Dusting for science: motivation and participation of digital citizen science volunteers

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

ABSTRACT
Digital citizen science offers a low-cost way to strengthen the scientific infrastructure, and engage members of the public in science. It is based on two pillars: (1) a technological pillar, which involves developing computer systems to manage large amounts of distributed resources, and (2) a motivational pillar, which involves attracting and retaining volunteers who would contribute their skills, time, and effort to a scientific cause. While the technological dimension has been widely studied, the motivational dimension received little attention to date. To address this gap, we surveyed volunteers at Stardust@home a digital citizen science project, in which volunteers classify online images from NASA's Stardust spacecraft. We found that collective and intrinsic motivations are the most salient motivational factors, whereas reward motives seem to be less relevant. We also found that intrinsic and norm-oriented motives are most strongly associated with participation intentions, which were, in turn, found to be associated with participation effort. Implications for research and practice are discussed.

Author keywords
Citizen science, motivation, participation, Stardust@home, crowdsourcing

ACM Classification Keywords
H5.m. Information interfaces and presentation (e.g., HCI): Miscellaneous.

INTRODUCTION
Many aspects of scientific research, such as observation, classification and analysis, are labor-intensive, time-consuming, and as a result, costly [21, 28]. Citizen science offers a participatory approach for conducting scientific research, following the crowdsourcing model, whereby volunteers participate in scientific research by contributing to data collection and analysis [24]. In times of research budget cuts and increasing cost of doing science [5, 38], citizen science represents a low-cost way to both strengthen the scientific infrastructure and engage members of the public in science. According to the director of the NSF's Information and Intelligent Systems Division, it is "a natural solution to many of the problems that scientists are dealing with that involve massive amounts of data" [38].

In digital citizen science projects [16] participation takes place online. Some examples include: volunteer computing projects, such as SETI@home and Rosetta@home (where volunteers run on their computers programs that search radio telescope signals for signs of extra terrestrial intelligence in SETI@home, and help determine the three-dimensional shapes of proteins in Rosetta@home) or web-based image classification projects, such as Stardust@home (where volunteers classify images from NASA's Stardust spacecraft, searching for interstellar dust particles); The story of Hanny van Arkel, a Dutch schoolteacher and Galaxy Zoo volunteer, who discovered in 2007 an unusual gas cloud that scientists believe could help explain the life cycle of quasars [20], illustrates the potential of the participatory approach to doing science.

iConference 2011, February 8-11, 2011, Seattle, WA, USA

BACKGROUND AND RELATED WORK
Scientists increasingly take advantage of digital citizen science [11, 16, 25]. For example, Galaxy Zoo - where volunteers classify telescope images from the Sloan Digital Sky survey - has reached 250,000 registered members in just three years.

Digital citizen science is founded on two pillars. The first is technological: developing computer systems to manage, allocate, and aggregate, large amounts of distributed resources. The second pillar – which this study investigates - is motivational: attracting and retaining people who would be willing to contribute their skills, time, and effort for a scientific cause. Yet, while the former was widely studied (e.g., [1, 2, 3]), the latter received little attention to date. Why do volunteers contribute their time, effort, and expertise to scientific projects? Understanding the motivational aspect is crucial for the design and management of such projects, especially given evidence that contributors often reduce their involvement after an initial period of experimentation, and that many digital citizen science projects suffer from an alarmingly high attrition rate.

The motivations underlying information and resource sharing in online peer-production venues has attracted substantial scholarly attention in recent years (e.g., [19, 36]), and studies were conducted in a variety of contexts, including open source software development (e.g., Linux), collaborative authoring (e.g., Wikipedia), collaborative tagging (e.g., delicious). These studies demonstrate that while in each of these domains participants are driven by a mix of motives, the motivational factors that are most salient differ between projects. These differences stem from a variety of sources, such the project's ideology and contributors' identification with it, the skills required for participation, or the availability of indirect monetary incentives. For example, a comparison of the motivations of open source software and Wikipedia volunteers revealed that while gaining professional reputation and learning new skills were the most salient motives for open source participation, the desire to help others in the community was the most important motivation for participation in Wikipedia [23].

