Manchester Metropolitan University

e-space

Manchester Metropolitan University's Research Repository

Please cite the Published Version

Vaast, E and Urquhart, C (2018) Building Grounded Theory with Social Media Data. In: The Routledge Companion to Qualitative Research in Organization Studies. Routledge Companions in Business and Management . Routledge. ISBN 9781138921948 (hardback); 9781032476643 (paperback)

Publisher: Routledge

Version: Accepted Version

Downloaded from: https://e-space.mmu.ac.uk/618230/

Additional Information: This is an Author Accepted Manuscript of a book chapter provided by Routledge of a paper accepted for publication in Routledge Companion to Qualitative Research in Organization Studies.

Enquiries:
If you have questions about this document, contact openresearch@mmu.ac.uk. Please include the URL of the record in e-space. If you believe that your, or a third party's rights have been compromised through this document please see our Take Down policy (available from https://www.mmu.ac.uk/library/using-the-library/policies-and-guidelines)

BUILDING THEORY WITH SOCIAL MEDIA:

EXPANDING THE HORIZON OF QUALITATIVE RESEARCHERS

Emmanuelle Vaast
McGill University
Montreal, QC
emmanuelle.vaast@mcgill.ca

Cathy Urquhart
Manchester Metropolitan University
Manchester, UK
c.urquhart@mmu.ac.uk

Abstract

The growing popularity and constant innovations of social media platforms and applications have transformed ways of interacting, working, creating value and innovating. We elaborate upon how building theory from case studies may be adapted to the opportunities and challenges of social media environments. We delve into key challenges of the research process: case study design, data analysis, and engaging in multi methods.

Keywords: Theory building, methods, social media, case studies, digital text, mixed methods.

INTRODUCTION

In the last few years, new generations of web applications have dramatically increased people's ability to interact with each other electronically, to generate content online, and to immerse themselves in alternative universes (boyd and Ellison 2008; Damer 2008; Messinger et al. 2009; O'Reilly 2007). These social media have recently become widespread, and led to new social practices (Koebler et al. 2010), sociability patterns (Van Den Eede 2010), learning practices (Greenhow and Robelia 2009; Kim and Abbas 2010), leisure activities, as well as social and political mobilization (Byrne 2008; Wattal et al. 2010). Social media use has also permeated the business domain (Culnan et al. 2010), triggering new business models (Lyons 2008), customer-relationship tactics (Di Gangi et al. 2010; Gallaugher and Ransbotham 2010) and managerial practices (Kaganer and Vaast 2010; Leidner et al. 2010).

The potential for social media research is huge, and there is a strong need for researchers to try and make sense of these new dynamics. Studies in sociology, marketing and Information Systems, have examined social media in some depth (e.g. Beer and Burrows 2007; Berente et al. Forthcoming; Cooke and Buckley 2009). Yet, much of this research has been *descriptive*, i.e. it has presented empirically, in sometimes great detail, what these new environments are. There is a growing opportunity for researchers to move beyond describing and towards theorizing these new contexts and associated practices (Majchrzak 2009).

Theory corresponds to "a statement of concepts and their interrelationships that shows how and / or why a phenomenon occurs" (Corley and Gioia 2011). We are interested in theory building that explicitly considers relationships between concepts as opposed to detailed description. This may encompass causal relationships. We also consider how theory building tools inherent in grounded theory methodology (such as theoretical sampling to expand the scope of a theory) may deserve to be used more with social media-based environments.

We focus on theory building from case studies in social media-based environments, and highlight the opportunities and challenges that qualitative researchers are likely to face in this pursuit. It seems urgent for researchers to develop ways of building theory for social media because many of them have embraced these environments as contexts for their research (e.g. (Huang et al. 2015; Vaast et al. 2013) and many more have been thinking about doing so. One can only anticipate that many more future publications will deal with social media environments. Consequently, there has been a growing concern, among social media scholars (e.g. (Kane and Fichman 2009; Kane et al. 2012; Majchrzak 2009; Te'eni 2009) that the *methodological* aspects of researching social media, and subsequent implications for theory building have not yet been sufficiently examined. In particular, there has not yet been much discussion of what it takes to build theories of these new environments. This paper elaborates on these emerging discussions by considering the theory building process in terms of the data that the researcher can collect from social media. We also examine new and possibly unexpected methodological dilemmas the researcher may face, and suggest ways to collect and analyze data from these environments in order to build theory. It is not our ambition to provide the last word on these important issues. This would be unrealistic, especially as social media and the intricately related social and technical conditions they generate are in flux, with a stream of new applications becoming available and adopted. Rather, building upon our own and other scholars' (Fleming and

Waguespack 2007) experience researching these contexts, we present documented arguments for taking theory building seriously when investigating social media.

SOCIAL MEDIA: A NEW FRONTIER FOR RESEARCH

Social media, or "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 and Haenlein 2010), have generated unprecedented opportunities for the development of new theories in research, as well as unanticipated methodological challenges for researchers. Opportunities lie in the multiplicity and diversity of methodological approaches that researchers might adopt in these environments: e.g. experiments (Antheunis and Scouten 2010; Minocha et al. 2010), ethnography (Garcia et al. 2009; Kien 2008), "cyber-archeology" (Zimbra et al. 2010), or case studies (Vaast and Walsham 2013; Veer 2011). The methodological challenges of dealing with these new environments are however also varied, and range from the technical (how to do it?) (Bollier 2010; boyd and Ellison 2008), to the legal (how legally accountable is the researcher?) (Allen et al. 2006; Lehmberg et al. 2008), and the ethical (what are the right courses of action?) (Beer 2008; Stanton 2010).

