Distributed Scientific Collaboration: Research Opportunities in Citizen Science

Andrea Wiggins
Syracuse University
Syracuse, NY 13244 USA
awiggins@syr.edu

Kevin Crowston
Syracuse University
Syracuse, NY 13244 USA
crowston@syr.edu

ABSTRACT
This paper introduces a conceptual framework for research on citizen science, a form of collaboration involving scientists and volunteers in scientific research. Designing CSCW systems to support this type of scientific collaboration requires understanding the effects of organizational and work design on the scientific outcomes of citizen science projects. Initial directions for future research are identified, with the goal of developing a foundation for research on and development of cyberinfrastructure and collaborative technologies for supporting citizen science

Author Keywords
design, citizen science, distributed work, scientific collaboration, public participation

ACM Classification Keywords
K.4.3 [Organizational Impacts]—Computer-supported collaborative work, H.5.3 [Group and Organization Interfaces]—Organizational design, Computer-supported cooperative work.

INTRODUCTION
Research which relies upon data about the natural world, and indeed the universe, is often hindered or rendered impossible by the high cost of data collection and analysis. In these areas, which include a wide array of environmental sciences, astronomy, and even genetics, scientific progress cannot keep pace with the demand for knowledge to address increasingly urgent global-scale problems. The real-world problems that fall into this category typically depend on either massive data sets that cannot be automatically generated, data collected over longer periods of time or wider geographic areas, or large-scale analyses that require human perceptual competencies; they range from climate change to the search for cures for cancer. To address these issues, as well as many other questions spanning a variety of disciplines, scientists are now employing citizen science as a solution to enable scientific research that is not feasible by any other means.

Citizen science projects involve the public with scientists in collaborative research. Many are virtual organizations, with geographically dispersed resources and members who work toward common goals through cyberinfrastructure. Key characteristics of this phenomenon and context are familiar in CSCW research, and most citizen science projects resemble the Community Data Systems and Open Community Contribution Systems models of scientific collaboratories [3]. Related research underscores the importance of understanding how organizational, task, and technology design requirements interact to affect participation and the scientific value of the work [13, 18]. However, suitable theoretical models are still needed to make sense of such complex phenomena and to provide a framework for further research.

This paper presents a design-oriented conceptual framework for organizing investigation into the interactions of organizational, task, and technology design in citizen science. The goal of the work is to develop a theoretical basis for research and development of cyberinfrastructure and collaborative technologies to support citizen science projects.

MOTIVATION
Citizen science projects conducted via web technologies are a form of massive virtual collaboration, based on voluntary contributions by diverse participants. Citizen science projects have some distinctive differences from other open production contexts that affect the processes and outcomes of participation. Designing CSCW systems to support this type of scientific collaboration requires understanding the effects of organizational and work design on the scientific outcomes of citizen science projects.

Citizen Science
The practice of citizen science is related to long-standing programs of volunteer monitoring for natural resource management, and projects are increasingly focused on balancing scientific and informal education goals. Public participation in scientific research can take a variety of forms; the dominant form of citizen science projects, found in the biological and environmental sciences, has focused primarily on monitoring ecosystems and wildlife populations (e.g., butterflies, birds). In these Community Data Systems, volunteers form a human sensor network for distributed data collection [5, 2]. By contrast, in projects organized by astronomers, such as NASA's Clickworkers [11], volunteers provide data analysis service by applying basic human perceptual capacities to computationally difficult image recognition tasks, in keeping with an Open Community Contribution System model. In many citizen science projects, more meaningful inclusion

CSCW, February 6–10, 2010, Savannah, GA, USA.

of volunteers in the scientific process arises through interaction, direct or indirect, with professional researchers; for example, the GalaxyZoo project's research into "green peas" galaxies came about at the insistence of volunteers who campaigned in the project's forums to "Give Peas a Chance" [1].

This type of organizational and work design is not new to science, but ubiquitous computing now makes broad public participation in scientific work a realistic research strategy for an increasing variety of projects. The evidence is clear that under the right circumstances, citizen science can work on a massive scale and is capable of producing high quality data as well as unexpected insights and innovations [2, 19], particularly when coupled with traditional scientific studies.