There are important differences between contributions made to digital citizen science and those made to other types of community-based peer-production projects. These differences stress the need to investigate motivations for participation in the specific context of digital citizen science: First, in digital citizen science there is a clear distinction between the volunteers making the contribution and those benefiting from the aggregate effort (i.e. the scientists who run the project). This asymmetric structure differs from most other community-based projects (e.g. Wikipedia, YouTube), where the distinction is blurred and contributors are also users of the community-generated. Second, it often takes a long time for the output of the scientific project to be made public, in contrast to community-based projects where the contributions are viewable immediately, which may provide instant gratification to contributors.

RESEARCH MODEL

For the empirical study of citizen scientists', we adopted the theoretical framework of voluntary participation in social movements [18, 31]. This framework includes four classes of volunteers' motivations for participation: collective motives (the importance attributed to the project's goals); norm-oriented motives (expectations regarding the reactions of important others, such as friends, family or colleagues); reward motives (benefits such as gaining reputation or making new friends); and collective identification (identification with the group, and following its norms). This conceptualization has been extended to include a fifth factor in studies of participation in open-source software development [17] and Wikipedia editing [30] – a hedonistic or intrinsic motivation, operationalized as the enjoyment associated with participation in the project. Given the broad range of possible 'reward motives' [17], we divided this factor to two specific motives, which were identified in studies of previous online communities: community reputation benefits and social interaction benefits [7, 27].

The model proposed in the present study follows two influential theories – the theory of reasoned action (TRA) [13] one of the most influential theories of human behavior [34], and its application in the technology adoption literature, the technology acceptance (TAM) (e.g., 12, 34, 35]). According to these theories, an intention to perform a certain behavior links the actual behavior to upstream antecedents. We used two different behavioral intentions, the intention to increase participation and the intention to continue participation [4], as the constructs linking motivations to behavior, and in our case, to citizen science participation (see Figure 1).

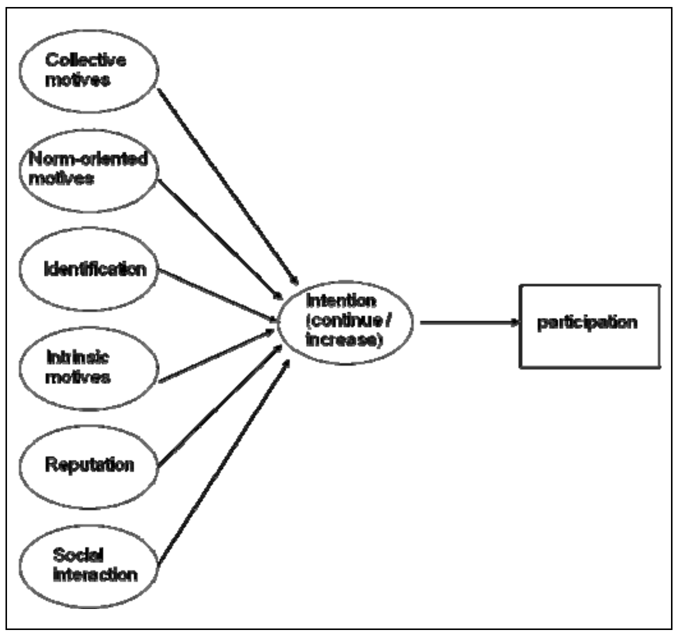

Figure 1. Research model

METHOD

Data Collection

We examined Stardust@home (stardustathome.ssl.berkeley.edu), a digital citizen science project in which volunteers (also known as "Dusters") classify online images from NASA's Stardust spacecraft. Using a virtual microscope developed by the Stardust@home team, dusters classify images using their home computers and search for tracks left by very small interstellar dust particles impacting Stardust’s aerogel tiles (see Figure 2 for a screenshot of the virtual microscope used by volunteers).