Social media environments vary substantially (see Table 1).

Table 1. Examples of widespread social media environments

Social media applications and environments	**Examples**	**Examples of publications**
Social networking sites	Facebook, Myspace	(Grasmuck et al. 2009)
Wikis	Wikipedia	(Niederer and van Dijck 2010)
Blogging and microblogging	Huffington Post Twitter	(Macias et al. 2009)
Virtual worlds	Second life	(Berente et al. Forthcoming; Schultze and Orlikowski 2010)
Tagging, bookmarking, online reviews	Del.ic.ious	(Scott and Orlikowski 2010) (Arriaga and Levina 2010)
Photo and video sharing sites	YouTube	(Lange 2008)

As a caveat, we note here that it is not our intention to claim that everything about social media is new, and that the challenges and opportunities they afford researchers are always unprecedented. There is a body of research on online communities (e.g. (Kraut et al. 1999; Ma and Agarwal 2007; Wilson and Peterson 2002) and open-source software development (e.g. (Fleming and Waguespack 2007; O'Mahony and Ferraro 2007; Von Hippel and Von Krogh 2003) that has for some time illustrated some of these challenges and opportunities. Vaast and

Walsham (2013), examined how grounded theorizing methods, in particular, might be adjusted in these computer-mediated and social environments. However, the fast growing and increasingly pervasive popularity of social media in many aspects of everyday life, and the highly dynamic character of social media applications (and of their waves of popularity) make it especially important for researchers to consider building theory of social media.

One of the major issues confronting the social media researcher is the question of what might be a legitimate unit of analysis when a whole range of data in the environment might (and perhaps should) be studied. The social media researcher has to deal, for instance, with web pages, chat threads, emails, and many visual images (see Table 2).

Table 2. Illustrative characteristics of digital texts

Characteristic	**Example**
Held in digital format	Emails, chat threads, photographs
Contained on a web site	Web content
Co produced by more than one person	Web forums, wikis
Ephemeral	Comments on a link, a news feed in Facebook
Embeds other discourses	Link within a webpage, linking digital text to one another
Contains images	Avatar, web content, photographs
Contains video	YouTube clips
Lack of context	Microblog posts (e.g. "tweets" of 140 characters or less).
Linguistic innovations	Emoticons, acronyms (e.g. "lol")

First, a digital text, by definition, is held in a digital format. The digital format can be held in any number of file formats, which can sometimes cause issues around data management.

The second characteristic, we think, is a common one – many social media researchers will be studying a particular web site, or type of web sites, e.g. social networking sites, dating sites, gambling sites and so on. The question then is what sort of context does the web site or sites provide? Can we see the web site as providing an overall frame for the study? Can we thus conceptualize the web site as providing the natural boundaries for the case study? Or does the boundary lie beyond the web site?

The third characteristic – co-produced by more than one person – produces various ethical issues for the researcher. For instance, is the discussion on a public forum deemed as being in the public domain and therefore, like, say, a text from a newspaper? If the forum is membership only, and discusses sensitive issues, should the people on that forum know that they are part of a research project? What are our responsibilities when we too, participate in these forums? Gaining permission to use, for instance, a stream of collaborative 'chats' about a project within an

organization is a different matter from considering what ethical concerns might operate in a web forum.

The fourth characteristic, the ephemeral nature of the text, also creates problems for the social media researcher. There needs to be a systematic way of capturing the texts so they are not lost. This can lead to the collection of large amounts of data, which in turn gives rise to two other problems – data management, and critically, deciding which texts might be worthy of analysis.

The fifth characteristic is that digital texts usually embed other texts through hyperlinks. A digital text is indeed usually not a "stand alone" text. One of the key features of digital environment is the ability to link content to each other. Therefore, digital texts are embedded in one another, creating new conditions for data collection.

The sixth characteristic, contains images, is, we think, an important characteristic. In such overwhelming visual environments such as those often afforded by social media (e.g. Facebook profile or pages) not to consider analyzing visual images that we encounter in such an environment may not do justice to that environment and its dynamics.

The seventh characteristic, contains video, underlines the previous characteristic. For instance, we know that video interviews give researchers much information from non-verbal cues. So we would suggest that the social media researcher might also gain insights from such video sources and should actively consider where, when and how they can be incorporated into the case study design.

The seventh characteristic, lack of context, is also one we think that social media researchers should pay particular attention too. Regardless of whether one is analyzing a digital text, or crunching some 'big data', it is impossible to infer patterns without context. For instance, if a stream of Skype chats gets recorded, it might be also good for the researcher to note some basic context (e.g. time of day, day of week). Deeper context can be grasp by examining who are the people involved in these chats and, perhaps, by interviewing them.

Finally, the last characteristic is that of linguistic innovations and includes the growing reliance upon acronyms (e.g. "lol" for "laugh out loud," "FYI" for "for your information," or "IMHO" for "In my Humble Opinion") in digital texts, as well as the tendency for digital "texts" to blur the distinction between iconic representations and discourse (Wolf 2000). Increasingly in web-based forums as well as in microblog posts, for instance, emoticons have complemented traditional language-based content and contributed to new ways of expressing complex ideas and sentiments online (Bos et al. 2007). They have transformed written text, making it in some ways closer to oral language, such as when an "lol" punctuates a digital text in a similar way that laughter punctuates an unmediated conversation (Spencer and Mandell 2007). Another similar linguistic innovation is the evolving use of hashtags on Twitter. From being simple tags on which a search can be performed, they also perform the function of a summarized and often humorous take on the tweet in question (Vaast et al., 2012).