Context

Citizen science projects are similar in some respects to massive virtual collaborations, but have scientific goals that pose particular constraints on task design. For example, assuring the reliability of data collection is critical to the value of a scientific project, but not a matter that can necessarily be left to the "wisdom of crowds". A number of projects, such as GalaxyZoo and the North American Bird Phenology Program, evaluate the contribution quality through multiple independent ratings; however, these and other similar mechanisms for validation are not yet in widespread use in the practitioner community. Including volunteers in scientific research projects also results in very different distributed organizational structures than those of scientific collaboratories, raising new challenges for scientists to manage [13]. For example, the design of scientific collaboratories may tacitly assume that participants have comparable and high levels of skill and will contribute relatively equally. This is rarely the case for citizen science volunteers, who have widely varying levels of skill or knowledge, and contribute at levels differing by orders of magnitude. Combined, these factors raise unique concerns for designing CSCW systems to support citizen science.

CONCEPTUAL FRAMEWORK

For our conceptual framework, shown in Figure 1, we chose to analyze citizen science projects as work teams, comprised of individuals acting as a social entity embedded in a larger organization and performing interdependent tasks [7]. A team differs from a community of practice because members have a shared output, instead of applying common practices to individual, independent tasks. A citizen science project has a shared goal and social identity, and interdependencies usually take the form of volunteers depending upon staff for training, frequently in collocated contexts for localized projects, and researchers depending upon volunteer contributions to data and analysis. Adopting this perspective allows us to draw from the extensive research on small groups, providing a theoretical starting point for further development. At the same time, the wide variations in the structure of social interaction and interdependency in many citizen science projects is not adequately encompassed by small group models. In future revisions of this initial model, we expect to better address these discrepancies through empirical research, currently in an early stage of development. In

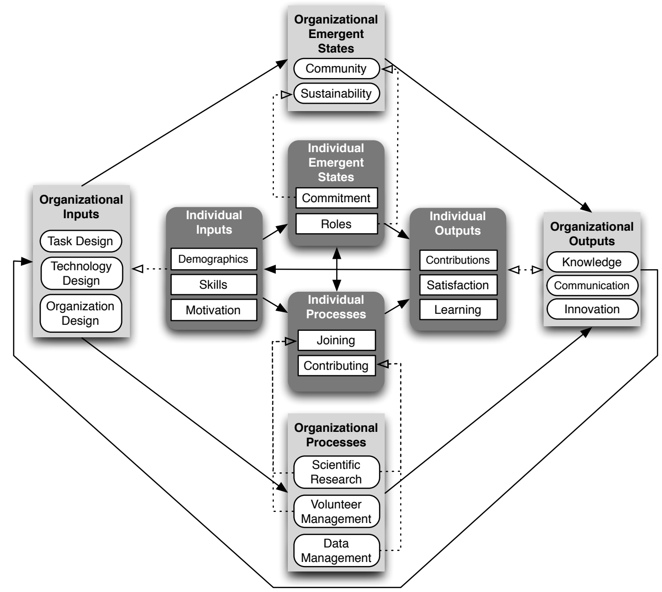

Figure 1. A conceptual model of citizen science virtual organizations.

this initial framework, synthesizing elements from organizational design, sociology and management with small group theory helps describe the antecedents of knowledge production in citizen science.

We organize our conceptual framework as an input-mediator-output-input (IMOI) model [10]. Inputs are the starting conditions, including member characteristics and project/task characteristics [9]. Mediators represent factors that mediate the influence of inputs on outputs, divided into two categories: processes and emergent states. Processes represent dynamic interactions among team members, leading to the outputs, which are task and non-task outcomes of team activity. Emergent states are constructs that characterize dynamic team properties, which vary by context.

Inputs

Inputs are the starting conditions of a project; at the individual level, both staff and volunteers come to the project with diverse demographics, levels of skill, and motivations for their individual contributions to the project. While demographics and skills will vary, prior research has consistently identified a set of common motivators that may have differential effects on individual experiences and performance [12, 16, 4].

At the organizational level, we focus on the effects of organizational, task and cyberinfrastructure technology design. Organizational design is a key point of differentiation between citizen science projects and other scientific collaboratories. These configurations vary widely, ranging from a single PI with a research assistant to an inter-organizational network of federal agencies, academic researchers and nonprofit organizations, each with different goals and resources to contribute. However, the overall structure is likely to mirror the core/periphery structure that describes many distributed projects with volunteer contributors: a core of highly involved project leaders, surrounded by a larger group of active volunteers and a still larger group of occasional contributors [6]. One important difference in citizen science projects is that there are often formal status differences that separate these groups: most core participants likely have graduate training and formal roles, while other participants are lay volunteers.

