review of the IS research literature thus suggests a lost opportunity. Quantitative researchers in IS are making good use of qualitative data from social media platforms, but qualitative researchers are not. Only a few qualitative researchers in IS are currently utilizing the vast amounts of qualitative data that are available from social media sites. Our survey of the research literature is consistent with the findings of Müller et al. (2016).

We believe one probable reason for this state of affairs is that there are few qualitative research methods papers about the use of social media and big data in IS. Hence our motivation in writing this paper – we want to encourage qualitative researchers in IS to start using this rich and potentially interesting source of data. Therefore, the purpose of this paper is to suggest how qualitative researchers in information systems can use social media data. Although the value of qualitative social media data has been addressed for specific purposes such as supplementing quantitative social networking studies (e.g. Whelan, Teigland, Vaast, & Butler, 2016), as far as we are aware, this is one of the first qualitative research methods contributions to the IS research literature about the use of social media. We discuss how qualitative IS researchers can design their studies to capitalize on social media and discuss some of the challenges of using social media data. This paper should be of interest to PhD students, to supervisors who are unsure about the conduct of research on social media and its implications, and to

researchers who are interested in combining traditional qualitative techniques with social media studies.

The paper is organised as follows. In Section 2 we discuss the nature of social media and how it might be possible to design IS research studies to capitalize on social media. In Section 3 we discuss some of the opportunities and challenges of using social media in IS research. In Section 4 we provide an illustrative example of using social media in IS. Section 5 makes some recommendations for using social media data. The final section is the discussion and conclusions.

2. Designing IS Research Studies to Capitalize on Social Media

As qualitative IS researchers, we need to figure out a way to design IS research studies to capitalize on social media data. What are the socio-technical boundaries of interest? Should we study the use of social media within an organization, between organizations, by a virtual community resident on a platform, or by a virtual community that spans platforms? These and many other scenarios are possible. Before we decide upon the boundaries, however, we first need to discuss the nature of the phenomenon we are intending to study.

Social media are computer-based tools (such as websites and apps) that enable people to create and share content with other people and/or participate in a community. Bradley (2010), from the Gartner group, says that at their foundation all types of social media are a set of technologies that can construct and enable a potentially large community of

participants to collaborate. Whereas IT tools to support collaboration have existed for decades, new social-media technologies enable collaboration on a much grander scale (Bradley, 2010).

The term “social media” is usually related to web 2.0 applications such as blogs, social networking sites, or video/image/file sharing platforms, and wikis (Fuchs, 2013). Kaplan and Haenlein (2010) define social media as "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" (p.61). This definition introduces us to two key concepts common to social media: technology and content. The authors see social media as dependent on mobile and web-based technologies. These create highly interactive media through which individuals and communities share, co-create, discuss, and modify user-generated content.

Shirky (2011, p. 20) says that social media are tools that “increase our ability to share, to co-operate, with one another”. Boyd (2009) claims that social media is a collection of software allowing people to collaborate, play, share, and communicate and is characterized by user generated content. Lovink (2011) says that social media facilitates social interactions, while Meikle and Young (2012) add that social media includes the creation of a profile, contacts, and the interaction between those contacts. Most researchers seem to agree or at least imply that social media can introduce major and pervasive changes to communications between individuals, organizations, and communities.

Using a set of theories related to social presence, media richness, self-presentation and self-disclosure, Kaplan and Haenlein (2010) suggest a classification scheme for social media (p.61). They categorize seven different types of social media as follows:

1. collaborative projects (for example, Wikipedia)
2. blogs and microblogs (for example, Twitter)
3. social news networking sites (for example, Digg)
4. content communities (for example, YouTube)
5. social networking sites (for example, Facebook)
6. virtual game-worlds (for example, *World of Warcraft*)
7. virtual social worlds (for example, *Second Life*)

However, while this categorization is useful, the boundaries between the different types are becoming increasingly blurred. For example, Shi et al. (2014) argue that Twitter is a combination of broadcasting service and social network and classify it as a "social broadcasting technology" (p.126). This means it would fit into three of the seven types of social media as identified by Kaplan and Haenlein (2010) – types 2, 3 and 5. New media and new apps are continually appearing which offer new functionality or combine the various types of social media in new ways.

Cohen (2011) defines social media as “platforms that enable the interactive web by engaging users to participate in, comment on and create content as a means of communicating with their social graph, other users and the public.” Acknowledging some of the aspects mentioned above, she adds a few additional characteristics of social media. She says that social media involves different levels of engagement by participants who can create, comment or lurk on social media networks; provides for one-to-one, one-to-many and many-to-many communications; enables communication to take place in real time or asynchronously; and extends engagement by creating real-time online events, extending online interactions offline, or augmenting live events

online. We find her emphasis on the interaction types interesting, especially her distinctions regarding communication directionality (i.e. one-to-one, one-to-many and many-to-many communications) and time (i.e. the distinction between synchronous and asynchronous communications).

As IS researchers, these various characteristics of social media imply that there are several ways in which we can design our IS research studies. For example, one dimension relates to the depth of involvement. Walsham (1995) considers two extremes in relation to the roles of the researcher in qualitative research: that of the *outside observer* and that of the *involved researcher*. In qualitative social media studies, the outside observer role could take the form of a passive observer on a social media platform. This could involve the use of web scraping software to extract user-generated data from social media platforms. An example of the outside observer role is provided by Vaast et al. (2013). They studied the discursive practices of a group of tech bloggers and obtained their data (blog posts) from an aggregator website. The *involved researcher* role is one where the researcher could become a member of the social network or organization and hence enable the researcher to have closer access to data which may be confidential or sensitive. In this case, the researcher might become an active contributor or participant in the social network. One example of this is our own research project discussed in the illustrative example below. A useful review of ethnographic studies of social media is provided by Coleman (2010). Walsham (2006) says that the role of the researcher is usually more of a spectrum, where the researcher's role lies somewhere in between the two extremes of complete outside observer and complete involved researcher, and this role may change over time.

Clearly, there are several ways to design qualitative studies of social media. A researcher has to consider their own role along with the topic and particular type of social media that they choose to study. However, any research design needs to take into account the various challenges that present themselves when using social media. We discuss these in the next section.

3. Challenges of Using Social Media in Qualitative IS Studies

Although qualitative research of any kind is challenging – whether it is of a traditional nature or a relatively new approach – there are additional challenges when using social media data (Hanna, 2012; Hunt & McHale, 2007; Jowett, Peel, & Shaw, 2011). We describe some of these challenges below.

3.1 Volume of data

The most obvious challenge in using qualitative data from social media platforms is the sheer volume of data. Although qualitative researchers tend to gather large amounts of data anyway (Myers, 2013), the size of the social media data set might be daunting, even for experienced researchers. For example, there are on average 500 million tweets per day on Twitter, amounting to around 200 billion tweets per year (Internet Live Stats, 2016). In one IS study, the authors collected a total of 1,915,429 tweets from 50,778 Twitter users on their chosen topic in just two months (Oh, Eom, & Rao, 2015). Electronic data management tools such as Nvivo or Atlas/ti are obviously necessary to

store, categorise and manage the data, but these tools in themselves are not sufficient. Since qualitative researchers tend to study a particular topic in depth with a focus on the context, they need to find some way of filtering or "cleaning" the data such that irrelevant data are ignored, while the richness of the story is revealed.