Taken together, these characteristics of social media environments provide unexpected challenges for researchers, who then need to adjust their theory building methods.

CHALLENGE 1 – CASE STUDY DESIGN

A key challenge is related to the possibility to collect large datasets, much larger than what many qualitative researchers would have considered for just a few years ago. In this section, we

consider how we might design a case study for a social media environment, given these large datasets. It is important to note here that we are defining a case study in broad terms, as we believe that a flexible definition of what a case is makes sense in social media studies. "*In the sociological and anthropological literature, a case is typically regarded as a specific and bounded (in time and place) instance of a phenomenon selected for study. The phenomenon of interest may be a person, process, event, group, organization, and so on*." (Schwandt 1997).

In particular, we consider how we might select data of interest within such large datasets, how we might decide on a case study boundary, how we might handle the issue of context, and whether single or multiple case studies are appropriate.

The availability of "big data" has arisen from two main conditions. The first one, obviously, and aforementioned, is related to the growth of social media, in particular in terms of available platforms and features as well as in terms of their increasing popularity that have contributed to new ways of communicating, working, etc. The second condition comes from increased computational abilities that have enabled companies to develop and populate huge databases as well as computer-mediated communications to become socially omnipresent and very cost-effective (Jacobs 2009).

From the possibility to collect huge data for research purposes, new challenges have arisen. In particular, how does the researcher select what might be significant data in a welter of logs, chats, emails and other data? The temptation for the unwary researcher is to collect as much data as possible – and of course, it is possible to collect a great deal of data in a social media environment. For instance, if collecting data on a virtual open source project, possible archival sources would include, for instance: emails; bug tracking data or software version control logs.

For illustrative purposes, let us consider a situation in which researchers decide that they actually do need to collect large volumes of data, say from social networking websites, blogs, or microblog archives. They are likely to rely upon an Application Programming Interface (API) that will help them access these data. One such publicly available API is Topsy, an openly accessible and relatively user-friendly search engine for Twitter, currently the most popular microblogging platform. Through Topsy, anyone, including researchers, can use the API to access the entire archive of Twitter (Topsy labs 2010). There are huge opportunities, of course, for researchers in being able to collect so much data and to be selective in their data collection (e.g. researchers can search for key terms and specific times). One problem here is that search engines and APIs can actually influence the data collection process: relying upon search engines and APIs can lead to biased samples, with the "bias" not always or easily understood by the researchers (Bollier 2010; boyd and Crawford 2011). In effect, rather than purposeful theoretical sampling, the data collection can become the result of an uncontrolled and little understood process of opaque algorithm-led, data extraction. Therefore, researchers need to be aware of how these algorithms might influence their dataset.

So how might we begin to select data for our theory building case study? First, we might consider the notion of 'central' texts, and less central texts, and how those texts should be analyzed. There is always a 'depth versus breadth' concern in any qualitative research design; the more deeply a text is analyzed, the less texts will be analyzed. One possible route here is to analyze some texts more deeply than others – hence the notion of having central texts to analyze. Less central texts can be used to provide corroboration or triangulation. Readers may say at this point that this is merely primary and secondary data. We would beg to differ, because in fact, if

theoretical sampling is used, 'data slices' could, and should, come from different sources – so we prefer the distinction between 'central,' and 'less central.' The logic of settling on a central text helps to provide an entryway for the researcher and to think about which data sources are important in the design.

We think that the idea of a digital text also has some reach when considering the unit of analysis of a study. Of course, the unit of analysis is always dependent upon the specific research questions of various projects. The idea of a digital text data forces researchers to think about what is the unit of analysis of their research, and what this unit of analysis means to them. For instance, in a recent study by one of the authors on bankers and online forums (Vaast and Levina 2015), the unit of analysis was the post, but other units could have been selected (e.g. threads as a whole).

Once central texts, and less central texts, have been identified, we should then arrive at the *boundary* of the case study. One important consideration for social media researchers is whether their boundary is the social media boundary – do they study purely virtual environments? While it is entirely appropriate for ethnographers to study 'virtual worlds' as ethnographic environments, we contend that to place the boundary within or on a social media environment is not always advisable. First, when we consider the permeability of the division between 'real' and 'virtual' worlds, it is perhaps unrealistic to simply study what goes on in a social media environment. For instance, friendships in Facebook have corresponding relationships in 'real' life. There is often a reflexive relationship between the two. Second, one major challenge of studying social media environments, we contend, is that of *context*. How can we interpret a digital text if we are not aware of its context?

If we agree that context is important in a social media environment, whether when using digital texts or when using 'big data' expressed as visualizations, how do we build a context for our case study design? If we consider what our central text might be, we can at the same time consider how context can be built in. For instance, if we are interested in how hotel operators respond to negative reviewers on Trip Advisor, we might opt to use their responses on the web site as our central text for the case study (Orlikowski and Scott 2014). It might also be worth interviewing those owners about their response to follow up on issues that come up from the analysis or simply to find out about other factors that were in operation at the time for that owner. Similarly, if we collect a stream of posts about the development of a software product, it would be sensible to relate this to field notes about what was happening in the project at the time. Even if the researcher focuses exclusively on the virtual environment, the same argument for context still applies. For instance, let us imagine that researchers are investigating microblogging (e.g. activity on Twitter) in various contexts and situations (Heverin 2011; Jansen et al. 2009; Vaast et al. 2012). Researchers in this situation are likely to select single microblogging posts (i.e. the "tweets") as their unit of analysis. At the same time, though, a tweet is very short (140 characters or less): making sense of a tweet in and of itself is difficult – tweets are to be understood within an ensemble (cf. the notion of "ambient awareness" (boyd et al. 2010; Marwick and Boyd 2010). For researchers, the implication is that it is sometimes difficult, but important, to access the context of digital text data. Following our previous example, when studying microblogging, researchers could consider what was trending in Twitter at the time; they could consider the surrounding tweets, and the sequence of tweets and retweets (boyd et al. 2010; Lotan et al. 2011; Namaan et al. 2011).