The second organizational input, "task design", encompasses several related concepts. Task design in this context includes the research design for the project, the job design for individual participants, and the task design for citizen science protocols, which must reflect careful consideration of job design and task design [5, 19]. Organizational design theories link individual-level inputs and outputs (motivation and performance) to the task design, as do theories of volunteerism [16].

Finally, technology design and use is of particular interest given the potential of cyberinfrastructure to support citizen science, especially for data management. In many cases, traditional place-based volunteer monitoring practices are being adapted for virtual participation with limited consideration of the change in context from face-to-face to virtual, which often involves moving from analog to digital technologies. Study designs and volunteer protocols for virtual citizen science projects may require applying a different set of design criteria to ensure scientifically valid results and sustainable participation. These issues are even more salient for projects involving large numbers of participants, as the scale of the activity fundamentally alters the types of relationships that scientists are able to forge with volunteers. Understanding the range of interactions between such diverse end users and technologies that support the scientific research is important to creating usable, robust CSCW systems for collecting useful independent contributions by distributed volunteers [14].

Processes

In the IMOI model, the inputs are conceptualized as influencing the effectiveness of projects through two sets of moderators, processes and emergent states. Processes are the dynamic interactions among group members leading to outputs. In this context, volunteer involvement varies widely, and may include tasks at almost any stage of the scientific research process. Understanding these work practices is the first key to designing CSCW systems to support knowledge production and innovation in citizen science.

At the organizational level, the processes include those of scientific research itself. The nature of the research will have an important influence on the kinds of data and analysis required, and the mapping of tasks to actors with different roles. Similarly, data management processes have a significant impact on project outcomes, particularly for inter-organizational projects that must ensure interoperability and reliability of data created by volunteers. Finally, a unique aspect of this context is the applicability of volunteer management processes often associated with nonprofit management, e.g., recruitment, training, supervision, recognition, and retention of volunteers [16].

Emergent States

Emergent states are dynamic properties of the group that vary as a function of inputs and processes. Potentially relevant emergent states that include task-related factors that describe the state of the group in terms of its progress on the scientific task, as well as social factors that describe social states of the group that enable that work [13]. Research on other kinds of virtual organizations has identified the importance of interpersonal relationships that affect the sense of group community, and thus long-term sustainability [15].

At the individual level, the evolution of volunteers through different roles in the group, from initial volunteer through sustained contributor, and potentially to more central roles, is relevant to organizational design. A related concern is volunteers' level of commitment to the project and how it influences task performance [4]. Understanding how these factors affect the social and technological barriers to and enablers of participation is important for effective technology and work design.

At the individual level, the input elements of organizational, task and technology design affect motivation and participation of distributed volunteers [12, 17]. At the project level, they may transform the means of production of scientific knowledge, shaping the demand for supporting cyberinfrastructure and potentially transforming organizational design.

Outputs

Finally, outputs represent task and non-task consequences of group activity, signaling effectiveness. At the individual level, the task outputs are contributions, most often raw or processed data. In addition to the individual-level outputs, outputs at the project level include scientific knowledge created from the data. Innovative findings, processes and tools can also emerge from involving the public in scientific research.

Hackman's model of group effectiveness [8] also includes non-task outputs. Satisfaction of individual participants' needs, such as individual learning and personal satisfaction, are measures of effectiveness closely related to the educational mission of many citizen science projects. Finally, Hackman also includes the group's continued ability to work together, speaking to project sustainability; a project is not effective if it achieves a research goal but drives away participants in the process.

An important feature of the IMOI model is that outputs become future inputs to the dynamic processes. Positive personal outcomes can motivate future participation, and individual learning can increase ability to contribute. Positive project outputs may lead to increased interest from researchers and volunteers alike. At the societal level, project success may affect public participation in and perception of science, create informal learning opportunities, and enable knowledge production at an unprecedented pace and scale [19, 5].

CONCLUSION AND FUTURE WORK

In summary, synthesizing elements of prior research on small groups with contextually relevant theory provides a theoretical foundation for future research on and development of technologies to support massive virtual collaboration in scientific research. Differences between prior work and the context of citizen science suggests the need to validate the applicability of existing theory and search for possible extensions, raising a number of questions for future research.

