Crowdsourcing for science: understanding and enhancing SciSourcing contribution

Position paper: ACM CSCW 2010 Workshop on the Changing Dynamics of Scientific Collaborations

Oded Nov
Polytechnic Institute of New York University,
New York, NY, USA
onov@poly.edu

Ofer Arazy
School of Business
University of Alberta
Edmonton, AB, Canada
ofer.arazy@ualberta.ca

David Anderson
Space Sciences Laboratory
University of California, Berkeley
Berkeley, CA, USA
davea@ssl.berkeley.edu

INTRODUCTION

Recent years have seen an increase in the number of citizen science initiatives, which harness the contribution from large numbers of volunteers to support a scientific project, much like other internet-based crowdsourcing applications. In some cases, citizen science projects are based on volunteer computing, where people contribute their computer resources to the project (rather than contributing their manual effort). The most notable example of these projects is SETI@home, where contributors allocate their personal computing resources to be used for detecting intelligent life outside Earth. Other citizen science initiatives, on the other hand, rely on the manual effort of contributors, for example Galaxy Zoo, a web-based distributed analysis project where contributors classify images of galaxies, or the Citizen Weather Observer Program (CWOP) where volunteers monitor the environment and contribute weather data. Common to all these initiatives is the use of the internet as the primary platform for enabling distributed, volunteered, citizen science contribution. These initiatives represent a paradigm shift in scientific research, reducing substantially operating costs and increasing the involvement of the general public. We term this new breed of internet-based citizen science 'Science Sourcing', or simply 'SciSourcing'. In this workshop position paper we briefly review the field, identify gaps in the current literature, introduce our own research program in this area, and describe preliminary results from two recent empirical studies.

SciSourcing is based on two pillars: the first is computational - developing information systems that can manage, allocate and aggregate large amounts of distributed resources. The second pillar is behavioral: enticing a large number of people to contribute their resources, and creating contribution systems that will encourage continuous contribution. Understating why people voluntarily contribute various resources (computing resources, skills, time, and effort) to such projects contribution is, therefore, of paramount importance. However, while the computational aspect of SciSourcing received much research attention [e.g. 1, 2, 3, 9], the behavioral aspect remains largely unexplored. What do we know about the factors driving SciSourcing contribution? How can a large scale scientific project provide an environment that would encourage resource contribution from many volunteers? These are the primary questions our research program aims to address.

In recent years, crowdsourcing has emerged as a powerful approach for harnessing resources contributed by large numbers of geographically distributed individuals. Underpinning the sustainability of projects such as Wikipedia, YouTube, Flickr, and many others, is the willingness of individuals to voluntarily contribute information, time and skills [5, 6]. Consequently, researchers have investigated the factors driving contribution in a wide range of online settings [e.g. 6, 8, 10, 11, 13]. However, there are some important differences between these crowdsourcing and SciSouring projects. First, in SciSourcing there is a clear distinction between those benefiting from the aggregated contributions (i.e. the scientists who run the project) and the volunteer contributors. In contrast, in most community-based projects (e.g. Wikipedia) this distinction is blurred, such that contributors are often users (i.e. readers or viewers) of others' contribution. In fact, much of the research in community-based projects, e.g. open source software development, has emphasized this feature. Second, each person's SciSourcing contribution often constitutes only a small and unidentifiable part of the larger scientific research project. In addition, there is a substantial delay from when the contribution is made to the time when the output of the project (e.g. scientific publications) is made public. In contrast, in other crowdsourcing applications, a user's contribution (whether text, software code, or photos) is an identifiable piece that is often associated with the contributor and is immediately viewable once published. These differences – we argue – have implications for the factors driving contribution, and we suspect that the motivations for contributing for SciSourcing projects may

differ from the motives in other types of crowdsourcing projects.

The success of SciSourcing projects depends largely on distributed contribution by volunteers, just as much as they rely on the technical infrastructure [7]. To date, the question of why people contribute to SciSourcing projects has been largely overlooked. The goal of our research program is, thus, to advance the understanding of peoples' motivations for participating in SciSourcing projects. The primary research questions we aim to address are: (1) What are the factors driving participation in SciSourcing projects? (2) How do these factors affect project performance? (3) Are there ways in which the human-computer interface of SciSourcing application can be modified to enhance contributors' motivations and participation?

The different types of SciSourcing projects are characterized by the different levels of task granularity - "the smallest possible individual investment necessary to participate in a project" [4] – that is required from contributors. SciSourcing projects differ greatly in the their task granularity (see Figure 1), ranging from almost passive contribution in the case of volunteer computing, to more active and demanding contribution in web-based scientific analysis projects. In our conceptualization of motivations for SciSourcing contributions, we distinguish between projects with low task granularity (e.g. volunteer computing) and those with high granularity (e.g. web-based analysis), since granularity is directly linked to motivations [4].