3.2 Digital texts

Another challenge relates to the type of data on social media platforms, which often contain new types of digital texts (Urquhart & Vaast, 2012). Examples are emails, chat threads, images, wikis, avatars, YouTube clips, microblog posts, and emoticons/emojis. Diaz Andrade, Urquhart, and Arthanari (2015) advocate the use of images in IS research, not just as contextual information to other data sources, but also as a source of information in their own right. Of course, most social media platforms contain images of one sort of another. Urquhart and Vaast (2012) suggest that there is a need to theorize these social media-related environments given that they contain a wealth of digital text data.

3.3 Visual cues

In traditional face-to-face interviews, there may be a variety of cues generated by both parties in the social encounter (facial expressions, jokes, encouraging sounds, mannerisms etc.). These can be useful in supplementing the words embodied in the transcripts and may tell us something about the demeanour of the subject and how open they appear to be at the time of the interview (Myers and Newman, 2007). Depending upon the nature of the social media platform, many or all of these cues may be absent or replaced by electronic ones.

3.4 New types of behaviour

Researchers may experience different types of behaviour in social media that are rarely found in face-to-face settings. Behaviour such as 'flaming' (Papacharissi, 2002), 'lurking' or 'whispering' (Garcia, Standlee, Bechkoff, & Cui, 2009) can occur in social media. For example, lurking is where participants in social media adopt passive behaviours: they listen to, observe, and perhaps record the "conversations," but do not engage with the contributors to the social media to any great extent (if at all). How does one study such passive behaviour if it is not visible or obvious?

3.5 Level of access

In traditional face-to-face interviews, the level at which the researcher enters the organization is crucial (Wasko, Teigland, Leidner, & Jarvenpaa, 2011) and will affect the researcher's ability to move around the site. Often a researcher might only talk to managers and other key people, but not employees at the front line. The same issue can apply with social media. For example, permission from a gatekeeper might be required to access certain sections of a site protected by a firewall. Without the required permissions, access to some areas of the site (such as a list of friends/ members) may be denied. In fact, in some social media environments it might not be sufficient to simply obtain permission to enter; rather, the researcher may need to create one or more avatars (Schultze, 2010) or create some other type of online presence. He or she may need to attain to a certain level of skill with their avatar in order to access some areas of the site. However, a researcher might be able to access more subjects using social media than would otherwise be possible in a traditional organizational setting.

3.6 Digital divide

The use of social media might exclude some members of a society who do not have access to or are uncomfortable with their use. In Iran, for example, blogging is very popular but only among a minority of the population, typically younger people or the educated elite (Ameripour et al., 2010). The existence of a digital divide might exclude some people from a study if the researchers use social media only, hence making the use of social media alone nonviable for a particular research project.

3.7 Origin of data

One key difference between the data obtained from interviews and that obtained via social media data is that the researcher has to generate interview data, whereas social media data is self-generated user content. In a traditional qualitative interview, the researcher has more control since they tend to direct the conversation with focussed questions (Myers & Newman, 2007). With user generated data, however, there is less control and less knowledge about the origin of the data, meaning there is potentially much more noise in the data (irrelevant data) which needs filtering. The social media data collected may not contain the specific points the researcher is looking for, or alternatively, there may be questions about the trustworthiness and authenticity of the data.

3.8 Authenticity

Participants may be anonymous or use pseudonyms on social media platforms (Christopher, 2009), which means that it may be difficult to ensure the authenticity of

the data. If we cannot be sure of the identity of the author, can the data be trusted? Some evidence suggests that anonymous use, while offering users a high degree of privacy, at the same time gives users the licence to "misbehave" e.g. by posting inappropriate, offensive or illegal content without fear of punishment (Tsang, Au, Kapadia, & Smith, 2010).

3.9 Ethics

Ackland (2013) says there are three main ethical concerns relating to social media research. The first is informed consent, which is the process of informing participants about the nature of the study so they can freely decide whether to participate or not. Given some of the issues mentioned above, it is simply unrealistic to obtain the consent of everyone. For example, with web-crawling (or scraping) of data from social media sites, the volume of data is huge. Trying to obtain consent from every contributor, let alone verify his or her identity, is infeasible. The consensus of most scholars seems to be that researchers are free to use data available in the public domain (e.g. websites, newsgroups and blogs,). However, researchers should obtain consent when they are conducting research on those sites where is some expectation of privacy. The challenge then becomes, who should one obtain consent from? The issue of informed consent along with other ethical issues is discussed in depth by Thelwall and Stuart (2006).

The second ethical concern that Ackland (2013) mentions relates to the distinction between public and private. The distinction between these two spheres can become blurred in online environments. For example, bloggers may reveal personal information about themselves in a public manner, but with the belief they are only interacting with a

small group of people. They have the perception that they are having a conversation in a private place, with the expectation that others will not use the information. But if the data is actually public, can this data be used? The distinction tends to be clearer in social networking sites because of the use of privacy controls.

The third ethical concern mentioned by Ackland (2013) is participant anonymity. In social media research, it is often not clear when to grant anonymity to participants. For example, in a study of dissident bloggers in Iran, the authors were asked by the editor and reviewers to provide additional evidence to confirm the authenticity of the subjects and validity of their data (Ameripour et al., 2010). However, the authors were unwilling to divulge the identities of their subjects as they feared for their safety (Ameripour et al. 2010). In the end, the authors were able to convince the editors and reviewers of the authenticity and plausibility of their research. Although this is perhaps an extreme example, it demonstrates that considering anonymity along with issues such as authenticity and validity can be a challenge. Light and McGrath (2010) and Zimmer (2010) discuss this and other social media ethical concerns in more detail.

In summary, there are many potential challenges in using social media for research purposes, over and above those that already exist for those doing qualitative research. However, despite these potential problems, we believe that social media holds much promise for qualitative researchers in information systems. We will now discuss some of these issues with an illustrative qualitative study conducted by one of the authors, adding our own experience where appropriate.

4. An Illustrative Qualitative Study: A Social Movement in the World of Warcraft

In this section, we describe one example of a qualitative study that used social media data. Please note that this example is illustrative only. Given the variety of social media studies, this example is not meant to be representative of social media studies in general.

