

# “Un-Googling”: Research Technologies, Communities at Risk and the Ethics of User Studies in HCI

Irina Shklovski  
IT University of Copenhagen  
Copenhagen, Denmark  
[irsh@itu.dk](mailto:irsh@itu.dk)

Janet Vertesi  
Department of Sociology  
Princeton University  
[jvertesi@princeton.edu](mailto:jvertesi@princeton.edu)

**Recent increase in volume of qualitative work on transnational technologies, HCI for development, virtual communities, and collaborative systems across a range of areas has resulted in focus on user communities whose very uniqueness may be of interest to HCI, but whose exposure in a research setting presents real threats to those individual's or community's livelihoods, work, or civil liberties. As the tools of research dissemination increasingly make scholarly publications more easily accessible to the public and other entities outside the academic community through simple search engines, scholars must grapple with new challenges to the ethics of exposure. We present a case-study of un-Googling publication of research results and consider potential problems with such an approach to minimizing risk to research participants.**

*Keywords: Un-googling, ethics, exposure, anonymity, publication*

## 1. INTRODUCTION

The HCI community is engaging ever more closely with users in a range of unexpected locations. Studies of communities or users whose perspectives have been rarely heard in HCI bring new design opportunities. However, they also pose new questions with respect to our engagements with these communities, as our own tools of research and dissemination may place our studies and our participants' safety at risk. Moving beyond simply “privacy” or Institutional Review Board applications, such engagements raise practical, ethical, methodological and publication concerns among researchers and practitioners.

As the tools of research dissemination are changing with ever more publication venues putting content online, scholars are arguing for more open-access publications. Open-access publishing is clearly beneficial, as it makes research results available to scholars beyond the rich educational institutions of the Western world. Yet as open-access publishing becomes more common and our Internet-based search tools continue to improve, sensitive findings reported in publication are made locatable and available to an audience far beyond the academic world. Traditionally, empirical scholars could report potentially controversial and even negative findings from their field sites and such publications rarely came to the attention of research participants themselves or the power struc-

tures within which they resided. In the post- Google era, however, reliance on obscurity is no longer an option.

In this paper we begin to articulate an approach to the ethical threat that new tools of research dissemination may pose for our ongoing engagements with communities at risk. Our goal is to outline the new challenges of this space, considering attempts at search engine anonymization (which we call “Un-Googling”) as an area of particular relevance to practical and ethical considerations of implementation and presentation of HCI research based on such populations. These concerns are relevant to a range of user studies research encompassing quantitative, qualitative and design research approaches. However, here our example and discussion focuses on qualitative approaches as perhaps most acutely implicated in these concerns.

## 2. EXPANDING “COMMUNITIES AT RISK”

Communities at risk have been the subject of anthropological and sociological study for decades, but are only recently becoming of interest to HCI researchers for a variety of reasons. They may present a case of “extreme users” whose specific needs demand that we broaden our design considerations. These users may inhabit a variety of power structures born of historical contingencies and socio-economic concerns. They may be Thai orphans (Williams, 2011), Nigerian spam artists (Burrell, 2008), or homeless people in the United States (Le Dantec and Edwards, 2008) who engage with global systems from powerless socio-economic or political positions. These users' ac-

cess to technologies of interest is often tricky, uncertain, inconsistent, at times even less than legal.

Considerations of risk go beyond studies of marginalized, underprivileged or underserved communities. For example, in countries where censorship is an active state practice, participation in technical communities may be limited to the economically privileged in society, but can articulate an implicit or explicit anti-government position (Shklovski and Kotamraju, 2011, MacKinnon, 2008). This is particularly acute in situations where users are under more direct government control than Western researchers expect. Yet studies of such communities at risk present powerful results for HCI not only in terms of technical work under such conditions, but also in terms of how such practices must be our consideration even “right here.” (Taylor, 2011)

In environments with fewer legal restrictions, publication of results concerning high-tech companies or organizations may affect financial support, status, or stock options. For example, high-visibility government science teams (Vertesi and Dourish, 2011) require continued funding and public support, which rests on maintaining a positive public image. Where analysis of challenges in internal processes may lead to important design insights for new collaborative technologies in other environments with similar status, exposure of these same issues through publication may impact the perceptions of these organizations and groups for stakeholders outside the field of HCI.

We argue that making certain kinds of practices public, even if they seem mundane to the academic community can be a political act, whether intentional or not. When we present the practices we observe to HCI audiences for the purposes of research and design, we also expose the people who participate in our studies to other types of scrutiny. Although a number of documents have recently addressed ethical issues in research, such as the Ethics Guide produced by the Association of Internet Researchers (Ess and The AoIR Ethics Working Group, 2002), much of the concern detailed in these documents focuses on informed consent and the use of information disclosed by the participants. But the users that we study, interview, and cite may be at risk of jail time, fines, family dishonor, or forms of state punishment because of the ideas we as researchers might publish rather than due to the information they disclose. Here current ethical guides fall short.