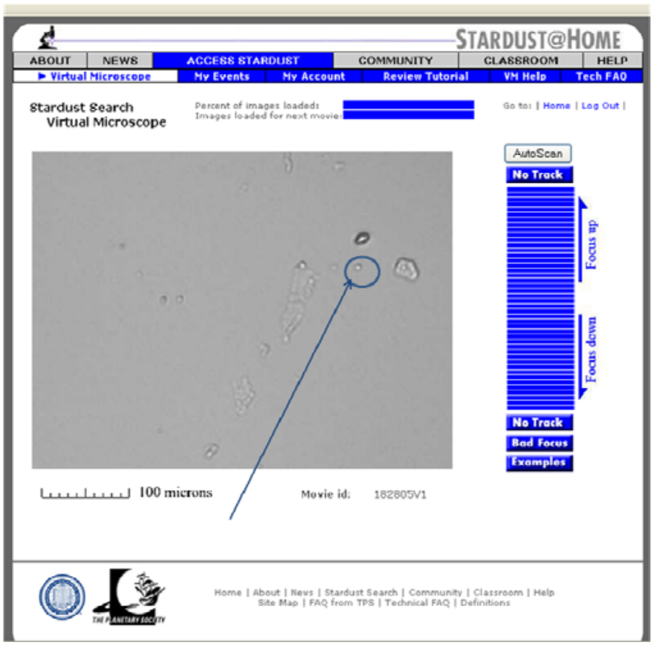

Figure 2. A screenshot of the Stardust@home Virtual Microscope; a volunteer’s categorization of an image as a potential dust particle is circled.

Survey items were adapted from previous studies and adjusted to the citizen science context. The survey was developed based on the Extended Klandermans Model [17, 18], and used a number of sources [7, 17, 27, 30] for specific items based on the model. Participation was measured as the number of weekly hours spent in participation, following the operationalization used in previous studies of online voluntary participation (e.g. [17, 19, 22]).

The survey was administered to volunteers in the project who were active in the 30 days prior to the survey time and respondents were asked to rate the importance of the different motives on a 1-7 Likert scale. 139 Stardust@home volunteers participated in the survey, representing a response rate of 27.1%, which is relatively high compared to similar studies (e.g., [37]).

Structural equation modeling (SEM) was used to analyze the survey results and estimate the relationships between the constructs. Partial Least Squares (PLS) was applied [9] using SmartPLS 2.0 [26] for the measurement validation and structural model testing. PLS is used extensively in information systems research as it offers a number of advantages that are pertinent to the present study: In addition to the verification of a complex model, PLS enables testing of individual hypotheses and provides amount of variance explained for each endogenous variable. Compared to covariance-based SEM and regression, it is better suited to dealing with data nonnormality and small sample size [9]. Similar to other structural equation modeling techniques, it allows measurement validation and model verification to be performed in a single step.

RESULTS

To confirm the reliability of survey items, we conducted a factor analysis. Eight factors emerged, corresponding directly to our framework of six motivational factors and two intention factors, with 89.9% of the total variance explained. All items factor loadings on the intended construct were higher than their cross-loadings, as expected (see Table 1). Further, to confirm convergent and discriminant validity, we calculated the average variance extracted (AVE) for each construct (Fornell). For each construct, AVE exceeded 0.5, and the square root of AVE (RAVE) exceeded the correlation with other constructs - thus displaying convergent and discriminant validity [8, 9]. In addition, all constructs exhibited Cronbach’s alpha values above the generally accepted level of 0.70 [32], indicating measures reliability.

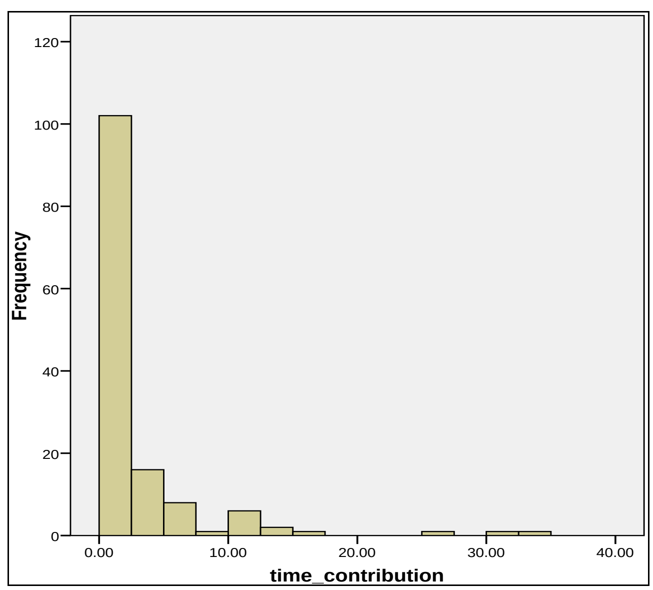

Figure 3. Distribution of Stardust@home volunteer participation levels.