One of the authors explored the co-evolution of a social movement within a virtual world. The virtual world was World of Warcraft (WoW), which is a massively multiplayer online role-playing game created in 2004 by Blizzard Entertainment. In 2010 WoW had over 12 million players globally (Blizzard, 2010). Within this virtual world, people create characters using avatars, meet new people and engage in new forms of social interaction. The social movement studied was the Lesbian, Gay, Bisexual, and Transgender (LGBT) movement, one of the largest social movements in the world and one of the largest within WoW. LGBT aims to create awareness for LGBT issues, both in game and out. By early 2013 LGBT had over 7,800 members (players) in WoW with over 15,000 characters (it is possible for one player to add multiple characters). LGBT was established on a WoW server in October 2006 to "*better service the LGBT community and offer a safe, inclusive place to game for members of any sexual orientation or gender identity*" (LGBT website, 2010). Even though the game of WoW revolves around fighting monsters or avatars from opposing teams, the LGBT group holds many regular activities inside WoW that do not involve fighting. These activities include an annual virtual pride parade with floats, model competitions, dance parties, group photographs, and events for Valentine's Day.

To study this social movement, the first author used a variety of data gathering techniques. Netnography, sometimes called virtual ethnography or online ethnography (Hine, 2000; Kozinets, 2010; Ruhleder, 2000), is a form of ethnography (Harvey & Myers, 1995; Myers, 1999) that involves participation and interaction with community members over the Internet. The author obtained textual data from other sources as well including blog posts, discussion fora and websites. Please note that in the subsequent discussion we are focusing solely on how we conducted the research, in order to illustrate how we addressed some of the challenges in using social media, and not the findings. The findings from the research have been and are being published elsewhere (McKenna, Gardner, & Myers, 2011, 2012).

As our research project progressed we realized that the activities of the social movement within WoW were influenced by patches, which are changes to the software released by the game designers. We followed how these changes to the technical ecosystem influenced the social activities of the LGBT movement within WoW. The purpose of our study was to understand how the technological artefact (the virtual world) and the social movement co-evolve.

The role of the researcher was that of an *involved researcher* (Walsham, 2006). After obtaining permission to conduct the research from LGBT leaders within WoW, the first author immersed himself in WoW. He joined the LGBT movement and participated in a number of movement activities such as virtual pride parades, dance parties, and group photographs. The leaders and many of the members of LGBT were aware of the

presence of the researcher playing the involved researcher role. During the fieldwork, field notes such as digital texts and images were taken, as suggested by Urquhart and Vaast (2012). In addition we found that the discussion forum data was an extremely valuable data source. The discussion forum data was downloaded from the LGBT website with a script to automatically extract the content of each forum post, the thread it belonged to, the entire thread, the dates stamps of each post, and the name of the poster. The researcher saved this dataset into a Microsoft Access database.

In total, the first author spent over 1,600 hours engaging with LGBT. Table 1 lists the multiple sources of data that were collected throughout the research project. The challenges of using this data will now be discussed, although we will not discuss the data obtained from LGBT's website or other websites, as the way we approached the analysis of this textual data was not particularly unique to social media and no different from traditional text analysis approaches.

Table 1. Data sources

Source of Data	Nature of Data Collected	Quantity Collected	Type
Fieldwork	Screen captures of LGBT activities during participant observation.	At least 50 screen images.	Images
WoW Patch Notes	Documentation describing the changes to the software implemented by a patch.	114 patches dating back to 2006.	Text
Discussion Forum Posts	Discussion posts from the LGBT website.	128,773 posts dating back to 2006.	Text
Chat Logs	Chat logs from movement in-game chat channels.	Approximately 1.5 years' worth of chat logs.	Text
LGBT website	Textual information relating to background information about the movement and rules of membership.	Approximately 20 pages.	Text
Other WoW websites	Textual information relating to aspects of WoW gameplay.	Over 100 pages.	Text

4.1. Fieldwork

Given that we were studying an online community, the field notes based on participant observations took a different form than in traditional ethnographic studies. For example, the researcher could not actively take field notes during interactions with members of the social movement, as the mouse and keyboard are needed for other things (e.g. moving an avatar). Therefore, the researcher chose to record what happened with a movie screen cam instead.

Most of the time members of the movement are just simply playing the game, and nothing interesting was happening from a research perspective. Hence, the fieldwork involved the researcher spending many hours just playing the game. On occasion,

however, members of LGBT would meet together and perform social movement activities, such as a parade. As mentioned above, it was at this point that the researcher chose to record movie screen cams of the activities, rather than take notes of the activities. This allowed the researcher to free himself from note taking during this time, and allowed him to continue participating in the activities, for example, "marching" in the parade. Marching in the parade involves controlling an avatar's movements with a keyboard and mouse, which means one's hands are not free for active note taking.

This approach to collecting data during fieldwork meant that images were collected from the participant observation, as well as textual data. The first author took screen shots, as recommended by Kozinets (2010). As Urquhart and Vaast (2012) point out, digital texts can include images or photographs of the virtual world, not the real world. By analyzing these images, we were able to better understand how avatars interact with each other and the virtual objects around them. For example, a video screen cam of the LGBT pride parade and dance party within WoW (created by the researcher), clearly demonstrates that no fighting took place at the time, even though the primary purpose of WoW is ostensibly to fight one's enemies (https://youtu.be/Vfko_sN5z40). Hence, the presentation of images might not only increase the contextual understanding of the reader, but can also provide empirical evidence to support a theoretical point (in this case concerning the interaction between the social movement and the technology).

4.2 Discussion forum posts

As we mentioned earlier, one of the biggest challenges in studying social media is analyzing the large volume of data. This turned out to be the case in our research

project. Given that we collected 128,773 discussion posts, we needed to find a way to filter the data so that we could focus our attention only on those posts considered important for answering our research questions.

In addition to common qualitative data analysis software like NVivo, there are also many software tools available for more advanced text mining, for example QDA Miner and KNIME. We found Leximancer, developed at the University of Queensland, Australia, to be useful for analyzing large amounts of text. Leximancer uses machine learning (content analysis) to analyze large qualitative data sets and to display the results in a visual format. Leximancer has been used in accounting and management (Crofts & Bisman, 2010), conceptual modelling (Davies, Green, Rosemann, Indulska, & Gallo, 2006), human-computer studies (Stockwell, Colomb, Smith, & Wiles, 2009), risk management (Martin & Rice, 2007), and event management (Scott & Smith, 2005). Leximancer has been evaluated for stability and reproducibility and its results so far have been reported to be reliable (Palmer, 2013; Rooney, 2005; Smith & Humphreys, 2006).

Leximancer creates visual output in the form of a conceptual map, which presents the main themes contained within the text, and information about how those themes are related. The themes are heat-mapped to indicate their importance. Therefore, the 'hottest' (most important) themes appear in red and the next most important theme in orange, and so on. Leximancer also allows the researcher to extract the actual pieces of text which were used to create the themes (McKenna, Myers, & Gardner, 2015).

Content Analysis in Leximancer can be supervised or unsupervised. If using the supervised approach, the researcher will construct a set of key terms (known as concepts) usually with some background knowledge within the domain, or with some theoretical sensitivity. Alternatively, in the unsupervised approach, the Leximancer software itself will discover the concepts via reading and re-reading the data. Of course, this approach relies on the algorithms in Leximancer to detect the main themes and concepts arising from the data, but it is the unsupervised approach which is considered by some to be the greatest strength of Leximancer, particularly when there is no prior model or set of factors by which to analyze the data (Davies et al., 2006; Palmer, 2013).