While it might seem that practically any community or user can now be conceived of as at risk in some way, this is not only because there is a rapidly growing category of such users in the world and in HCI research. It is also a product of our methods

and the socio-technical conditions of our own research practice and dissemination. This is why, we argue, concerns for ethics in data collection are only part of the story. Our publications in our own research communities are increasingly available to anyone with access to Google, making what and how we publish based on our work with communities at risk an additional major concern. As a research community attuned to issues of data integrity and the rapid increases in data availability, we must consider the ethics of new research methods such as access and use of public online behavioral data in social network sites (Zimmer, 2010). We discuss a case study of a particular real ethical dilemma encountered in the course of research in such a “community at risk” as an example of such a tension in the practice of research. We then describe four methodological and ethical challenges that we faced under these new sociotechnical research conditions.

### **3. CASE STUDY**

In 2009 and 2010 the first author conducted fieldwork in a country where the authoritarian government engages in significant surveillance and blocking of Internet access for its citizens. While the research itself was focused on the role of technology in relational maintenance, the first author was struck by the palpable fear of retribution from the state expressed by some of the most active online content contributors, such as bloggers, discussion forum participants and writers. Interviews and informal interactions with owners of local blogging platforms, forums and informational sites revealed a detailed picture of the threat of surveillance and the practices of regular users. These findings highlighted the external structural conditions relevant to how and why users contribute content online, presenting a range of practices users engaged in to manage their exposure to the whims of the state.

Given the increasing reliance of social technologies on user-generated content, the relevance of these findings to the HCI community was clear. However, the fact that the local IT community under study was very small made specific participants identifiable through the attitudes they expressed in the quotes used as supporting data in the paper regardless of anonymization efforts (van den Hoonaard, 2003). The project in question was also in the first phase of a multi-year study and the first author herself could potentially become a person of interest in a state that publicly denies limiting freedom of speech while clearly doing so online. These concerns became relevant because the paper, if accepted, would be placed in a public online, keyword-searchable repository. The question then became: how might we share important research in-

sights with the HCI community while protecting participants and the integrity of the ongoing project?

### 3.1 The IRB, informed consent and its limits

In conducting the study the first author followed the practices required by an Institutional Review Board (IRB) of an affiliated North American university as well as acknowledged and adapted to the conditions and expectations of the population under study. Thus all study participants were presented with a study information sheet that detailed the goal of the study, participants' rights, and steps that will be taken to protect participant's confidentiality in the course of data collection, storage and use. No signed informed consent was collected in order to minimize potential for exposure. The researchers worked hard to ensure that study participants retained control over all research-related encounters.

Yet when it came to publication, it became clear that (IRB) requirements were not adequately tailored to addressing the issue of presentation of research results. Participants in this study were aware that their statements were either written down or recorded with their permission during interviews. They were also aware of the researcher's intent to publish results of her investigation. The information study sheet detailed the names and affiliations of the researchers that would have direct access to raw data. Researchers used pseudonyms when using representative quotes in academic publications and limited the amount of demographic information revealed about participants. Yet these efforts did not seem enough.

### 3.2 The challenges of anonymization

Challenges to common anonymization practices apply well beyond the individual level described above. For example, anonymization presents challenges for suites of authors from a company like IBM who, although they describe their study in writing as located at "a large technology company," cannot expect actual anonymity. However, the specificity of the socio-technical context under study is often too important to the story of why these users matter to leave out for the purposes of publication. As research in such field-sites cannot be kept invisible from other employees or managers, published results may directly imperil participants' work-life. This then becomes not simply an issue of "privacy" but one of accountabilities of research practices (Troshynski et al., 2008).

This may not, at first glance, be an issue where the study is concerned with technology users in a particular country rather than a more bounded space of an organization or an online interaction space such as a forum or a game. Yet as HCI follows computing devices from offices into the world, the anonymized encounters of users at large may not

provide shelter for user privacy either. In our case, although the population of the country is in tens of millions, less than 30% were Internet users. This made the communities of users who generated online content relatively small. Despite the fact that fieldwork was conducted in three cities hundreds of miles apart, the user communities under study were tractable, small, and visibly connected online. Thus individuals could be identified just by the sentiments they expressed rather than through direct quotes. Moreover, the very fact that research was conducted through users' face-to-face encounters with a foreign researcher, made these encounters more locally visible.

### 3.3 Un-Googling published work

Our solution in this case was to anonymize not only the individual participants but also the country where the research was conducted. To fully limit the paper's traceability through online means, we removed any references that could potentially identify the location. After all, there are plenty of authoritarian governments in the world that engage in surveillance and blocking on the Internet. We called this practice *un-Googling* the paper. Interestingly, this approach generated criticism from some of the paper's reviewers who argued that: "*The anonymity of the research site hampers the informativeness and trustworthiness of the paper. Research in these restrictive states is relatively rare, that any kind of cited information would be a real contribution to the field.*" Nevertheless, the paper was ultimately accepted for publication in its anonymized form.

