Extending the Field – Challenges to Qualitative Research in Large Scale Online Environments

Dana Rotman and Jennifer Preece
College of Information Studies - Maryland’s iSchool
Human-Computer Interaction Lab
University of Maryland
<drotman, preece>@umd.edu

AUTHORS’ BACKGROUND

Dana Rotman is a PhD candidate at the College of Information Studies - Maryland’s iSchool. Her research focuses on online crowdsourcing, and examines motivational frameworks for social participation in large scale online collaborations. She is the author of several journal articles, conference papers and book chapters. She won the 2009 Yahoo! Key Scientific Challenges award and the CHI 2010 Student research competition.

Jennifer Preece is a Professor and Dean at the College of Information Studies – Maryland’s ischool. She is a co-author of Interaction Design: Beyond HCI (2011, 3rd Ed.) and author of Online Communities: Designing Usability, Supporting Sociability, both published by John Wiley & Sons: Chichester, UK. She is particularly interested in how to motivate continuing participation for contribution and curation of information in citizen science online organizations.

In our research we employ qualitative methods including observations and semi-structured interviews, which draw from traditional ethnography but adapt to the limitations posed by the studied fields. We extensively use grounded theory as an analysis method. We believe these methods provide valuable insights into experiences, motivations and social processes. Our qualitative work is the basis for two NSF-supported projects: SES 0968546 – “Biotracker - Melding human and machine intelligence to create large-scale collaborative systems”, and IIS 11019993 EAGER – “Extreme Ethnography: When Content and Tools Change Continually on Vast Scales, How Must Our Research Methods Change?” These two projects, and especially the latter, provide the impetus for us to submit a proposal to your workshop. We believe we can contribute valuable insights from our work, but most of all we believe we have lots to learn and will benefit enormously from being part of the community brought together by this workshop.

RESEARCH EXPERIENCES

Understanding what motivates citizen scientists to contribute data to biodiversity projects

Citizen scientists are people who contribute their time, knowledge and skills towards scientific projects, although they may not have formal training in science. Consequently they and their contribution are often not recognized and dully appreciated by scientists, even though some citizen scientists may reach high levels of expertise gained by experience, and despite the fact that collaborative scientific projects are becoming popular in recent years. In this study we sought to find out what motivates citizen scientists to contribute data to biodiversity websites (e.g., www.eol.org, www.ispot.org.uk). We employed a mixed method approach, starting from a questionnaire addressing both citizen scientists and scientists; some of the questions were open-ended and focused on why participants contribute data to biodiversity projects – in other words, what motivates them to contribute. We also followed the activity on these websites and conducted semi-structured interviews with both scientists and citizen scientists who were involved in these and other collaborative projects. The data was then analyzed using grounded theory [1]. We selected grounded theory because we wanted to gain a bottom-up, organic, participant-oriented understanding of their motivation. Several themes emerged, which at a high-level all involved being recognized, valued and respected, especially by scientists as well as their peers. One aspect that emerged from the data which we might have missed had we not opted to analyze it using grounded theory was that motivation changed with the length of involvement with the site. From this data we have proposed a motivation cycle [2] that we will now evaluate with different biodiversity communities involving citizen scientists. This study was focused on a specific community, which although geographically dispersed, had similar practices and closely related norms across disciplines and localities. Translating the findings from the specific citizen science community to a particular tool may pose challenges speaking to the constant tension between the in-depth, limited scope nature of qualitative work and the need for generalization in design.

Understanding video sharing practices

The practice of video-sharing creates immense opportunities for collaboration and interaction. From mashups and remixes of popular videos, topical collaborations based on mutual interests, to personal interaction grounded in comments, threaded conversations and befriending other users. In order to understand the practices and motivations of video sharing we conducted a study of sub-communities within YouTube [3]. The study
Qualitative data is much harder to amass. However, qualitative data is what allows us to go beyond mere patterns of interaction and gain a deeper understanding of motivation, interaction, closeness and affection. Some of the challenges that we’ve faced during this study led to our EAGER project which aims to unpack the challenges that large scale online environments pose to researchers. We outline some of these challenges next.

**Challenges to conducting qualitative work in large scale online environments**

Large scale online environments include – among others - social networks (Facebook), peer production networks (Wikipedia), information dissemination tools (Twitter), outlets for creativity (Flickr, YouTube) and mass collaboration platforms (Encyclopedia of Life). From a research perspective it is no longer sufficient to look at interaction and information seeking through the narrow prism of individual or local small group activity on relatively small websites; but in order to fully realize the potential that exists in large scale online environments we need to adjust exiting research methods and formulate rigorous methodologies that will enable researchers to go beyond data trails left from users’ interactions. Qualitative methods offer some of the most compelling options. Ethnography, for example, with its hallmark “thick descriptions” [4] of intricate social structures, behaviors, symbols, and language [5], offers researchers an opportunity to go beyond mere traces of interaction and activity, and gain a better understanding of implicit concepts of motivation, meaning and attitudes that are sometimes different or divergent from explicit behaviors.

