Crowdsourcing for science: understanding and enhancing SciSourcing contribution

Submitted to the CSCW workshop on the Changing Dynamics of Scientific Collaborations

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 incentive 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. WWW 2, 3, 4, 19, 30, 46], 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 [Brian S. Butler, Chao-Min Chiu, J. Koh]. Consequently, researchers have investigated the factors driving contribution in a wide range of online settings [e.g. Gee-Woo Bock, Irene Y. L. Chen, Chao-Min Chiu, Lars Jeppesen, Meng Ma, Nov 2007, Naren Peddibhotla, Joachim Schroer]. However, there are some substantial 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. Third, many rowdsourcing projects, e.g. open source software development and Wikipedia, are inherently social, while contribution to SciSourcing projects is not. 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. For example, since social factors play an important role in crowdsourcing projects more than in Scisourcing, social motives are likely to have a smaller impact on contribution to SciSourcing.

The success of SciSourcing projects depends largely on distributed contribution by volunteers, just as much as they rely on the tecnicaal infrastructure [Paul David]. 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 can be characterized by the different levels of task granularity - "the smallest possible individual investment necessary to participate in a project" [Benkler 2006] – that is required from contributors. SciSourcing projects differ greatly in the their task granularity (see Figure X), 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. SETI@home) and those with high granularity (e.g. Galaxy Zoo), since granularity is directly linked to motivations [Benkler 2006].

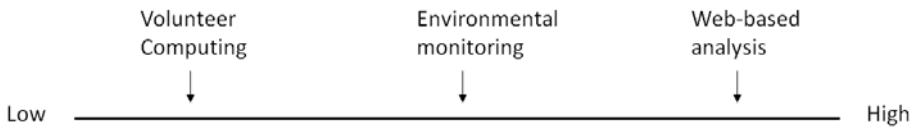

Figure 1. Contribution task granularity (based on [Benkler, 2006])

Based on existing literature on crowdsourcing contribution and our preliminary interviews with thet 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. A preliminary (and certainly not exhaustive) list of factors is presented in Table 1.

Possible driving factors		Description/Example
Individual Factors	Personal Motivations	e.g. intrinsic to extrinsic ; collective motives vs. norm-oriented motives vs. reward motives [e.g. Butler 2007, Hertel 2003]
	Team affiliation	Being part of a team within a project.
	Identification with the project, the team and other contributors	e.g. identification with project and / or identification with other project contributors, teams within the project.
	Demographics	e.g. age, sex, education.
Technology Factors	Participants' perceptions and beliefs about the project information system	e.g. ease of use, compatibility with systems used in the past.
	Project system's features and affordances	e.g. system interface, virtual co-presence of other contributors, individual/group ranking of contribution, feedback.
Contextual Factors	Task granularity	e.g. how much time does it take to make a minimal contribution.
	The objective of the project, and contributor's interest in it	e.g. research in a particular field such as astronomy.

Table 1. Potential driving factors: a preliminary list

Preliminary findings

We have performed two preliminary studies, where we linked survey data regarding volunteers motives to system logs data on these subjects actiual contribution levels. The first study 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 on the results from these studies.

Our study at SETI@home 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). Enhancement motivation (defined as the …) was found to

Comment [SoB1]: Oded – please completre

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 theis study - values motivation (defined in terms of 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. This result, too, stands in contrast to previous findings in the contexzt of crowdsourcing projects. Two other factors at the individual level – affiliation to a team and tenure – were found to be significant. Affiliation to a team exerted significant positive impact on contribution level (the coefficient was 0.16), while tenure exhibited a statistically significant negative impact (the coefficient was -0.27). 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).

Variable	Standardized Coefficients	t	Sig.
(Constant)		48.524	.000
Enjoyment	0.092	1.394	.165
Reputation	0.030	0.534	.594
Values	-0.144	-2.218	.027
Enhancement	0.182	2.779	.006
Team Affiliation	0.160	2.892	.004
Tenure	-0.291	-5.299	.000
Team Affiliation x Tenure	0.113	2.058	.041
Control (no. of computers)	0.272	4.961	.000

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 CWOP and found that the most salient factors for driving participation were (in order of importance): learning new information, values, norms, intrinsic motivations (e.g. enjoyment), and career advancement. Other factors – e.g. identification and reputation – on the other hand proved insignificant.

These preliminary findings demonstrate that not only that motives for SciSourcing contribution differ from the reported motives in crowdsourcing projects, there are also some substantial differences between SciSourcing projects based on granularity (e.g. in terms of the effect values). These findings higfhlight the need to develop a better understanding of the factors that influence contributors in SciSourcing initiatives. As part of our ongoing research program, we plan to continue exploring the behavioral aspects of SciSourcing participation in a variety of settings. We encourage others to join us in researching this important phenomenon.