The analysis of the results reveals a diverse set of volunteers (see Figure 3 for a distribution of time spent on participation; the Y axis represents the number of volunteers), with the majority of volunteers spending less than two hours per week, and a minority of dedicated volunteers spending more. This observed power-law participation pattern is highly common in peer-production communities.

The results' descriptive statistics are presented in Table 1. The two most salient motivational factors were collective and intrinsic motives (6.45 and 5.98 out of 7, respectively); identification and norm-oriented motives were found to be of secondary importance, and the reward motives of reputation and social interaction did not seem to play an important role.

Table 1: Item means, standard deviations, and factor loadings

Construct	Item	1	2	3	4	5	6	7	8	Mean	S.D
Continue participation	CONT1	.830	.374	.113	.062	.056	.173	.192	.076	5.94	1.07
	CONT2	.834	.339	.034	.092	.056	.222	.182	.091		
Increase participation	INCR1	.294	.879	.134	.066	.175	.132	.126	.082	5.12	1.42
	INCR2	.327	.862	.141	.132	.107	.156	.129	.032		
Reputation	REP1	.073	.164	.840	.031	.286	.167	.023	.240	3.70	1.62
	REP2	.077	.097	.899	-.014	.261	.130	.098	.035		
Collective motives	COL1	-.005	.090	.006	.924	-.049	-.051	.106	.097	6.45	0.72
	COL2	.133	.059	.007	.923	.119	.023	.021	-.006		
Social interaction	SOC1	.048	.125	.193	.047	.886	.068	.023	.182	3.30	1.30
	SOC2	.047	.121	.360	.032	.820	.132	.097	.061		
Norm-oriented motives	NORM1	.153	.221	.194	-.061	.044	.850	.043	-.032	4.52	1.12
	NORM2	.179	.042	.083	.026	.149	.849	.161	.180		
Intrinsic motives	INT1	.171	.177	.064	.097	.088	.127	.911	.086	5.98	0.83
	INT2	.575	.075	.095	.077	.027	.122	.660	.149		
Identification	IDNTP	.160	.090	.229	.102	.251	.134	.168	.880	4.42	1.37

The PLS results of testing the two models (a separate model testing for each of the behavioral intention factors) are presented in Table 2. We used the log-transformed participation data in the analysis because of the highly skewed distribution of the participation variable (see Figure 3), as is common in studies of online participation [6]. In both models, intention was significantly related to participation, and intrinsic motives and norm-oriented motives were most strongly related to participation intentions (the path coefficients were 0.294 and 0.185 for Model 1, and 0.509 and 0.243 for Model 2).

Table 2: PLS analysis results

	Variable	Model 1	Model 2
Participation motivations	Reputation	0.148†	0.036
	Collective motives	0.129*	0.071
	Social interaction	0.134	-0.015
	Intrinsic motives	0.294***	0.509***
	Norm-oriented motives	0.185*	0.243**
	Identification	-0.037	0.033
Participation intentions	Continue participation	--	0.483***
	Increase participation	0.282***	--
R^2	R^2 (intention)	0.313	0.454
	R^2 (contribution)	0.080	0.233

†p<0.1 *p<0.05 **p<0.01 ***p<0.01

DISCUSSION AND CONCLUSION

Digital citizen science is emerging as a powerful way to conduct scientific research by drawing on large numbers of geographically distributed volunteers. In the present study we proposed and tested a framework of the antecedents of contribution in digital citizen science projects, by building on the body of literature on online volunteer participation, and adapting it to the citizen science context. We tested our proposed framework and used structural equation modelling to examine the influences of antecedents identified on participation levels. The findings reveal the factors that are most important for motivating volunteers to participate in digital citizen science activities.

The findings show similarity to recent studies of the motivations of Wikipedia contributors. [30] identified intrinsic and collective motives as the most salient motivations. [22] found that the two most salient motivations were enjoyment and Wikipedia-related ideology (i.e. “information should be free”). While the Wikipedia study employed a different survey instrument, both motivations identified represent similar participants’ concerns: the top motivation in both studies is enjoyment, and the second involves identification with the goals of the projects. This similarity of the results suggests that the salience of the top motives is a relatively general phenomenon, not dependent on a particular participation context. Further research can help determine whether this holds in other peer-production venues as well.