4.3 Chat logs

WoW has a chat feature that enables communication between members of a guild. Since it is possible to save these chat logs, the researcher recorded chat logs of conversations between LGBT members. However, it proved to be difficult to extract useful data from the chat logs. Having collected 1.5 years of chat logs, the data set was massive. Therefore, we used keyword searches to extract useful text. The extracted text was then loaded into Leximancer and analyzed alongside the discussion forum text as detailed in the next section.

4.4 Data Analysis

Data analysis involved the use of two qualitative data analysis software programs, NVivo and Leximancer. Figure 1 depicts this process, with the numbers in the text indicating the process flow in the figure. The first step (1) was to load the entire dataset

into Leximancer. As discussed above, and given that our study was exploratory, the unsupervised approach was executed. This created a set of Leximancer themes. However, the dataset contained far too many posts which were irrelevant to our original research questions. Therefore, we had to find a way to reduce the dataset.

The second step (2) was therefore to reduce the dataset by reading the patch notes of 114 patches in an attempt to discover which patch had the most impact on LGBT (since we were interested in the co-evolution of the social movement along with the virtual world, the patch notes documented any changes to the software made by Blizzard). After analysis of the patches, we discovered three patches which had a strong influence. Often Blizzard releases the patch notes before the patch is implemented into the game. Therefore, we were able to filter the data from the discussion forum by looking at the time stamps and extracting only those posts made about a certain patch immediately before and after the release of a patch. We also performed keyword searches using keywords from our theoretical approaches along with keywords based on our knowledge of the game and our analysis of the patch notes. Therefore, we were able to disregard most of the posts, which gave us a final count of 405 posts which we considered to be useful for answering our main research question. This research question was concerned with the co-evolution of the technological artifact (the virtual world) with the social movement.

The third step (3) was to load this reduced dataset into NVivo for manual coding with theoretical sensitivity to actor-network theory (ANT). Please note that a discussion of

theory choice is beyond the scope of this paper, but is discussed elsewhere (McKenna et al., 2012).

The fourth step (4) was to load the newly coded dataset back into Leximancer. This created a new set of themes and knowledge pathways which were then more manageable.

The fifth step (5) was to load the text which created those knowledge pathways back into NVivo. This text was further coded manually with theoretical sensitivity to ANT. In the sixth step (6) the codes from both rounds of NVivo coding were compared with each other. This process is illustrated in Figure 1.

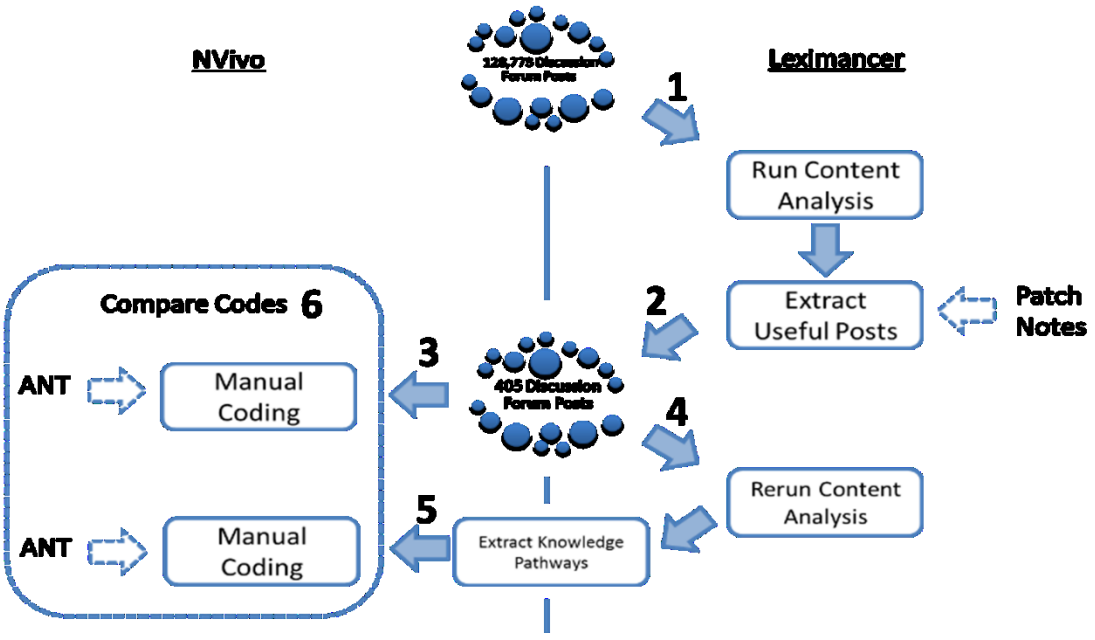

Figure 1. The coding process

An example of the Leximancer output is illustrated in Figure 2. This example shows the results of the text mining analysis of the discussion forum and chat log data where LGBT members discussed a specific patch which had a severe impact on the group. In the example provided, some common themes were extracted. These themes were directly related to when a new patch was released; this patch placed a cap on the size of guilds. On the left hand side is the conceptual map. Each circle within the map represents a theme. We can see from this image that the most important theme was "guild," followed by "members." Each theme contains multiple concepts (nodes) which make up that theme. The solid line indicates the knowledge pathway, which shows the connections between concepts. These pathways were used to empirically link concepts together which provided a useful way of understanding the data and analyzing relationships between concepts. The right hand side indicates the actual text extracted from Leximancer which creates the knowledge pathway, i.e. the text which supports the relationship between the concepts 'guild' and 'Blizzard'. Note that for the sake of simplicity only one knowledge pathway is shown, but a pathway can be created between any concepts in the map.