While thus far in this chapter, the plethora of data available to the social media researcher can be seen as carrying some disadvantages, there is one respect in which it can be seen as a distinct advantage; many digital texts lend themselves to *theoretical sampling* for theory building. The beauty of this solution is clear, as the developing theory determines which data 'slice' is examined next. In a social media environment, the ability to sample 'data slices' is extremely flexible. It will probably not require consent (although terms of usage for social media applications change and need to be checked regularly), or another phase of a study for 'member checking' as in a traditional case study.

Theoretical sampling can proceed in two ways. The first is exemplified in Eisenhardt (1989) where successive case studies are chosen on the basis of similarity or difference with the previous case and within case patterns. The second is a more systematic view from Glaser and Strauss (1967) where group differences are not only minimized or maximized, but sampling also unfolds with concept development. For instance, if the analysis of digital texts reveals many instances of a concept (e.g. self-presentation through Facebook status messages), one could choose to go on and sample more of that concept by interviewing individuals. Similarly, one could choose to sample further on a concept that has only occurred in a particular group of people, or is unusual (e.g. use of a Facebook status to communicate with only a few people in a very personal way). Glaser and Strauss suggest that sampling along diverse concepts quickly develops the theory and delimits the scope of the theory. Theoretical sampling in this systematic manner, we suggest, is underutilized in most management disciplines, including our own. Sampling using not only 'different' and 'similar' cases, but also looking for guidance in terms of concept development, means that we develop a much better, more grounded theory, with better scope.

Again, such flexible theoretical sampling of digital texts could be seen as a double-edged sword. When digital texts contain links to other digital texts that have some conceptual relevance, where should the researcher set the boundary for their case study? An obvious pitfall of is that researchers might soon find themselves overwhelmed by potentially relevant data. At the very extreme, link by link, the whole Internet could become their research setting, making theory building all but impossible and meaningless. There is thus an increased need to set boundaries for the cases under investigation.

CHALLENGE 2 – DATA ANALYSIS

Generally, when analyzing textual data, there are two main options: either the researcher can code the text at a detailed level, or they can apply a thematic framework of some kind. This of course still applies in the social media environment, but there are many more texts that can potentially be analyzed. So, distinguishing between these two options is important because they have different analytical 'loads'; if analyzing a text in detail, there will be less chance to analyze many texts. This of course does not mean that, in practice, the researcher should not employ multiple methods of analysis – but that the time needed to apply different methods should be considered. In short, there may be a 'depth versus breadth' issue. As suggested in the case study section, we suggest that the social media researcher opts for some 'central' texts that can be analyzed in depth, because in the opinion of the authors, there is no substitute for such in depth analysis (with one important proviso – that this analysis is subsequently theoretically integrated).

When coding at a detailed level, there are three options for coding. First, bottom up coding, where codes are suggested by the data. Grounded theory method is a very good example of this

approach, and can yield rich results because of a close tie with the data. Grounded theory method also gives a systematic way to theoretically sample different digital texts, as discussed previously.

Second, codes can be suggested by the literature and constructs used in instruments – this is 'top down' coding. Thirdly, codes can be applied which are a mixture of codes suggested by the data, *and* the literature. Overall, the coding approach gives the researcher assurance that what they have in their data really is in their data, because coding confers a more systematic approach to analysis. It also helps qualitative researchers defend themselves from charges of being selective about what they analyze.

Thematic frameworks are also frequently used to analyze qualitative data, and there are many possibilities available. For instance, there are many frameworks from discourse analysis and critical discourse analysis (e.g. (Fairclough 1992) that could be applied to digital texts. Thematic frameworks or models built from relevant literatures, for instance a 'sensitizing framework' as suggested by Klein and Myers (1999) is also a frequent strategy used by researchers. An alternative is to build a thematic framework from the data, as suggested by (Braun and Clarke 2006).

We would also urge social media researchers to seriously consider how they might analyze *visual* digital texts they encounter, and the role that such analyses can extend and enrich the emerging theory. There are various ways of analyzing such visuals. First, the text can be coded as if it was any other type of digital text. Given that most qualitative data analysis packages do have the capability to both store such visuals, and code them, there seems to be no obstacle to this type of analysis except a lack of familiarity, and a possible fear that this type of analysis might not be published. We could also view images as not neutral, but as constructed texts (Banks 2007). This makes sense particularly with social media, where people manage how they appear on-line, and there are web sites, for instance, devoted to 'photobombs' where the subject of the photograph is sometimes accidently, sometimes deliberately upstaged by another element in the photograph. We can take a Foucauldian view of the image– who is doing the looking, whom does society empower to look at and be looked at, and what knowledge does this produce? Of course, visual analysis is not new, it has a long pedigree in cultural studies, for instance. One useful perspective comes from (Ball and Smith 1992) who point out that it is important to distinguish between manifest and latent content for analytical purposes. For instance, an old photograph may contain a man with a mutton chop beard (manifest content) but the latent content concerns the social meaning of that beard at the time the image was produced (Ball and Smith 1992) in (Banks 2007). This brings us back to a key element which needs to be considered by the social media researcher when analyzing – context. This is why we would recommend that the issue of context be considered early, in the case study design.

