
Advancing an Industry/Academic Partnership Model for Open Collaboration Research

Jonathan T. Morgan
Wikimedia Foundation
jmorgan@wikimedia.org

Aaron Halfaker
Wikimedia Foundation
ahalfaker@wikimedia.org

Dario Taraborelli
Wikimedia Foundation
dtaraborelli@wikimedia.org

Tim Hwang
Imgur
tim@imgur.com

Sean Goggins
University of Missouri, Columbia
gogginss@missouri.edu

Abstract

We propose a full day workshop focused on characterizing areas of research within the domain of Open Collaboration Systems (OCS's) where partnerships between academic and industry researchers can both increase our scientific understanding of OCS's and also support those systems through research. This workshop's purpose is to bring together researchers on both sides of the "data divide" to identify current challenges and opportunities for future research within these research areas, and to develop a preliminary set of requirements for improved resource sharing and collaboration between OC enterprises and academic research institutions.

Author Keywords

Open Collaboration, Open Data, Cyberinfrastructure

ACM Classification Keywords

H.5.m. Information interfaces and presentation (e.g., HCI)

Introduction

An Open Collaboration System (OCS) is an online environment that supports the collective production of an artifact through a technologically mediated

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the Owner/Author. Copyright is held by the owner/author(s).

*CSCW '15 Companion, Mar 14-18 2015, Vancouver, BC, Canada
ACM 978-1-4503-2946-0/15/03.
<http://dx.doi.org/10.1145/2685553.2685559>*

collaboration platform, presents a low barrier to entry and exit, and supports the emergence of persistent but malleable social structures[2] The open collaboration process as supported by information technology has emerged as a powerful and effective form of content production. The so-called "Web 2.0" generation of user-generated content systems distribute production work by blurring the line between consumer and producer.

Open collaboration systems (OCS's) like Wikipedia, Reddit, Imgur, etc., have shown that networked communities of volunteer contributors can produce and maintain immense, valuable public resources.

Open collaboration work patterns have become the subject of focused study in the CSCW community over the last decade. Through a growing body of work exploring the nature of open collaboration, the research community has come to recognize novel opportunities and problems that openness brings with it. For example:

- Distributed production and the need for quality control[6]
- Maintaining volunteer motivation to participate[5]
- Distributing the processes for articulation[1]
- Community maintenance via socializing newcomers[3][4]

These themes have lead to both interesting threads of work in academic spaces and substantial concern over the viability of real world communities for current practitioners. However, while these studies have described many important phenomena related to

OCS's, findings from these academic studies do not always permeate the boundaries between scholarship and practice. Furthermore, despite the proliferation of both OCS research and OCS's, many other important phenomena related to peer production are not yet fully understood.

We propose a full day workshop¹ designed to explore possible models for formal partnership between OCS scholars and practitioners that could both increase our scientific understanding of open collaboration systems and also support those systems through state-of-the-art research that can help provide guidance for designing and sustaining successful Open Collaboration Systems.

Open Collaboration and the Data Divide

Both academic researchers and industry researchers have resources and expertise that are necessary to further our understanding of open collaboration. However, these two camps of researchers sit on either side of a "data divide". Academic research progress is often stymied by lack of access to data from OCS's and the ability to perform interventions within real, live systems. Similarly, better mechanisms for communicating the implications of CSCW research for creating and maintaining successful OCS's are needed if industry researchers and OCS enterprises are to benefit from CSCW research into the systems they design and maintain, and the volunteer communities they support.

Finally, the broader research community needs more opportunities to discuss how we can address

¹ https://meta.wikimedia.org/wiki/Research:CSCW15_workshop

controversial issues, such as the use of potentially sensitive user data for research purposes and the deployment of design interventions inside living systems to test research hypotheses, in a principled and responsible way. OCS enterprises and academic Institutional Review Boards would both benefit from better communication across this data divide as they attempt to develop policies that address the ethical, legal, and practical considerations around data sharing while maintaining the benefits of academic/industry research collaborations.

On either side of the divide, academics have the mandate to build knowledge and practitioners have the mandate to benefit their community. But this divide is not necessary. It's our firm belief that, through improving the coordination between these two groups, we can achieve both improved Theory and improved practice.