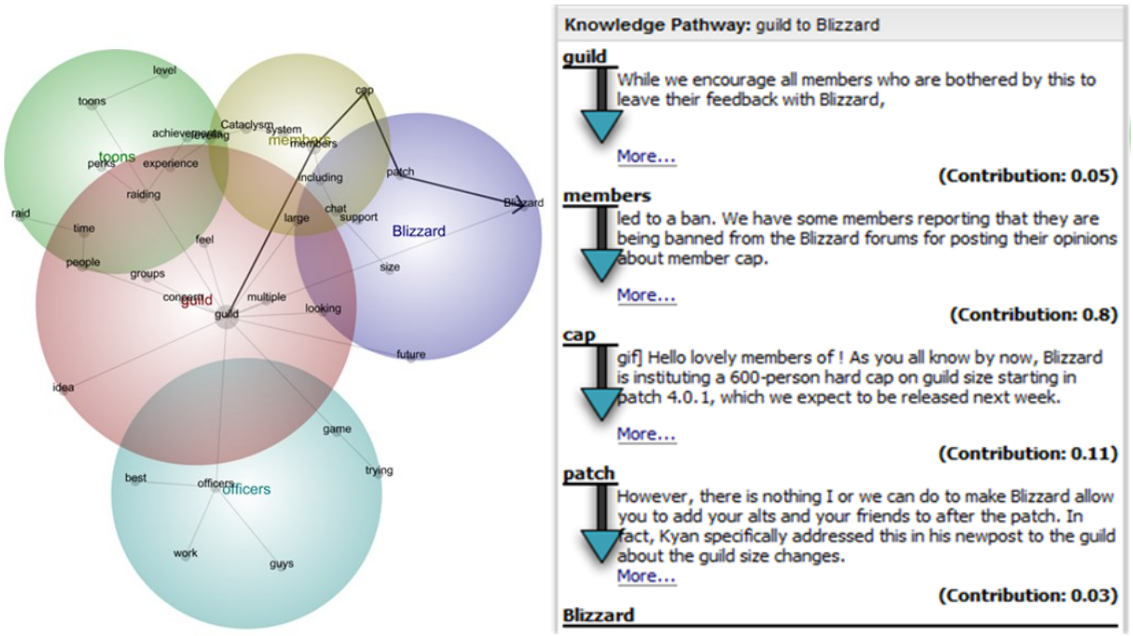

Figure 2. Leximancer output

By reducing the dataset in this way, we ended up with a rich set of direct quotes from the leaders and members of the movement, along with quotes from the software developer (Blizzard). This rich set of qualitative data was then used to create a narrative of the key events (software changes) and the adaptation of the social movement to these changes. We were successful in providing an answer to our research question.

As we said earlier, our WoW example above is meant to be illustrative only since there are many different types of social media along with various ways to study them. However, many of the issues we faced are likely to be found elsewhere. For instance, in a study that looked at the relationship between Internet social networks and societal change in Iran, the researchers collected a large volume of data, most of which was of a digital nature (Ameripour et al., 2010). As in our study, they took advantage of the automatic archiving feature in social media. Instead of the researcher gathering the primary data themselves (e.g. via a tape recorder), the social media platform did it for

them - the conversations, events, and almost anything that happened online was logged and recorded automatically by the system. With user-generated data, the users of the social media platform are in effect recording their own words and actions themselves. For example, if someone posts a comment in a discussion forum, the comments are already typed up and transcribed by the user, they are date and time stamped automatically, and they are immediately available to others who want to read them online.

In both the WoW study and the Iran study, both sets of authors took great pains to establish the authenticity of their data. One of the ways they did this was to use virtual ethnography to study the online behaviour of people on the social media platform. Another way was to use mixed methods and triangulate one form of data with another e.g. by supplementing discussion forum posts with chat logs. Lastly, in both studies the authors were able to negotiate appropriate access. In the Iran study this access was enabled by the first author being fluent in the Farsi language and being familiar with Iranian culture (Ameripour et al., 2010). In the WoW study, access was enabled mostly by the first author becoming adept at playing World of Warcraft – only a player with the required level of expertise is able to access certain restricted areas of the virtual world, areas which were important for participating in and observing the activities of the LGBT movement.

5. Recommendations for using social media data

Based on the lessons learnt from our own research project and our insights from the IS research literature, we now make some recommendations for using social media data in qualitative research in IS. These recommendations are summarised in Table 2.

Table 2 - List of recommendations

Challenge	Challenge	Recommendation
Volume of data	The researcher obtains huge amounts of data which needs to be analyzed (Boellstorff, Nardi, Pearce, & Taylor, 2012). There is likely to be much noise in the data.	1. Use a filtering or data mining technique 2. Use qualitative data analysis software
Digital texts	Social media platforms contain many different kinds of digital texts and images (Urquhart & Vaast, 2012).	3. Images (such as screen shots and videos) may need to be gathered and analyzed (sometimes while actively participating in the online fieldwork)
Visual cues	Visual cues may be in digital form rather than face-to-face	4. Become familiar with and socialized into the world of the social media platform
New types of behavior	People using social media platforms may exhibit different types of behaviour than in face-to-face settings	4. Become familiar with and socialized into the world of the social media platform
Authenticity	Participants may be anonymous or use pseudonyms, potentially raising questions about the authenticity of the data	5. Use mixed methods to triangulate different types of data 6. Develop research questions where the identity of participants is not important
Level of access	The researcher needs to gain access to certain (possibly restricted) areas of the social media platform	7. Obtain permission from the gatekeeper (if needed) 8. If needed, create and use one or more avatars
Digital divide	The use of social media might exclude some people	9. If needed, supplement social media data with traditional data gathering techniques

5.1 Recommendation 1: Use a filtering or data mining technique

Our first recommendation is related to the challenge of collecting a huge *volume of data.* We believe it is essential to use data mining techniques to filter or reduce the vast amount of text. There is some disagreement among researchers about how much data to code (Saldaña, 2009). For example, Lofland, Snow, Anderson, and Lofland (2006) recommend coding the entire dataset, whilst Seidman (2006) argues that only the most important data must be coded. Our recommendation for using social media data in qualitative research is to follow Seidman's approach. With such a large data set, it is simply impossible for a

qualitative researcher to analyse it all, and in any case much of the data might be irrelevant or at least not particularly useful for answering the research question(s). Although we recognize that sometimes this could be like finding a needle in a haystack, we believe it is more practical for the researcher to filter the data and weed out what is not relevant to the study.

For example, in our WoW study we focused our attention only on those discussion forum posts that discussed patches to the software (since these patches were directly related to our research question). Since the patches were implemented at particular times, we were able to filter the discussion forum posts by focusing only on the dates surrounding the patch implementation. Researchers could use other methods, such as filtering for keywords, events, research question specific concepts, or theoretical sensitivity. Payton and Kvasny (2012) organized their blog posting data chronologically to create a timeline of events.

5.2 Recommendation 2: Use qualitative data analysis software

Given the large *volume of data*, it is simply not feasible to analyse this data manually. Hence our second recommendation, closely related to the first, is that it is essential to use a qualitative data analysis (QDA) software package to help in the management and analysis of data. However, since not all QDA software packages have the same features, it is necessary to choose the most appropriate one.

In the WoW study we realized that coding 128,773 discussion forum posts would be time consuming if we used a qualitative data analysis software package like NVivo (auto-

coding is possible in nVivo, but only to a limited extent). Therefore, of the available text mining software, we decided to use Leximancer, which provides automated qualitative data analysis.

Of course, researchers need to be careful when using automated text analysis tools – one needs to be aware of the 'garbage in garbage out' problem, which is especially problematic when using live data such as discussion forums. Discussion forum data contain many threads and posts about an unlimited number of topics, most of which might be irrelevant to the research project. Therefore, our first recommendation needs to be read in conjunction with our second - the researcher may need to filter the data before using the software. In our case, we found that the Leximancer analysis was producing many inappropriate and irrelevant results before filtering. It was also necessary to filter the data once again after the Leximancer analysis. For example, the data from the discussion forums often had HTML tags embedded within them. As these tags are text, they were included in the results. Leximancer contains a pre-defined set of stop words which are skipped over by the algorithm. Since it is possible to edit the stop word list, we added the HTML tags to the list and ran the algorithm again.