CHALLENGE 3 – ENGAGING IN MIXED METHODS

Social media environments also constitute unprecedented research settings, because they generate a wealth of digital text data that, technically at least, researchers may collect exhaustively and with little to no interference on the socio-technical dynamics at play. This new situation has generated tremendous opportunities but also challenges for researchers (Bollier 2010; boyd and Crawford 2011; Lazowska 2008; Manovich 2011). Some of these challenges are legal, as detailed in (Allen et al. 2006) seminal essay on regulatory implications of automatic electronic data collection. Other key challenges are of an ethical nature, and deal with, for

instance, the blurring of the distinction between what is private and what is public, or questions regarding whether electronic data are "people" or "public" (Buchanan 2010; Buchanan and Ess 2008; Hudson and Bruckman 2004; Schultze and Mason 2011; Sveningsson 2004). Aware of these critical issues, in this essay, we focus on another important, and so far less discussed, challenging implication of "big data" for theory building, the need to engage in mixed methods.

In this regard, (Anderson 2008) provocatively called for a drastic change in the scientific method, given the rise of electronic data becoming publicly available to many. He predicted that "*the end of theory*" was being brought about by the "*data deluge*" of the Internet, search engines, and social media. Intrigued by this provocative thesis, we however believe that the scientific community can still contribute to society not only by discovering and explaining correlations, but also by theorizing causations, and multiple-order effects between concepts. Others have discussed how electronic data might transform both the process of developing new theories and the resulting theories. John Seely Brown, for instance, as reported in (Bollier 2010), addressed how newly available data can help researchers discover "generators" for new theories if researchers are able to make sense of what in their data, corresponds to "outliers" and what reveals meaningful patterns:

> "How can you invent the 'theory behind the noise' in order to de-convolve it in order to find the pattern that you weren't supposed to find? The more data there is, the better my chances of finding the 'generators' for a new theory."

We argue that for researchers to be able to discover the "generators" for a new theory, they need more than powerful computational abilities, although they obviously do need such resources to be able to sift through huge volumes of data. To do so, researchers interested in theory building, and not just in descriptive analyses or in testing existing theories with electronic data, would do well to engage in mixed methods analyses.

The huge volume of data that researchers can collect in social media environments creates data processing, reduction, and management challenges. Obviously, such a volume of data can be impressive, but is also utterly meaningless in and of itself. Collecting huge volumes of data, per se, does not guarantee a theoretical contribution for the research. If anything, big data might make building theory more difficult, because theorizing patterns and outliers becomes especially challenging in a wealth of decontextualized data. Researchers might not always feel well equipped to handle these datasets. Researchers who are usually engaged in qualitative research might find themselves intimidated by the volume of data, and might not know how to "attack" it to develop, for instance, a grounded theory of their phenomenon of focal interest. Getting deeply into a corpus of semi-directed interviews of reasonable duration is already formidable enough; how to make sense of years of archives from online discussion forums, for instance? For qualitative researchers dealing with such new contexts and new data, then, being able to delve into data from a more quantitative angle, thanks to descriptive visualization techniques, can be helpful.

Quantitative researchers are not immune to the challenges of big data for theory building either. A major issue for quantitative researchers is that, when huge volumes of data have been collected, the significance level loses some of its meaning, and putting too much weight on it for theory justification purposes, as opposed to the power of the analyses, might weaken the resulting theory (Bollier 2010; boyd and Crawford 2011). To build theories of these new environments, quantitative researchers therefore cannot rely singlehandedly on well-established

statistical indicators, and need to be able to give meaning to their data. Adoption of more qualitative approaches to theory building may assist with achieving this sensemaking.

Therefore, handling digital text data for theory-building purposes might force researchers out of their typical (qualitative/quantitative) comfort zone, and lead them to embrace methodological creativity. Qualitative researchers might try to deal with larger data sets with descriptive quantitative methods, and quantitative researchers might delve into the meaning of their data, rather than focus mostly on the significance level of their statistical procedures. Obviously, calls for triangulation for better theory building are far from new (see, in organization research, (Jick 1979; Shah and Corley 2006; Van Maanen 1979). Triangulation, in the sense of combination of different methods, especially involving qualitative and quantitative aspects, is particularly critical for theory building of social media contexts. For one, it can help researchers deal with huge volume of data and sift through them, relying upon multiple techniques, to identify patterns to be theorized. Moreover, mixed methods offer researchers the ability to develop complementary perspectives on the same phenomenon and to discern previously hidden connections among concepts.

On a practical note, many researchers are more familiar with qualitative or quantitative research, although many of us would also not consider ourselves exclusive proponents of a single method over other ones. Engaging in multi-method theory building research projects can help us reach beyond what we have usually become more familiar and, for some of us, even expert on, methodologically. This, obviously, makes such research demanding, because it is not enough to develop "qualitative" or "quantitative" research, and researchers have to develop research projects that combine various methodological "ideal types." We therefore hope that the opportunities and challenges of developing mixed methods for theory-building purposes will lead researchers to reach out and develop collaborations across the traditional methodological divides. Researchers at ease with the latest computational methods can summarize vast amount of data and develop impressive, descriptive visualizations (Lazowska 2008). Other researchers, more attuned to discovering theories, could work with them in order to make sense of empirical patterns and soundly the "creative leap" (Langley 1999) that is necessarily involved in any theory building effort. Such mixed methods research projects would still be challenging to develop, because there are still currently fewer established criteria to evaluate mixed methods than for more traditional methods. We would like to see researchers involved in mixed methods projects be more explicit about their methodological choices, and engage in conversations about those choices.