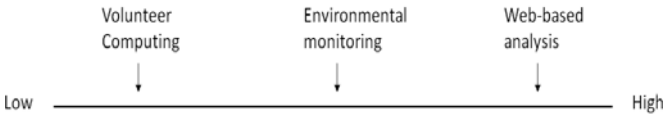

Figure 1. Contribution task granularity (based on [4])

Based on existing literature on crowdsourcing contribution and our preliminary interviews with leaders of SciSourcing projects, we have identified a set of factors that could potentially impact volunteers' motivations and participation levels. We classified these factors into three primary categories: Individual, Technological, and Contextual factors.

PRELIMINARY FINDINGS

We have performed two preliminary studies, where we correlated survey data of volunteers' motives, and system log data on the actual contribution levels of the survey participants. The first study [12] was conducted in SETI@home (volunteer computing; low task granularity) and the second was performed at CWOP (contribution of weather data; medium task granularity). Below we briefly report the results from these studies.

Comparing SciSourcing motivations with motivations examined in prior research on crowdsourcing, our SETI@home study [12] revealed that two motives that were shown to have significant impact in crowdsourcing participation – Enjoyment and Reputation – were not related to SETI@home contribution levels (i.e. the coefficients were 0.09 and 0.03 respectively and were not statistically significant). The Enhancement motivation (the satisfaction from seeing the project's findings published) was found to be positively related to contribution in a statically significant way (the coefficient was 0.18). It is interesting to note that, at the same time, this enhancement motivation received the lowest average score of the four motivational factors. The fourth motive we explored in this study – the Values motivation (the extent to which the contributor shares the stated goals and values of the project) - exhibited a statistically-significant negative effect on the outcome variable. A likely explanation is that sharing the project's objectives is a characteristic that helps to explain why people join the project in the first place, however once active contributors, sharing the project's objectives and values is not linked to contribution levels.

Two other factors – affiliation to a team and tenure – were also analyzed. Affiliation to a team was found to be significant related to the contribution level (the coefficient was 0.16), while tenure exhibited a statistically significant relation (the coefficient was -0.27) with contribution. The control variable, the number of computers allocated to the project, was also significantly related to the contribution level. Together, these variables explained 21.6% of the variance in the outcome variable (average daily contribution).

Based on the findings from the SETI@home study, we revised our conceptualization and have decided to focus on a slightly different set of variables. We have tested this revised conceptualization at the Citizen Weather Observation Program, and found that the most salient factors for driving participation continuance intentions were (in order of importance): learning new information, values, norms, and intrinsic motivations (e.g. enjoyment). Other factors – e.g. identification and reputation – proved insignificant.

These preliminary findings demonstrate not only that motivations for SciSourcing contribution differ from the reported motivations in non-science crowdsourcing projects, but that there are some substantial differences between the different SciSourcing projects, based on their task granularity. These findings highlight the need to develop a better understanding of the factors that influence SciSourcing contributors. As part of our ongoing research program, we continue exploring the behavioral aspects of SciSourcing participation in a variety of settings. We encourage others to join us in researching this important phenomenon.

REFERENCES

1. Anderson, D., Cobb, J., Korpela, E., Lebofsky, M.,Werthimer, D. 2002. SETI@home: An Experiment in Public-Resource Computing. Communications of the ACM, 2002.
2. Anderson, D. Volunteer computing: Planting the flag, in: Proceedings of the PCGrid 2007 Workshop, Long Beach, CA, USA, March 2007
3. Anderson D. and Fedak. G. The computational and storage potential of volunteer computing. CCGRID'06, 2006, 73-80.
4. Benkler, Y. The Wealth of Networks. Yale University Press, 2006.
5. Butler, B. Membership size, communication activity, and sustainability: A resource-based model of online social structures. Info. Sys. Res., 12(4) 346-362, 2001.
6. Chiu, C. Hsu, M and Wang. E. Understanding knowledge sharing in virtual communities: an integration of social capital and social cognitive theories. Decision Support Syst. 42(3):1872-1888, 2006.
7. David, P. (2005). Towards a cyberinfrastructure for enhanced scientific collaboration: Providing its "soft" foundations may be the hardest part. International conference on Advancing Knowledge and the Knowledge-Economy, National Academy of Science, Washington, DC.
8. Jeppesen, L. and Frederiksen, L. Why do users contribute to firm-hosted user communities? Organ. Sci. 17(1) 45–63, 2006.
9. Kondo, D. Anderson, D and McLeod, J. Performance Evaluation of Scheduling Policies for Volunteer Computing. International Conference on e-Science, pp. 415-422, 2007.
10. Ma, M. and Agarwal, R. Information Technology Design, Identity Verification, and Knowledge Contribution in Online Communities. Info. Sys. Res. 18(1), 42–67, 2007.
11. Nov, O. What motivates Wikipedians? Comm. of the ACM, 50(11):60-64, 2007.
12. Nov, O., Anderson, D. and Arazy, O. Volunteer Computing: a Model of the Factors Determining Contribution to Community-Based Scientific Research. *Proceedings of the 19th International World Wide Web Conference (WWW 2010).* Raleigh, NC, 2010.
13. Nov, O., Naaman, M. and Ye, C. What Drives Content Tagging: The Case of Photos on Flickr. *Proceedings of CHI 2008: ACM Conference on Human Factors in Computing Systems*, Florence, Italy, 2008.