This iterative process was completed many times to remove words unnecessary for the analysis. Other words were also added to the stop word list based on the prior knowledge of the researcher. It is claimed that the strength of Leximancer is its unbiased analysis of the data, however we found that the researcher needs to actively intervene in order to produce meaningful results. This requires researchers to have the relevant contextual

knowledge of the subject matter and sufficient familiarity with the dataset *prior to* processing in Leximancer.

When using a qualitative data analysis tool to analyse text, it is important to consider the language that might be used by participants. For example acronyms such as LOL (laugh out loud), ROFL (rolling on the floor laughing), and WDYM (what do you mean?) might be used. If the software tool does not understand these acronyms, some level of training for the algorithm might be necessary. In the WoW study there were many terms such as "toon" or abbreviations such as "LGBT" which were relevant to the game and/or to the LGBT group, but which were not understood automatically by the software.

After completing all the required steps to clean the data, we recommend that researchers should compare the final version produced by the software with codes created in another software tool and/or compare the results with some manually coded text. Assuming the results are reasonably consistent, this is one way of increasing a researcher's confidence in the results provided by the QDA software.

5.3 Recommendation 3: Images (such as screen shots and videos) may need to be gathered and analyzed

Although images have been underutilized in IS research so far (Diaz Andrade et al., 2015), social media platforms contain many different kinds of *digital texts* (Urquhart & Vaast, 2012). Hence our third recommendation is that IS researchers may need to gather and analyse some of these texts. Digital texts can include images, sounds, text, instant

messaging, or videos. Learning how to capture and code digital texts (such as a screenshot or video) may require the researcher to learn new skills.

In the WoW study we decided to capture what was happening in the field by taking screenshots. These screenshots were loaded into NVivo where sections of the image could be coded by the researcher. For example, screenshots taken during the pride parades were coded to record the avatars who were parade participants or parade observers, and to record some other aspects of the interaction. The researcher then triangulated this data with other sources of data.

5.4 Recommendation 4: Become familiar with and socialized into the world of the social media platform

Our fourth recommendation is that the researcher needs to become familiar with and socialized into the world of the social media platform. This is similar to many other forms of qualitative research in which the researcher "immerses himself or herself in the life of the social group under study" (Myers, 1999, p. 4). In the case of social media, this might mean finding out who the most influential people are on the site; it might also mean negotiating with them to obtain access to certain restricted areas. Becoming familiar with the culture of a social media platform involves being able to understand *visual cues*, possibly learning the language of the social media platform (as mentioned earlier), and perhaps discovering *new types of behaviour*.

In our WoW study this meant that the field researcher had to immerse himself in WoW. Since a certain level of expertise is needed in order to access some LGBT events in WoW,

the only way to gain this expertise is to spend time playing the game. Hence, the first author spent over 1,600 hours playing and interacting with members of LGBT within the game. This increased familiarity with the LGBT movement within WoW meant that he built up an intimate knowledge of their activities, which then enabled him to gather data relevant to his research project.

We discovered one rather surprising new type of behaviour during the fieldwork: during the pride parade and a few other LGBT events, fighting was expressly forbidden; any breaking of this rule could result in someone being expelled from membership of the movement. This peaceful behaviour, however, is exactly the opposite of what is normally expected within WoW – the game is explicitly designed to be a game of war. The discovery of this new type of behaviour provided more insight about the co-evolution of the technological artefact and the social movement.

In the Iran study, the first author was already fluent in the Farsi language and familiar with Iranian culture. This familiarity with the culture and the world of the social media platform meant that the participants were willing to share their opinions, thus increasing our confidence in the findings (Ameripour et al., 2010). This leads directly to our next recommendation.

5.5 Recommendation 5: Use mixed methods to triangulate different types of data

Given that participants on social media platforms may be anonymous or use pseudonyms, questions might arise about the *authenticity* of the data. We therefore recommend researchers should consider using mixed methods to triangulate different types of data in

order to increase our confidence in the findings or to provide additional insights. One way to do this is to triangulate different types of data within the social media platform. In the WoW study, we used six different types of data including chat logs, discussion forum posts and participant observation. Each of the data types can be used to confirm or provide a different perspective one the same phenomenon under exploration. For example, in the WoW study we matched the patch notes with text downloaded from the discussion forums and chat logs. Additionally, the researcher could experience the issues relating from the patch directly through his participant observations. Combined, all these data sources provided a richer picture of the situation.

Alternatively, a researcher could triangulate the data obtained from the social media platform with data obtained external to the platform (e.g. web sites or interviews). In the Iran study, the researchers supplemented the social media data with email and telephone interviews.

5.6 Recommendation 6: Develop research questions where the identity of participants is not important

Another way to overcome the problem of *authenticity* (besides triangulation of data) is to develop research questions where the identity of participants is not important. In the WoW study, knowing the identity of the participants was not essential for answering our primary research question. However, the first author did discover the identity of the leaders of the LGBT movement in WoW. This is mostly because our institutional ethics review board required us to obtain permission from the leaders to conduct the study. We can imagine that some other institutions might not require this, particularly in cases where

the identity of participants is irrelevant to the study. Of course, there are many social media sites where the users are not anonymous and hence this recommendation does not apply.

5.7 Recommendation 7. Obtain permission from the gatekeeper (if needed)

Although permission may not be needed on some publicly available social media sites, on others the researcher may need to gain access to certain (possibly restricted) areas of the social media platform. This may mean obtaining permission from one or more gatekeepers in order to gain the necessary *level of access*.

5.8 Recommendation 8: If needed, create and use one or more avatars

In order to gain the right *level of access* on some social media platforms, it might be necessary for the researcher to create and use an avatar (Schultze, 2010). This avatar becomes the researcher's identity when conducting the research. Avatars can be in human form or some other fantasy based form. Researchers might also have to choose an online name (Hagström, 2008). We suggest that a researcher should take into account the values or cultures of the participants when choosing an identity. In some social media platforms, it may be more appropriate to create an account with the researcher's real identity, but in others an assumed one might be fine. In some online worlds, it is possible for a researcher to create and use multiple avatars, which might enable an understanding of the research problem from multiple perspectives.

In our WoW study the first author created and used six characters (avatars) within WoW. Each character had to be enrolled separately into the LGBT group. Using multiple avatars

is considered normal behaviour in WoW and hence would not be seen as unusual by the leaders of the LGBT movement.

5.9 Recommendation 9. If needed, supplement social media data with traditional data gathering techniques

In order to address the potential problem of the *digital divide* in some situations, we recommend that researchers should consider supplementing social media data with traditional data gathering techniques. These traditional techniques include interviews, participant observation and the use of documents. This recommendation is similar to our 5th recommendation, except that in the former case the data that is triangulated need not be data obtained from traditional data sources. Another difference between this recommendation and the earlier one is in relation to its purpose: recommendation 5 is concerned with addressing the challenge of ensuring authenticity, whereas recommendation 9 is concerned with trying to ensure that some members of a society are not excluded (if their participation would be relevant to the study).