The findings have important implications for the design and management of digital citizen science projects; we recommend that designers and leaders of such projects focus their recruiting and retention efforts on motivational factors that are more salient and have a positive relation with intention and participation. The high levels of collective motives suggest that citizen science projects should strive to increase volunteers’ commitment to the project and its goals. This could possibly be done by communicating the project’s mission and achievement to the volunteer-base (e.g. through social media). The salience of intrinsic motives stresses the need to develop game-like contribution channels, such as the one used in Foldit [10], a multiplayer online game in which citizen scientists compete by folding proteins into a chemically stable configuration. The moderate levels observed for identification and norm-oriented motives suggest that - although of secondary importance – administrators should not neglect the necessity to establish a community of volunteers who share beliefs, interact regularly, possibly using social media outlets, and work collectively towards a common goal.

Another key insight from this study is the need to create dynamic contribution environments that allow volunteers to start contributing at lower-level granularity tasks, and gradually progress to more demanding tasks and responsibilities. Many community-based projects, such as open source software development and Wikipedia, have long realized this, and they allow interested contributors to progress in the ladder of responsibilities. This mechanism is currently absent in digital citizen science projects, where volunteers’ tasks are usually restricted in their scope, and the governance and decision making is left in the hands of the scientists managing the projects. Adopting a more symmetric governance structure, closer to the one in community-based projects such as Wikipedia, represents a major paradigm shift, even for those scientists who appreciate the potential benefits of citizen science. However, as digital citizen science develops, and competition for volunteers’ resource increase, such a trend toward greater empowerment of volunteers may be inevitable.

The present study has a number of limitations that can be addressed in future research. The study was conducted in a specific citizen science project, Stardust@home. Studies of other citizen science projects, in different fields or with different goals, could help verify the generalizability of the findings. Further, we applied a cross-sectional research design, which allows establishing correlations between constructs, and thus arguments regarding causal relationships should be taken with caution.

In conclusion, the present study advances the understanding of the motivational pillar of digital citizen science. Still a number of questions remain open and warrant future research. Some of the future research directions we identify are (a) examining the mechanisms by which participants increase or decrease their participation levels over time, (b) identifying additional factors that may have an effect on contribution, such as personality traits, other motivations not studied here, and (c) exploring how changes in the design of digital citizen science systems could possibly help increase volunteer participation.