CONCLUSION

The purpose of this chapter was to reflect on the building theory of social media-related environments, and to articulate the challenges and opportunities for researchers in these new environments.

Theory building remains an enduring challenge for many researchers in disciplines that deal with 'applied' environments. The advent of social media has ushered in two major changes in how we use the Internet. First, it is a much more collaborative era, that allows people to state their opinions, and for their opinions to be gathered, via crowdsourcing. This gives the researcher an increased opportunity to study varying viewpoints, and gain corroboration of particular viewpoints. Second, visual images are now so much of our social worlds in social media that we think we could, and should, as researchers, be analyzing those images.

The advent of 'big data' cannot be understated as a development either; there is a huge potential to collect a large amount of data, but with a corresponding need to make sense of that data. The role of context in a social media should in this regard be acknowledged, as it is required to make sense of either a large data set or a 'slice' of data. We recommend then, that, when embarking on research in social media environments, there be an active consideration of how context will be accounted for.

Finally, the availability of large datasets has generated a critical need for researchers to go beyond their familiar methodological choices and to embrace mixed methods. Engaging in mixed methods could trigger productive collaborations among researchers and help generate thought-provoking conceptualizations.

This paper represents a first step in opening a dialogue among researchers interested in theorizing about social media-related environments. We look forward to engaging with our colleagues, and to further elaborating on these ideas about possible methodological approaches.

REFERENCES

Allen, G. N., Burk, D. L., and Davis, G. 2006. "Academic data collection in electronic environments: defining acceptable use of internet resources," *MIS Quarterly* (30:3) September, pp 500-510.

Anderson, C. 2008. "The end of theory: The data deluge makes the scientific method obsolete," *Wired*) 06/23/2008.

Antheunis, M., and Scouten, A. P. 2010. "The effects of other-generated and system-generated cues on adolescents' perceived attractiveness on social network sites," *Journal of computer-mediated communication* (16), pp 391-406.

Arriaga, M., and Levina, N. 2010. "Bourdieu's Practice Theory in Organizational Studies: A call for quantitative methods use," in *AOM Annual Meeting*: Montreal, Canada.

Ball, M., and Smith, G. 1992. *Analysing Visual Data* (Sage: London.

Banks, M. 2007. *Using Visual Data in Qualitative Research*, (Sage: London.

Beer, D. 2008. "Researching a confessional society," *International journal of market research* (50:5), pp 619-629.

Beer, D., and Burrows, R. 2007. "Sociology and, of, and in Web 2.0: Some initial considerations," *Sociological research online* (12:5) 17.

Berente, N., Hansen, S., Pike, J., and Bateman, P. J. Forthcoming. "Arguing the value of virtual worlds: Patterns of discursive sense making of innovative technology," *MIS Quarterly*).

Bollier, D. 2010. *The promise and peril of big data*, (Aspen Institute: Washington DC.

Bos, D., Bos, A. E. R., and von Grumbkow, J. 2007. "Emoticons and social interaction on the Internet: The importance of social context," *Computers in Human Behavior* (23:1), pp 842-849.

boyd, D., and Crawford, K. Year. "Six provocations for Big Data," A Decade in Internet Time: Symposium on the Dynamics of the Internet and Society, Oxford, UK, 2011.

boyd, D., Golder, S., and Lotan, G. Year. "Tweet, tweet, retweet: Conversational Aspects Of Retweeting On Twitter," 43rd Hawaii International Conference On System Sciences, Hawaii, 2010.

boyd, D. M., and Ellison, N. B. 2008. "Social network sites: Definition, history, and scholarship," *Journal of computer-mediated communication* (13), pp 210-230.

Braun, V., and Clarke, V. 2006. "Using thematic analysis in psychology," *Qualitative Research in Psychology* (3), pp 77-101.

Buchanan, E. A. 2010. "Internet research ethics: Past, present, and future," in *The handbook of Internet studies,* R. Burnett, M. Consalvo and C. Ess (eds.), Blackwell: Malden, MA, pp. 83-108.

Buchanan, E. A., and Ess, C. 2008. "Internet research ethics: The field and its critical issues," in *The handbook of information and computer ethics,* K. E. Himma and H. T. Tavani (eds.), Wiley: New York, NY, pp. 273-292.

Byrne, D. N. 2008. "Public discourse, community concerns, and civic engagement: Exploring black social networking traditions on Blackplanet.com," *Journal of computer-mediated communication* (13), pp 319-340.

Cooke, M., and Buckley, N. 2009. "Web 2.0, social networks and the future of market research," *International journal of market research* (50:2), pp 267-292.

Corley, K. G., and Gioia, D. A. 2011. "Building theory about theory building: What constitutes a theoretical contribution?," *Academy of Management Review* (36:1), pp 12-32.

Culnan, M. J., McHugh, P. J., and Zubillaga, J. I. 2010. "How large U.S. companies can use Twitter and other social media to gain business value," *MIS Quarterly Executive* (9:4) Dec., pp 243-259.

Damer, B. 2008. "Meeting in the ether: A brief history of virtual worlds as a medium for user-created events," *Journal of virtual worlds research* (1:1) July, pp 1-17.

Di Gangi, P. M., Wasko, M. M., and Hooker, R. E. 2010. "Getting customers' ideas to work for you: Learning from Dell how to succeed with online user innovation communities," *MIS Quarterly Executive* (9:4) Dec., pp 213-228.

Fairclough, N. 1992 "Discourse and Text: Linguistic and Intertextual Analysis within Discourse Analysis," *Discourse Society* (Vol. 3:no. 2), pp 193-217.