6. Discussion and Conclusions

Social media holds much potential for qualitative researchers in IS. The advent of big data on the Internet (most of which is unstructured textual data) means that a potentially valuable new source of qualitative data is now available for analysis. However, as we have seen, few qualitative researchers in information systems are currently utilizing the

vast amounts of qualitative data that are available from social media sites. Quantitative researchers in IS are making good use of qualitative data on social media platforms, but qualitative researchers are not. Hence, this paper has looked at the potential use of social media in qualitative research in information systems, discussed some of the challenges of using social media, and made some recommendations.

The challenges of using social media in qualitative research are many. These challenges are related to the large volume of data, the nature of digital texts, visual cues, and types of behaviour on social media sites, the authenticity of the data, the level of access obtained, and the digital divide in some situations. In an attempt to address these challenges, we have made nine recommendations for conducting qualitative research using social media data. While some of our recommendations are similar to those that might be made for more traditional kinds of research, others are quite different. We can summarize these differences as follows.

First, researchers might have to learn new skills to gather and analyse social media data. Even if a researcher, adopting the *outside observer* role, uses commercially available web scraping software for gathering the data, the volume of data will make it essential to use some filtering technique. To do this a researcher will need learn and use a qualitative data analysis (QDA) software package. However, since not all QDA software packages have the same features, it might be necessary for the researcher to become proficient in more than one.

Second, while it is common practice for fieldworkers to immerse themselves in a field site, learn a new language/ jargon and learn about new types of behaviour, this is likely to be insufficient for conducting research on social media (i.e. when adopting the *involved researcher* role). For example, the only way to gain access to the activities of the social movement in WoW was by the researcher earning sufficient access level "points." The way to earn these points is to play the game and attain a reasonably high level of proficiency. Without a certain level of proficiency, the researcher would not have been able to observe any of the activities of the social movement. Hence, this reinforces the point that conducting research on a social media platform may require a certain set of skills, skills that may take time and commitment to master.

Third, conducting research on social media tends to lead to a much greater use of visual images. Visual images can be useful and can enhance our understanding, but many publishers are unwilling to include colourful images in the print version of the journal paper, given the additional expense. Perhaps researchers using social media can consider creative ways of including images and animations in their paper by, for example, embedding links to video clips on YouTube in their manuscripts or providing an online addendum. The use of images is not such a problem with purely electronic journals, of course, but it can be a problem for print journals.

Fourth, the use of social media introduces new ethical challenges for qualitative researchers. It can be a challenge trying to strike the right balance between ensuring the plausibility, validity and rigor of the research findings, while at the same time adhering to ethical principles such as informed consent.

We acknowledge a few limitations. First, while we believe we have mentioned most of the challenges that qualitative researchers in IS are likely to face when studying social media, we do not claim to have mentioned them all. In fact, we are sure that new challenges will emerge in future. Second, we have discussed only one example of a qualitative study using social media, and hence the challenges and lessons learnt may not necessarily be generalizable to other types of social media. Third, we acknowledge that our recommendations might not apply in every given research project. It is up to each researcher and his or her colleagues to decide which ones apply. We hope that future research methods papers about the use of social media data will address some of these limitations and perhaps explore new opportunities e.g. the potential for collaboration between qualitative and quantitative researchers.

In conclusion, our motivation in writing this paper was simply to encourage qualitative researchers in IS scholars to start using social media data and to suggest how to conduct such research. As social media software continues to evolve, we are sure that new opportunities and challenges for qualitative researchers will emerge.

References

Ackland, R. (2013). *Web social science: Concepts, data and tools for social scientists in the digital age*: Sage.

Ameripour, A., Nicholson, B., & Newman, M. (2010). Conviviality of Internet social networks: An exploratory study of Internet campaigns in Iran. *Journal of Information Technology, 25*(2), 244-257.

Blizzard. (2010). World Of Warcraft® Subscriber Base Reaches 12 Million Worldwide. Retrieved 15 November 2010, from http://us.blizzard.com/en-us/company/press/pressreleases.html?101007

Boellstorff, T., Nardi, B., Pearce, C., & Taylor, T. L. (2012). *Ethnography and Virtual Worlds: A Handbook of Method*. New Jersey: Princeton University Press.

Boyd, D. (2009). *Social media is here to stay... Now what?* Paper presented at the Microsoft Research Tech Fest. Retrieved from http://www.danah.org/papers/talks/MSRTechFest2009.html

Bradley, A. (2010). A new definition of social media. *Gartner blog network* Retrieved April 2014, 2014, from http://blogs.gartner.com/anthony_bradley/2010/01/07/a-new-definition-of-social-media/

Christopher, T. (2009). In-Game Identities and Meatspace Mistakes. In L. Cuddy & J. Nordlinger (Eds.), *World of Warcraft and Philosophy: Wrath of the Philosopher King* (pp. 165-171). Chicago: Open Court Publishing.

Cohen, H. (2011). Social media definitions. *Actionable marketing guide* Retrieved April 2014, 2014, from http://heidicohen.com/social-media-definition/

Coleman, G. (2010). Ethnographic Approaches to Digital Media *Annual Review of Anthropology, 39*, 487-505.

Crofts, K., & Bisman, J. (2010). Interrogating accountability: An illustration of the use of Leximancer software for qualitative data analysis. *Qualitative Research in Accounting & Management, 7*(2), 180-207.

Davies, I., Green, P., Rosemann, M., Indulska, M., & Gallo, S. (2006). How do practitioners use conceptual modeling in practice? *Data and Knowledge Engineering, 58*, 358-380.

Diaz Andrade, A., Urquhart, C., & Arthanari, T. S. (2015). Seeing for Understanding: Unlocking the Potential of Visual Research in Information Systems. *Journal of the Association for Information Systems, 16*(8), 3.

Fuchs, C. (2013). *Social media: A critical introduction*. London: Sage.

Garcia, A., Standlee, A., Bechkoff, J., & Cui, Y. (2009). Ethnographic Approaches to the Internet and Computer-mediated Communication. *Journal of Contemporary Ethnography, 38*, 52-84.

Germonprez, M., & Hovorka, D. S. (2013). Member engagement within digitally enabled social network communities: new methodological considerations. *Information Systems Journal, 23*(6), 525-549.

Hagström, C. (2008). Playing with Names: Gaming and Naming in World of Warcraft. In H. G. Corneliussen & J. W. Rettberg (Eds.), *Digital Culture, Play, and Identity* (pp. 265-285). Cambridge, Massachusetts: The MIT Press.

Hanna, P. (2012). Using internet technologies (such as Skype) as a research medium: a research note. *Qualitative Research, 12*(2), 239-242.

Harvey, L., & Myers, M. D. (1995). Scholarship and practice: the contribution of ethnographic research methods to bridging the gap. *Information Technology & People, 8*(3), 13-27.

Hine, C. (2000). *Virtual Ethnography*. Thousand Oaks, CA: Sage Publications.

Hunt, N., & McHale, S. (2007). A Practical Guide to the e-mail Interview. *Qualitative Health Research, 17*(10), 1415-1421.