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* 45, 11, 56-61.
2. Anderson, D. (2006). Volunteer computing: Planting the flag. *PCGrid 2007 Workshop*, Long Beach, CA, USA, March 2007.
3. Anderson D. and Fedak. G. (2006). The computational and storage potential of volunteer computing. *CCGRID'06*, 2006.
4. Bhattacherjee, A. (2001). Understanding information systems continuance: an expectation-confirmation model. *MIS Quarterly* 25, 3, 351-370.
5. Bonetta L. Tough Challenges for the Next NIH Director (2008) *Cell* 135 (4), pp. 583-585.
6. Burke, M., Marlow, C., and Lento, T. (2009). Feed me: motivating newcomer contribution in social network sites. *Proc. CHI '09*. 945-954.
7. Butler, B., Sproull, L., Kiesler, S., and Kraut, R. (2002). Community effort in online groups: Who does the work and why? In S. Weisband and L. Atwater (Eds.) *Leadership at a distance*. Mahwah, NJ: Lawrence Erlbaum.
8. Chin, W. (1998). The partial least squares approach for structural equation modeling. In: Marcoulides G, ed. *Modern Methods for Business Research*. Lawrence Erlbaum Associates.
9. Chin, W., Marcolin B, Newsted P. (2003). A partial least squares latent variable modeling approach for measuring interaction effects. *Information Systems Research* 14(2): 189-217.
10. Cooper, S., Khatib, F., Treuille, A., et al. (2010). Predicting protein structures with a multiplayer online game. *Nature* 466, 756-760.
11. Das R, Baker D. (2007) Automated de novo prediction of native-like RNA tertiary structures. *Proceedings of the National Academies of Science*; 104:14664–69.
12. Davis, F. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly*, 13(3): 319-339.
13. Fishbein, M., Ajzen, I. Belief, *Attitude, Intention and Behavior: An Introduction to Theory and Research,* Addison-Wesley, Reading, MA, 1975.
14. Fornell, C. D. Larker. Evaluating structural equation models with unobservable variables and measurement error. Journal of Marketing Research, 18, 39–50 (1981).
16. Hand, E. (2010) Citizen science: People power. *Nature* 466, 685-687.
17. Hertel, G., Niedner, S., & Herrmann, S. (2003). Motivation of software developers in Open Source projects: an Internet-based survey of contributors to the Linux kernel. *Research Policy*, 32, 1159-1177.
18. Klandermans, B. (1997). *The social psychology of protest.* Oxford, UK: Blackwell.
19. Lakhani, K., & Wolf, R. (2005). Why hackers do what they do. In J. Feller, B. Fitzgerald, S. Hissam, & K. Lakhani (Eds.), *Perspectives in Free and Open-Source Software.* MIT Press.
20. Lintott, C., Schawinski, K., Keel W., et al. (2009). Galaxy Zoo: 'Hanny's Voorwerp', a quasar light echo? *Monthly Notices of the Royal Astronomical Society*.
21. Maxwell, S., Meliker, J., et al. (2009). Use of land surface remotely sensed satellite and airborne data for environmental exposure assessment in cancer research. *Journal of Exposure Science and Environmental Epidemiology*.
22. Nov, O. (2007). What motivates Wikipedians. *Communications of the ACM*, 50 (11) 60-64.
23. Oreg, S. and O. Nov (2008). Exploring motivations forcontributing to open source initiatives: The roles of contribution context and personal values. *Computers in Human Behavior* (24) 5, 2055-2073.
24. Raddick, M. J., Bracey, G., Carney, K., et al, (2009). Citizen Science: Status and Research Directions for the Coming Decade. *ASTRO2010 Decadal Survey Position Paper.*
25. Raman, S., Lange, O. F., Rossi, P., et al. (2010) NMR structure determination for larger proteins using backbone-only data. *Science*, 327, 1014-1018.
26. Ringle, C.M, Wende, S, SmartPLS 2.0. www.smartpls.de (accessed 20 July 2010).
27. Roberts, J., Hann, I. H., & Slaughter, S. (2006). Understanding the motivations, participation, and performance of open source software developers: A longitudinal study of the apache projects. *Management Science*, 52(7) 984–999.
28. Rose, A. (2009). Temporal and Individual Variation in Offspring Provisioning by Tree Swallows: A New

Method of Automated Nest Attendance Monitoring. *PLoS ONE* 4(1): e4111.

30. Schroer, J., & Hertel, G. (2009). Voluntary engagement in an open Web based encyclopedia: Wikipedians, and why they do it. *Media Psychology*, 12, 1–25.

31. Simon, B., Loewy, M., Sturmer, S., et al. (1998). Collective identification and social movement participation. *Journal of Personality and Social Psychology*, 74, 646-658.

32. Straub, D. M.-C. Boudreau, Validation guidelines for IS positivist research. *Communications of the AIS*, 13, 380–427 (2004).

33. Venkatesh, V., Morris, M, User Acceptance of Information Technology - Toward a Unified View. *MIS Quart*, 2003; 27,3, 425-478.

34. Venkatesh, V. (2000). Determinants of perceived ease of use: Integrating control, intrinsic motivation, and emotion into the technology acceptance model. *Information Systems Research*, 11(4), 342–366.

35. Venkatesh V, Davis F. (2000). A theoretical extension of the technology acceptance model: Four longitudinal field studies. *Management Science*, 46, 186-204.

36. von Hippel, E., & von Krogh, G. (2003). Open source software and the “Private-Collective” innovation model: Issues for organization science. *Organization Science*, 14(2), 209–223.

37. Wu, C. and Gerlach, Y. (2007). An empirical analysis of open source software developers’ motivations and continuance intentions. *Information & Management*, 44, 253–262 (2007).

38. Young, J. Crowd science reaches new heights. *The chronicle of higher education*, 28 May 2010.