Fleming, L., and Waguespack, D. M. 2007. "Brokerage, boundary spanning, and leadership in open innovation communities," *Organization Science* (18:2) March-April, pp 165-180.

Gallaugher, J., and Ransbotham, S. 2010. "Social media and customer dialog management at Starbucks," *MIS Quarterly Executive* (9:4) Dec., pp 197-212.

Garcia, A. C., Standlee, A. I., Bechkoff, J., and Cui, Y. 2009. "Ethnographic approaches to the Internet and computer-mediated communication," *Journal of Contemporary Ethnography* (38:1), pp 52-84.

Grasmuck, S., Martin, J., and Zhao, S. 2009. "Etho-racial identity displays on Facebook," *Journal of computer-mediated communication* (15), pp 158-188.

Greenhow, C., and Robelia, B. 2009. "Old communication, new literacies: Social network sites as social learning resources," *Journal of computer-mediated communication* (14), pp 1130-1161.

Heverin, T. 2011. "Use of microblogging for collective sense-making during violent crises: A study of three campus shootings," *Journal of the American Society for Information Science and Technology* (63:1), pp 34-47.

Huang, J., Baptista, J., and Newell, S. 2015. "Communicational ambidexterity as a new capability to manage social media communication within organizations," *The Journal of Strategic Information Systems* (24:2), pp 49-64.

Hudson, J. M., and Bruckman, A. 2004. ""Go away": Participant objections to being studied and the ethics of chatroom research," *The Information Society* (20), pp 127-139.

Jacobs, A. 2009. "The pathologies of big data," *Communications of the ACM* (52:8) August, pp 36-44.

Jansen, B. J., Zhang, M., Sobel, K., and Chowdury, A. 2009. "Twitter power: tweets as electronic word of mouth," *Journal of the American Society for Information Science & Technology* (60:11), pp 2169-2188.

Jick, T. D. 1979. "Mixing qualitative and quantitative methods: Triangulation in action," *Administrative Science Quarterly* (24:4), pp 602-611.

Kaganer, E. A., and Vaast, E. Year. "Responding to the (almost) unknown: Social representations and corporate policies of social media," International Conference of Information Systems, Saint Louis, MI, 2010.

Kane, G. C., and Fichman, R. G. 2009. "The shoemaker's children: Using wikis for Information Systems teaching, research and publication," *MIS Quarterly* (33:1) March, pp 1-17.

Kane, G. C., Levina, N., Pikorski, M. J., Ransbotham, S., Griffith, T., Majchrzak, A., Borgatti, S., Faraj, S., and Jarvenpaa, S. L. 2012. "Researching the informal economy: Opportunities and challenges of social media research," in *Academy of Management Conference*: Boston, MA.

Kaplan, A. M., and Haenlein, M. 2010. "Users of the world, unite! The challenges and opportunities of social media," *Business Horizons* (53) January-February, pp 59-68.

Kien, G. 2008. "Technography = Technology + ethnography, an introduction," *Qualitative Inquiry* (14:7) October, pp 1101-1109.

Kim, Y.-M., and Abbas, J. 2010. "Adoption of library 2.0 functionaliies by academic libraries and users: A knowledge management perspective," *The journal of academic librarianship* (36:3), pp 211-218.

Koebler, F., Riedl, C., Vetter, C., Leimeister, J. M., and Krcmar, H. Year. "Social Connectedness on Facebook :An Explorative Study on Status Message Usage," Proceedings of the Sixteenth Americas Conference on Information Systems, Lima, Peru, August 12-15, 2010.2010.

Kraut, H., Steinfield, C., Chan, A., Butler, B., and Hoag, A. 1999. "Coordination and virtualization : the role of electronic networks and personal relationships," *Organization Science* (10:6) Nov - Dec, pp 722 - 740.

Lange, P. G. 2008. "Publicly private and privately public: Social networking on YouTube," *Journal of computer-mediated communication* (13), pp 361-380.

Langley, A. 1999. "Strategies for theorizing process data," *Academy of Management Review* (24:4) Oct., pp 691 - 710.

Lazowska, E. 2008. "Envisioning the future of computing research," *Communications of the ACM* (51:8) August, pp 28-30.

Lehmberg, T., Rehm, G., Witt, A., and Zimmerman, F. 2008. "Digital text collecions, linguistic research data, and mashups: Notes on the legal situation," *Library trends* (57:1) Summer, pp 52-71.

Leidner, D. E., Koch, H., and Gonzales, E. 2010. "Assimilating generation Y IT new hires into USAA's workforce: The role of an enterprise 2.0 system," *MIS Quarterly Executive* (9:4) Dec., pp 163-176.

Lotan, G., Graeff, E., Ananny, M., Gaffney, D., Pearce, I., and Boyd, D. 2011. "The revolutions we tweeted: Information flows during the 2011 Tunisian and Egyptian revolutions," *International journal of communication* (5), pp 1375-1405.

Lyons, K. 2008. "Toward a theoretically-grounded framework for evaluating immersive business models and applications: Analysis of ventures in Second life," *Journal of virtual worlds research* (1:1) July, pp 1-19.

Ma, M., and Agarwal, R. 2007. "Through a glass darkly: Information Technology design, identity verification and knowledge contribution in online communities," *Information Systems Research* (18:1) March, pp 42-67.

Macias, W., Hilyard, K., and Freimuth, V. 2009. "Blog functions as risk and crisis communication during Hurrican Katrina," *Journal of computer-mediated communication* (15), pp 1-31.