Foremost among these questions is to what degree these projects truly resemble small groups, communities of practice, and crowdsourcing; this issue is particularly important for further revision of the conceptual framework presented here. In addition, the increasing number of virtual citizen science projects raises concern over the translation of analog to digital participation; often the extent of the adaptation is creation of an online data submission form that directly mimics paper data forms and provision of presentation slides for training purposes, but it is questionable whether these materials and related research protocols are crafted to overcome the lack of direct interaction. Further investigation is also needed to better understand the nature of the volunteer experience, which influences related factors that directly impact outcomes, such as who chooses to volunteer, what they are willing to do, how long they remain involved, and the quality of their contributions.

The paper contributes a multi-level design-oriented conceptual model of participation in citizen science virtual organizations, and identifies initial directions for inquiry into the phenomenon. Finally, it provides a basis for future research on and development of cyberinfrastructure and collaborative technologies for supporting citizen science.

ACKNOWLEDGEMENTS

This work has been partially supported by NSF Grant 0943049, under the American Recovery and Reinvestment Act of 2009.

REFERENCES

1. GalaxyZoo Blog. Give peas a chance. http://www.galaxyzooblog.org/2008/08/19/give-peas-a-chance/, November 2009.
2. R. Bonney and M. LaBranche. Citizen science: Involving the public in research. *ASTC Dimensions*, page 13, May/June 2004.
3. N. Bos, A. Zimmerman, J. Olson, J. Yew, J. Yerkie, E. Dahl, and G. Olson. From shared databases to communities of practice: A taxonomy of collaboratories. *Journal of Computer-Mediated Communication*, 12(2):652–672, 2007.
4. R.A. Cnaan and T.A. Cascio. Performance and commitment: Issues in management of volunteers in human service organizations. *J. Soc. Serv. Res.*, 24:1–38, 1999.
5. Jeffrey P. Cohn. Citizen science: Can volunteers do real research? *BioScience*, 58(3):192–107, March 2008 2008.
6. K. Crowston and J. Howison. The social structure of free and open source software development. *First Monday*, 10(2-7), 2005.
7. R.A. Guzzo and M.W. Dickson. Teams in organizations: Recent research on performance and effectiveness. *Ann. Rev. Psy.*, 47(1):307–338, 1996.
8. J.R. Hackman. The design of work teams. *Handbook of organizational behavior*, 315:342, 1987.
9. J.R. Hackman and C.G. Morris. *Group Tasks, Group Interaction Process, and Group Performance Effectiveness: A Review and Proposed Integration.*, pages 45–99. Group Processes. Academic Press, 1974.
10. D.R. Ilgen, J.R. Hollenbeck, M. Johnson, and D. Jundt. Teams in organizations: From IPO models to IMOI models. *Ann. Rev. Psy.*, 56:517–543, 2005.
11. B. Kanefsky, N.G. Barlow, and V.C. Gulick. Can Distributed Volunteers Accomplish Massive Data Analysis Tasks? *Lunar and Planetary Science*, 1, 2001.
12. A. Lawrence. 'no personal motive?' volunteers, biodiversity, and the false dichotomies of participation. *Ethics, Place & Environment*, 9(3):279–298, 2006.
13. C.P. Lee, P. Dourish, and G. Mark. The human infrastructure of cyberinfrastructure. In *Proc. CSCW 2006*, pages 483–492, 2006.
14. K. Luther, S. Counts, K.B. Stecher, A. Hoff, and P. Johns. Pathfinder: an online collaboration environment for citizen scientists. In *Proc. CHI 2009*, pages 239–248, 2009.
15. M.L. Markus, B. Manville, and C.E. Agres. What makes a virtual organization work? *Sloan Management Review*, 42(1):13–26, 2000.
16. J.L. Pearce. *Volunteers: The organizational behavior of unpaid workers*. Routledge, 1993.
17. L. Sproull and S. Kiesler. Public volunteer work on the Internet. *Transforming Enterprise: The Economic and Social Implications of Information Technology*, page 361, 2005.
18. S.L. Star and K. Ruhleder. Steps towards an ecology of infrastructure: complex problems in design and access for large-scale collaborative systems. In *Proc. CSCW 1994*, pages 253–264, 1994.
19. D.J. Trumbull, R. Bonney, D. Bascom, and A. Cabral. Thinking scientifically during participation in a citizen-science project. *Science Education*, 84(2):265–275, 2000.