Certainly anyone in the academic community who knew the first author and was familiar with her recent research could easily identify the country in question upon discovering the paper. The goal was not to hide the country under study from the academic community. On the contrary, this research was performed in a relatively under-studied area of the world and revealing the country name likely would have made the contributions of the paper that much more interesting and compelling. The paper was anonymized with a simple goal in mind: to prevent individuals outside the academic community, more specifically, the power-elite of the country in question, from perhaps accidentally coming upon the paper by inputting 'country name' and 'surveillance' or 'internet blocking' into an internet search engine. While the paper itself is freely accessible to the public at large, we hope it is now less likely to be used against the individuals who participated in this research. This was a kind of "the least we could do" solution.

### 3.4 The Ongoing Work of Un-Googling

*Un-Googling* does not end with simple anonymization at the point of publication. In order for this prac-

tice to be successful, the first author has had to meticulously monitor when and how the paper, her name and the name of country might have been mentioned together in a range of documents. For example, putting this paper on her CV along with other papers that do mention the country name would eliminate any benefit gained from the initial *un-Googling*. Mentioning this paper as part of the specific project on the authors' website would undo the anonymization as well. The care with which this paper must now be handled is an on-going process rather than a one-time issue to be solved.

Research within socio-technical communities presents awkward challenges. Ongoing connections with our communities of study through Facebook friendships, or live Twitter feeds of our research talks may put the researchers themselves in a difficult situation. Whatever the level of involvement with the users and communities under study, the researcher and the participants can remain virtually connected well past the duration of the research project. In fact, such digital connections may appear after the interviews and observations have been completed, putting the researcher in a tricky situation. Many of the participants in our study were highly educated individuals imminently interested in what the findings the author published. We continue to receive requests for papers and inquiries about our future activities. The question then is can this *un-Googled* paper be shared or will that once again undo our efforts?

## 5. LOOKING FORWARD

HCI is at a turning point in socio-technical systems research. A wide variety of user communities continue to attract research interest, and new techniques and tools are being applied to better understand those communities. HCI has long been aware of the importance of ethical considerations to research practices (Anderson, 1992). Ironically, as we improve technologies of search discovery and fight for open access to the outcomes of our publicly funded research activities, we also complicate the ethical considerations of our research practice and the reporting of results. Research conducted on the Internet must confront the fact of archived digital traces of online interactions, often easily discoverable through search, and of the legal complications of Terms of Use (Zimmer, 2010). This presents new challenges to the ethics of exposure. In qualitative work, complete anonymization is impossible precisely because of the level of interpersonal involvement required. The challenge now is not only how to present one's field site to an academic community but also how to balance revealing the necessary specificities of the site with the needs for anonymization. If our methods and our research dissemination practices place users at risk, then as

a community we have a responsibility to address these issues and to present some resolution, both for the purposes of the continued presentation of robust results in HCI publications, and for continued emphasis on user safety in the very communities we seek to support.

## ACKNOWLEDGEMENTS

This research was funded by NSF grants SOCS-0968616 and IIS-0917401. Any opinions, findings, conclusions or recommendations expressed in this material are those of the authors and do not reflect the views of the National Science Foundation

## REFERENCES

- Anderson, R. E. (1992) Social Impacts of Computing: Codes of Professional Ethics. *Social Science Computer Review*, 10, 453-469.
- Burrell, J. (2008) Problematic empowerment: West african internet scams as strategic misrepresentation. *Inf. Technol. Int. Dev.*, 4, 15-30.
- Ess, C. & The AoIR Ethics Working Group (2002) Ethical decision-making and Internet research: Recommendations from the AoIR ethics working committee. <http://aoir.org/documents/ethics-guide/>, The Association of Internet Researchers (AoIR).
- Le Dantec, C. & Edwards, W. K. (2008) Designs on dignity: perceptions of technology among the homeless. *Proceedings of CHI 2008*. Florence, Italy, ACM.
- Mackinnon, R. (2008) Flatter world and thicker walls? Blogs, censorship and civic discourse in China. *Public Choice*, 134, 31-46.
- Shklovski, I. & Kotamraju, N. (2011) Online contribution practices in countries that engage in internet blocking and censorship. *In proceedings of CHI 2011*, 1109-1118
- Taylor, A. S. (2011) Out there. *Proceedings of CHI 2011*. Vancouver, BC, Canada, ACM.
- Troshynski, E., Lee, C. & Dourish, P. (2008) Accountabilities of presence: Reframing location-based systems. *in proc of CHI 2008*. NY, ACM.
- Van Den Hoonaard, W. C. (2003) Is Anonymity an Artifact in Ethnographic Research? *Journal of Academic Ethics*, 1, 141-151.
- Vertesi, J. & Dourish, P. (2011) The value of data: considering the context of production in data economies. *Proceedings of CSCW 2011*. Hangzhou, China, ACM.
- Williams, A. (2011) Transnational Networks and Slum Mobilities. *Paper presented at Transnational HCI workshop, CHI 2011*.
- Zimmer, M. (2010) "But the data is already public": on the ethics of research in Facebook. *Ethics and Information Technology*, 12, 313-325.