

Challenges to Effective Anonymization

Julia Bullard

University of Texas at Austin
Austin, USA
julia.a.bullard@gmail.com

INTRODUCTION

My research follows the actions and design choices of a large and unique online community. In this workshop, I would like to discuss our responsibility to protect the identity of the communities we study and the strategies to do so when the design choices under investigation are identifiably unique.

STUDYING LARGE & UNIQUE COMMUNITIES

I study the collaborative design of the organizing schema for an online fan work repository. Relevant details about the design process—that it is performed by volunteers, that it modifies a user-generated folksonomy into a kind of controlled vocabulary—are sufficient to identify the site among its peers. I have made the choice to avoid naming the community in publications based on my ethnographic research, but my participants and I are aware that this achieves little more than plausible deniability with regards to their collective identity. There would be no doubt from members of this community that my anonymized site is their own.

Though it detracts from the rich picture developed through ethnographic work, it is possible in a single publication to minimize identifying details about the community. Particularly given the volume of scholarly work that is the intended output of prolonged ethnographic research, triangulation across studies would negate such an effort. For example, the combination of details from publications on different elements of the design process, such as volunteer recruitment, controversial decisions, and assessments of outputs, would be sufficient to identify the site.

Similar challenges face social science researchers who study large organizations with few peers, such as in automotive manufacturing [2] and state institutions [1]. In such research, the organization may be unique in its domain, structure, and reach, while its members as participants are still due the protection of confidentiality and anonymization. An additional layer of difficulty in protecting the identity of online participants is the visibility of public-facing documents and their indexing in search engines. For example, in quoting community guidelines we may make even obscure communities identifiable through a simple Google search.

By presenting my case to the workshop, I hope to spark discussion on how to design and write-up similar online communities research responsibly as well as how to fairly

disclose to participants the barriers to effective anonymization. I recognize as a broader challenge the need to develop sufficient resources and oversight to address this and other ethical challenges for our research community.

SHAPING ETHICS POLICIES

Among the most pressing questions facing our research community I would include: What is our role in shaping ethics policies in our research institutions? Since institutional review board policies for online research are developing and in flux, researchers are in position to shape reviewers' understandings of the ethical landscape. To what extent should we educate reviewers with regards to risks at the expense of effectively advocating for our own study designs? For example, though aggregated, anonymized public data may be judged according to an existing standard of publicly available information, analysis applied to such data potentially identifies individuals while connecting them to sensitive behaviors. Researchers setting out on such a study can try to persuade reviewers that their and similar studies require more thorough consideration given the unexpected outcomes of analyzing aggregate data.

If IRBs, at the present time, are not adequately equipped to address the complexity of ethical issues regarding online research, how can we as a community provide the oversight necessary to encourage ethical research? This workshop can advance our networked resources for providing guidance and review. In particular, I hope to discuss how we can work within existing processes, such as IRB, to inform how they respond to new types of research design, and how we can develop parallel processes to inform research design norms.

AUTHOR BIOGRAPHY

Julia Bullard is a doctoral candidate at the University of Texas at Austin's School of Information. Her co-advisors are Melanie Feinberg and Diane Bailey. Through knowledge organization, infrastructure studies, and values-in-design frameworks, she studies the choices designers make in the construction of classification systems. Her dissertation, an ethnographic study of volunteer classification design for a fan work repository, focuses on how designers manage conflicting responsibilities to authors, readers, and the community's ethical position.

REFERENCES

1. Jane Gruning, Julia Bullard, and Melissa Oceppek. 2015. Medium, access, and obsolescence: What kinds of objects are lasting objects? *Proceedings of the 33rd Annual ACM Conference on Human*

Factors in Computing Systems - CHI '15, ACM Press, 3433–3442.
<http://doi.org/10.1145/2702123.2702238>

2. Paul M. Leonardi and Diane E. Bailey. 2008. Transformational technologies and the creation of new work practices: Making implicit knowledge explicit in task-based offshoring. *MIS Quarterly* 32, 2: 411–436.