This workshop builds on an informal "wiki breakfast" session held at CSCW 2014, in which the first three authors facilitated a roundtable discussion with 20 members of the CSCW community on the state of Wikipedia research and the research needs of academic researchers who study similar systems. This informal discussion lead us toward exploring strategies for improved collaboration between the Wikimedia Foundation and the academic research community. The workshop we propose to hold at CSCW is intended to explore improved collaboration more generally.

Workshop plan

Participant survey. In the months leading up to the CSCW 2015 conference, potential participants will be

surveyed about their interests, concerns and view of the state-of-the-art in OCS theory, as well as the expertise and resources they have available. This survey will serve as a kind of application from which we will select participants and construct a set of discussion themes based on common concerns/interests.

Lightning talks. In the mid-morning of the workshop day, we will invite some participants to give brief lightning talks in order to anchor and stimulate subsequent discussion. Industry practitioners (members of the organizing team as well as other invited industry participants from outside of the CSCW community) will be invited to discuss their OCS, their data resources, and the challenges that they face. Selected academic research participants will be also invited to discuss their study of OCS phenomena, the data resources they make use of in the research projects, and the implications of their findings for the practice of creating and sustaining an OCS.

Roundtable discussion. In the late morning and through lunch, participants will discuss the common themes we will have identified from their survey responses (as well as new, interesting themes that come to light in the morning). Examples of discussion themes include any of the research areas described in the first section of this document, as well as other partnership considerations such as:

- dissemination of research findings
- data access (public dumps, databases and APIs)
- subject recruitment and field experiments

Break-out group sessions. In the early-mid afternoon, participants will break out into small discussion groups for focused discussion on themes of their choosing. We will make use of the open space protocol² for enabling participants to organize around which themes & discussions that they find most relevant. We will advocate for including both academic and industry practitioners in each group. Groups will be asked to produce a digital artifact capturing the product of their discussion, which will be published on the workshop website. Examples of artifacts may include:

- methodological requirements (data, intervention, etc.) for exploring a particular theme
- a list of literature relevant to that theme and open questions
- descriptions of desirable industry/academic collaboration practices

Report out/summing up. At the end of the day, groups will be asked to complete their artifacts and present them to the larger group (5-10 minutes/theme group). As a final activity, we will ask every participant to make a statement about what they learned from the workshop and how it might affect their work (1-2 minutes/person).

References

- [1] Forte, A., Larco, V., & Bruckman, A. (2009). Decentralization in Wikipedia governance. *Journal of Management Information Systems*, 26(1), 49–72.

- [2] Forte, A., & Lampe, C. (2013). Defining, Understanding, and Supporting Open Collaboration: Lessons From the Literature. *American Behavioral Scientist*, 57(5), 535–547.

- [3] Halfaker, A., Geiger, R. S., Morgan, J. T., & Riedl, J. (2013). The Rise and Decline of an Open Collaboration System: How Wikipedia's reaction to sudden popularity is causing its decline. *American Behavioral Scientist*, 57(5), 664–688.

- [4] Gary Hsieh, Youyang Hou, Ian Chen, and Khai N. Truong. 2013. "Welcome!": social and psychological predictors of volunteer socializers in online communities. In *Proceedings of the 2013 conference on Computer supported cooperative work (CSCW '13)*. ACM, New York, NY, USA.

- [5] Ling, K., Beenen, G., Ludford, P., Wang, X., Chang, K., Li, X., Cosley, D., Frankowski, D., Terveen, L., Rashid, A. M., Resnick, P. and Kraut, R. (2005), Using Social Psychology to Motivate Contributions to Online Communities. *Journal of Computer-Mediated Communication*, 10:00.

- [6] Andrew G. West, Sampath Kannan, and Insup Lee. 2010. Detecting Wikipedia vandalism via spatio-temporal analysis of revision metadata?. In *Proceedings of the Third European Workshop on System Security (EUROSEC '10)*. ACM, New York, NY, USA

² <http://www.mindviewinc.com/Conferences/OpenSpaces>