Given the rapid development of large scale online environments, the question remains, why has ethnography not been more widely adopted to study them? A general answer to this question is that these environments present many challenges to qualitative researchers wishing to remain true to the essence of qualitative methods, and especially ethnography. Some of these challenges are:

**Challenge 1 – Scope and size**

The immense scope of the data, comprised of interactions between millions of users, requires the researchers to make *a-priori* difficult decisions about the segments of data that are pertinent for addressing their research questions. For example, the boundaries of the field are difficult to chart. Furthermore, Hammersely [6] warned against the notion of producing an understanding of a culture when looking at just a sub-group of that culture – a warning that is even more relevant when the scope and size of the field are impossible to encompass, and where assuming knowledge of an entire network, community or practice from a focused study of one of its sub-section is somewhat speculative.

**Challenge 2 – Choosing the unit of analysis**

Traditionally, the unit of analysis in ethnography – was a culture, a community, and in the case of early online ethnographies - a group of people that were brought together to use a specific platform by a certain shared activity or goal. Large scale online environments require renewed thinking about what is the relevant unit of analysis and how do we find it among the millions of users and artifacts. When Anderson coined the term “long tail” [7] he noted that although heterogeneity flourishes in the short run, ultimately homogeneity prevails. This is true from the perspective of ethnographic research in large scale online environments as well. Although theoretically such environments enable us to look at all kinds of online behavior, selection will result, in many cases, in data collected from the most prominent and talkative users or popular pages. The structure of social networks and other collaborative environments does little to enable access to the singleton or small groups that are not highly connected but may present an important phenomenon nonetheless.

**Challenge 3 – Interface structure**

Many large scale online environments are based on a matrix-like structure that augments layers of ego-centric personal ties and content that transcends several layers of the human-computer interface. For instance, on YouTube, users can friend each other, become connected by their reciprocal comments and topical interaction, but they can also provide content that is a fertile ground for ethnographic exploration; the same structure can be seen on other user-generated content sites. Although these layers are sometimes inter-related, they yield different questions and provide different data. Had these environments been on a smaller scale, as they were at the time online ethnography was first emerging, it would have been relatively easy to capture the activity on both layers. Yet again size makes such exploration extremely complex.

This begs a related, but different, issue: that of the nature of participant observation. Atkinson and Hammersley [8] suggest a continuum of participant observation, where the ethnographer has to be intellectually committed and immersed in the studied environment, and at the same time he can be known as a researcher to all those being studied, some of them, or none at all. Large scale online environments create a state where the option of the researcher being known to all those who are studied, is all but impossible. In most cases it is even difficult to make the researcher known to most of the people studied: although the researcher can take part in the activities happening on the site, through posting, connecting, sharing information and making her presence visible, large scale online environments will not warrant the researcher the same visibility she will have on smaller online spaces.
Due to technological advances, commercial reasons, competition, improvement efforts and – not the least – trendiness, large scale online environments constantly change. While changes to the interface can be beneficial to researchers (e.g. introduction of new tools), sometimes allowing serendipitous findings that would have otherwise been ignored, they make long or even short-term immersion in the field problematic. A tool is here today, gone tomorrow, and along with it disappears all the interaction and content that is produced, and the links that tie them to other content and users (e.g. Friendster and LiveJournal which were incredibly popular not long ago became obsolete and deserted rapidly; Facebook changes its interface and privacy policy periodically; and, YouTube went through several major interface changes in the past year alone). Every change in the interface causes a shift in the content and the structure, and alters the research context. A different type of change occurs when users manipulate the content they provide or their connections, or remove them altogether, independently of other users’ actions. In doing so they leave fragmented interactions, orphan responses, and social networks with missing links. While these actions carry meaning and can present an important facet of a living and breathing online culture, they simultaneously hinder the researchers’ ability to gain insights into the interaction that preceded this change and the meanings that are embedded in it.

**Challenge 5 – Ethics**

In large scale virtual environments there are also ethical issues to consider as in any research study, but because of the scale of participants’ distribution - physically, temporally and geographically, it may be more difficult to apply commonly used ethical practices.

**FACING THE CHALLENGES**

We have outlined several challenges to conducting qualitative work, and specifically ethnothography, within the extreme conditions of large scale online environments. One can argue that these challenges are not novel, that similar challenges faced researchers doing offline and online qualitative work for a while, as both the offline world and the online one are complex systems that necessitate meticulous choices of field, scope and context. However, we believe that today’s large scale environments certainly exacerbate known issues while creating some new ones. Some ways to face these challenges can be through borrowing principles from other methods, and combining qualitative work with other methods. For example, adopting a multi-sited ethnographic perspective [9, 10] and treating one large scale online environment as a multi-faceted site. This necessitates multiple entry-points, fragmented micro-accounts that build up into a macro-understanding of a bigger “world system”, and juxtaposing layers of interaction, will allow researchers to justify selection choices of sub-networks, specific environments or informants, and the particular viewpoints that are garnered through this type of research.

Other avenues may be combining qualitative work with structural analyses (e.g. social network analysis, log analysis and visualizations) or with natural language processing tools, in order to identify points of interest within the researched landscape, highlight clusters of activity and interpersonal ties or personal networks of informants. They can also surface phenomena that will not be observable though the inherently limited scope of an qualitative inquiry.

These are but a few ways to address some of the challenges outlined above (a fuller discussion of these issues is available in a paper to be presented at the 2012 iConference [11]). We hope to have the opportunity to broaden the discussion of both challenges and opportunities with the community of HCI qualitative researchers at the workshop.

**ACKNOWLEDGMENTS**

Our work is supported through NSF grants: IIS 11019993 and SES 0968546.

**REFERENCES**