Internet Live Stats. (2016). Twitter usage statistics. Retrieved August 8, 2016, from http://www.internetlivestats.com/twitter-statistics/

Jowett, A., Peel, E., & Shaw, R. (2011). On-line interviewing in psychology: reflections on the process. *Qualitative Research in Psychology, 8*, 354-369.
Kaplan, A. M., & Haenlein, M. (2010). Users of the world, unite! The challenges and opportunities of social media. *Business Horizons, 53*(1), 59-68.
Kozinets, R. V. (2010). *Netnography. Doing Ethnographic Research Online*. London: Sage Publications Ltd.
Lofland, J., Snow, D., Anderson, L., & Lofland, L. H. (2006). *Analyzing social settings: a guide to qualitative observation and analysis* (4th ed.). Belmont, CA: Thomson Wadsworth.
Lovink, G. (2011). *Networks without a cause: A critique of social media*. Cambridge: Polity Press.
Martin, N. J., & Rice, J. L. (2007). Profiling Enterprise Risks in Large Computer Companies using the Leximancer Software Tool. *Risk Management, 9*, 188-206,.
McKenna, B., Gardner, L., & Myers, M. D. (2011, August 4-7). *Social Movements in World of Warcraft.* Paper presented at the Americas Conference on Information Systems, Detroit, MI.
McKenna, B., Gardner, L., & Myers, M. D. (2012). *The Co-Evolution of the "Social" and the "Technology": A Netnographic Study of Social Movements in Virtual Worlds.* Paper presented at the International Conference on Information Systems, Orlando, FL.
McKenna, B., Myers, M., & Gardner, L. (2015). *Analysing qualitative data from virtual worlds: using images and text mining.* Paper presented at the European, Mediterranean & Middle Eastern Conference on Information Systems (EMCIS).
Meikle, G., & Young, S. (2012). *Media convergence: Networked digital media in everyday life*. Basingstoke: Palgrave Macmillan.
Müller, O., Junglas, I., vom Brocke, J., & Debortoli, S. (2016). Utilizing Big Data Analytics for Information Systems Research: Challenges, Promises and Guidelines. *European Journal of Information Systems, 25*, 289-302.
Myers, M. D. (1999). Investigating Information Systems with Ethnographic Research. *Communications of the AIS, 2*(23), 1-20.
Myers, M. D. (2013). *Qualitative Research in Business & Management* (2nd ed.). London: Sage Publications.
Myers, M. D., & Newman, M. (2007). The qualitative interview in IS research: Examining the craft. *Information and Organization, 17*(1), 2-26.
Oh, O., Eom, C., & Rao, H. R. (2015). Role of Social Media in Social Change: An Analysis of Collective Sense Making During the 2011 Egypt Revolution. *Information Systems Research, 26*(1), 210–223.
Palmer, P. D. (2013). Exploring attitudes to financial reporting in the Australian not-for-profit sector. *Accounting and Finance, 53*, 217-241.
Papacharissi, Z. (2002). The Virtual Sphere: The internet as the public sphere. *New Media & Society, 4*(1), 5-23.
Payton, F. C., & Kvasny, L. (2012). Considering the political roles of Black talk radio and the Afrosphere in response to the Jena 6: Social media and the blogosphere. *Information Technology & People, 25*(1), 81-102.
Rooney, D. (2005). Knowledge, economy, technology and society: the politics of discourse. *Telematics and Informatics, 22*, 405-422.
Ruhleder, K. (2000). The Virtual Ethnographer: Fieldwork in distributed

electronic environments. *Field Methods, 12*(1), 3-17.
Saldaña, J. (2009). *The coding manual for qualitative researchers*. London: Sage.
Schultze, U. (2010). Embodiment and presence in virtual worlds: a review. *Journal of Information Technology, 25*(4), 434-449.
Scott, N., & Smith, A. E. (2005). Use of automated content analysis techniques for event image assessment. *Tourism Recreation Research, 30*(2), 87-91.
Seidman, I. (2006). *Interviewing as qualitative research: A guide for researchers in education and the social sciences* (3rd ed.). New York: Teachers college press.
Shi, Z., Rui, H., & Whinston, A. B. (2014). Content Sharing in a Social Broadcasting Environment: Evidence from Twitter. *MIS Quarterly, 38*(1), 123-142.
Shirky, C. (2011). The political power of social media: Technology, the public sphere, and political change. *Foreign affairs, 90*(1), 28-41.
Smith, A. E., & Humphreys, M. S. (2006). Evaluation of Unsupervised Semantic Mapping of Natural Language with Leximancer Concept Mapping. *Behavior Research Methods, 38*(2), 262-279.
Stockwell, P., Colomb, R. M., Smith, A. E., & Wiles, J. (2009). Use of an Automatic Content Analysis Tool: a Technique for seeing both Local and Global Scope. *International Journal of Human Computer Studies, 67*(5), 424-436.
Thelwall, M., & Stuart, D. (2006). Web crawling ethics revisited: Cost, privacy, and denial of service. *Journal of the American Society for Information Science and Technology, 57*(13), 1771-1779.
Tsang, P. P., Au, M. H., Kapadia, A., & Smith, S. W. (2010). BLAC: Revoking repeatedly misbehaving anonymous users without relying on TTPs. *ACM Transactions on Information and System Security (TISSEC), 13*(4), 39:31-33.
Urquhart, C., & Vaast, E. (2012). *Building Social Media Theory from Case Studies: A New Frontier for IS Research*. Paper presented at the Thirty Third International Conference on Information Systems (ICIS).
Vaast, E., Davidson, E. J., & Mattson, T. (2013). Talking about Technology: The Emergence of a New Actor Category Through New Media. *Mis Quarterly, 37*(4), 1069-1092.
Vaast, E., & Levina, N. (2015). Speaking as one, but not speaking up: Dealing with new moral taint in an occupational online community. *Information and Organization, 25*(2), 73-98.
Vijayan, J. (2015, June 25, 2015). Solving the Unstructured Data Challenge. *CIO Magazine*, from http://www.cio.com/article/2941015/big-data/solving-the-unstructured-data-challenge.html
Walsham, G. (1995). Interpretive case studies in IS research: nature and method. *European Journal of Information Systems, 4*(2), 74-81.
Walsham, G. (2006). Doing interpretive research. *European Journal of Information Systems, 15*(3), 320-330.
Wang, Y., Min, Q., & Liu, Z. (2014). *A Meta-analytic Review of Social Media studies.* Paper presented at the Proceedings of the Pacific Asia Conference on Information Systems (PACIS), Chengdu, China.
Wasko, M., Teigland, R., Leidner, D., & Jarvenpaa, S. (2011). Stepping into the Internet: New Ventures in Virtual Worlds. *MIS Quarterly, 35*(3), 645-652.
Whelan, E., Teigland, R., Vaast, E., & Butler, B. (2016). Expanding the horizons of digital social networks: Mixing big trace datasets with qualitative approaches. *Information and Organization, 26*(1), 1-12.

Zimmer, M. (2010). “But the data is already public”: on the ethics of research in Facebook. *Ethics and information technology, 12*(4), 313-325.