Majchrzak, A. 2009. "Comment: Where is the theory in wikis?," *MIS Quarterly* (33:1) March, pp 18-21.

Manovich, L. 2011. "Trending: The promises and challenges of big social data," in *Debates in the Digital Humanities,* M. K. Gold (ed.), University of Minnesota Press: Minneapolis, MN.

Marwick, A. E., and Boyd, D. 2010. "I tweet honestly, I tweet passionately: Twitter users, context collapse, and the imagined audience," *New media & society* (13:1), pp 114-133.

Messinger, P. R., Stroulia, E., Lyons, K., Bone, M., Niu, R. H., Smirnov, K., and Perelgut, S. 2009. "Virtual worlds - Past, present, and future: New directions in social computing," *Decision support systems* (47), pp 203-228.

Minocha, S., Tran, M. Q., and Reeves, A. J. 2010. "Conducting Empirical Research in Virtual Worlds: Experiences from two projects in Second Life," *Journal of virtual worlds research* (3:1), pp 1-21.

Namaan, M., Becker, H., and Gravano, L. 2011. "Hip and trendy: Characterizing emerging trends on Twitter," *Journal of the American Society for Information Science & Technology* (62:5) May.

Niederer, S., and van Dijck, J. 2010. "Wisdom of the crowd or technicity of content? Wikipedia as a sociotechnical system," *New media & society* (12:8), pp 1368-1387.

O'Mahony, S., and Ferraro, F. 2007. "The emergence of governance in an open source community," *Academy of Management Journal* (50:5), pp 1079-1106.

O'Reilly, T. 2007. "What is web 2.0: Design patterns and business models for the next generation of software," *Communications & Strategies*:1) First quarter, p 17.

Orlikowski, W. J., and Scott, S. V. 2014. "What Happens When Evaluation Goes Online? Exploring Apparatuses of Valuation in the Travel Sector," *Organization Science* (25:3) May-June, pp 868-891.

Schultze, U., and Mason, R. O. 2011. "Ethics of online research: Inquiry in a fishbowl," in *Acadademy of Management Conference*.

Schultze, U., and Orlikowski, W. J. 2010. "Research commentary - Virtual worlds: A performative perspective on globally distributed, immersive work," *Information Systems Research* (21:4) December, pp 810-821.

Schwandt, T. 1997. *Qualitative Inquiry*, (Sage: London.

Scott, S. V., and Orlikowski, W. J. Year. "Reconfiguring relations of accountability: The consequences of social media for the travel sector," Annual meeting of the Academy of Management, Best paper proceedings, Montreal, Canada, 2010.

Shah, S. K., and Corley, K. G. 2006. "Building Better Theory by Bridging the Quantitative-Qualitative Divide," *Journal of Management Studies* (43:8), pp 1821-1835.

Spencer, R. J., and Mandell, D. L. 2007. "Emotional Expression Online Emoticons Punctuate Website Text Messages," *Journal of language and social psychology* (26:3), pp 299-307.

Stanton, J. M. 2010. "Virtual worlds, the IRB and a user's bill of rights," *Journal of virtual worlds research* (3:1) November, pp 3-15.

Sveningsson, M. 2004. "Ethics in internet ethnography," in *Readings in virtual ethics: Issues and controversies,* E. A. Buchanan (ed.), Idea group: Hershey, PA, pp. 45-61.

Te'eni, D. 2009. "Comment: The wiki way in a hurry - The ICIS anecdote," *MIS Quarterly* (33:1), pp 20-22.

Topsy labs 2010. "Using influence to tune signal to noise on the social web."

Vaast, E., Davidson, E. J., and Mattson, T. 2013. "Talking about technology: The emergence of new actors with new media," *MIS Quarterly* (37:4) December, pp 1069-1092.

Vaast, E., and 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), pp 73-98.

Vaast, E., Negoita, B., and Safadi, H. 2012. "Symbolic action through microblogging during and after a crisis: Tweeting the oil spill," in *Academy of Management Conference, OCIS Division*: Boston, MA.

Vaast, E., and Walsham, G. 2013. "Grounded theorizing for electronically-mediated social contexts," *European Journal of Information Systems* (22:1), pp 9-25.

Van Den Eede, Y. 2010. ""Conversation of manking" or "idle talk"? A pragmatist approach to social networking sites," *Ethics and Information Technology* (12), pp 195-206.

Van Maanen, J. 1979. "Reclaiming qualitative methods for organizational research: A preface," *Administrative Science Quarterly* (24:4) Dec, pp 520 - 526.

Veer, E. 2011. "Staring: How Facebook Facilitates The Breaking Of Social Norms," *Research in Consumer Behavior* (13), pp 185–198.

Von Hippel, E., and Von Krogh, G. 2003. "Open source softare and the "private-collective" innocation model: Issues for organization science," *Organization Science* (14:2) March-April, pp 209-223.

Wattal, S., Schuff, D., Mandviwalla, M., and Williams, C. B. 2010. "Web 2.0 and politics: The 2008 U.S. Presidential election and an e-politics research agenda," *MIS Quarterly* (34:4) Dec., pp 669-688.

Wilson, S. M., and Peterson, L. C. 2002. "The anthropology of online communities," *Annual Review of Anthropolgy* (31), pp 449-467.

Wolf, A. 2000. "Emotional expression online: Gender differences in emoticon use," *Cyberpsychology & behavior* (3:5), p 827.

Zimbra, D., Abbasi, A., and Chen, A. 2010. "A cyber-archaelogy approach to social movement research: Framework and case study," *Journal of computer-mediated communication* (16), pp 48-70.