9-14-2010

Developing a Conceptual Model of Virtual Organizations for Citizen Science

Andrea Wiggins
Syracuse University, wiggins@unomaha.edu

Kevin Crowston
Syracuse University

Developing a Conceptual Model of Virtual Organizations for Citizen Science

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

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

Abstract

This paper develops an organization design-oriented conceptual model of scientific knowledge production through citizen science virtual organizations. Citizen science is a form of organization design for collaborative scientific research involving scientists and volunteers, for which Internet-based modes of participation enable massive virtual collaboration by thousands of members of the public. The conceptual model provides an example of a theory development process and discusses its application to an exploratory study. The paper contributes a multi-level process model for organizing investigation into the impact of design on this form of scientific knowledge production.

Keywords: conceptual models, virtual organizations, citizen science, cyberinfrastructure, massive virtual collaborations, theory development, scientific knowledge production

Biographical Notes:

Andrea Wiggins is a Ph.D. Candidate in Information Science and Technology at the School of Information Studies at Syracuse University. Her research interests include social computing and virtual collaboration. Her dissertation is a comparative case study of virtual organizing and public participation in scientific knowledge production. She received her M.S. in Information (2005) from the University of Michigan School of Information.

Kevin Crowston is a Professor in the School of Information Studies at Syracuse University. His research examines new ways of organizing made possible by the extensive use of information and communications technology. Specific research

topics include the development practices of Free/Libre Open Source Software teams and work practices and technology support for citizen science research projects, both with NSF support. He received his Ph.D. (1991) in Information Technologies from the Sloan School of Management, Massachusetts Institute of Technology (MIT).

1 Introduction

Virtual organization (VO) designs have been discussed in the literature for some time. One promising area of application for VOs is to scientific research, using information and communications technologies (ICT) to enable collaboration among scientists (Hey & Trefethen 2005, Freeman, et al. 2005); however, the organization of such VOs is still a new area of research. Designing effective VOs to support scientific collaboration requires the application of social science and computer science research and practice to the study and implementation of new organization designs, including the integrated structuring, modeling, development and deployment of systems and people.

Prior research on VOs for scientific work has focused primarily on distributed collaboration among scientists and related professionals, leading to a rich stream of research on scientific collaboratories (e.g., Finholt 2002, Chin & Lansing 2004). However, the widespread deployment of ubiquitous computing technologies has enabled new options for distributed collaboration. A variety of phenomena that can be loosely described as massive virtual collaboration (e.g., social networking, open source software development, Wikipedia (Crowston & Scozzi 2008, Nov 2007, Forte & Bruckman 2008)) prompt us to consider the potential of VOs for supporting large-scale, distributed and heterogeneous participation in scientific research.

We are interested in particular in the phenomenon of citizen science (Clark & Illman 2001, Cohn 2008), that is, research projects involving "partnerships between volunteers and scientists that answer real-world questions" (Bonney & Shirk 2007). These project-based partnerships are a form of VO, fitting the U.S. National Science Foundation's definition of "a group of individuals whose members and resources may be dispersed geographically, but who function as a coherent unit through the use of cyberinfrastructure"[1] . There are few studies of this form of VO, however, so design principles and potential benefits for science are still being established.

Conceptualizing processes of participation is a first step to developing a theoretical basis for designing virtual organizations for citizen science. This paper draws from social science theory to develop an interdisciplinary design-oriented conceptual model of massive virtual collaboration for scientific knowledge production. Although organizational emergence does not always

[1]from the Virtual Organizations as Sociotechnical Systems (VOSS) Program Solicitation, NSF 09-540

unfold in the top-down fashion that this design-oriented perspective suggests, the model is suited to the goals of providing a foundation for future research and practical decision-making. The model contributes a multi-level model for organizing investigation into organization, task and technology design in virtual citizen science projects.

2 Motivation

Citizen science projects conducted via web technologies can yield massive virtual collaborations based on voluntary contributions by diverse participants. The increasing scale of citizen science projects, some of which involve hundreds of thousands of members of the public in distributed data collection and analysis, suggests a need for additional research. In particular, designing organizations to support this form of scientific knowledge production requires understanding the effects of organization and task design on the scientific outcomes of citizen science projects.

Citizen science is related to long-standing programs employing volunteer monitoring for natural resource management (Cooper, et al. 2007, Firehock & West 1995) and is often employed as a form of education and outreach to promote public understanding of science (Brossard, et al. 2005, Krasny & Bonney 2005). However, modern citizen science projects are increasingly focused on benefits to the scientific research as well (Bonney & LaBranche 2004, Baretto, et al. 2003). The evidence is clear that in 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 (Trumbull, et al. 2000, Fore, et al. 2001).

Public contributions to scientific research can take a variety of forms, with participation ranging from nearly passive to deep engagement in the full process of scientific inquiry. Diverse volunteer populations can contribute to scientific research through a variety of activities, from primary school students engaging in structured classroom projects, to families volunteering together in “bioblast” one-day organism census events, to geographically-distributed individuals monitoring wildlife populations over time.

In the ecological sciences, citizen science projects have focused primarily on observation of ecosystems and wildlife populations (e.g., monarch butterflies, birds, reef fishes), where volunteers form a human sensor network for data collection. For example, in the *Great Sunflower Project*[2] , volunteers grow sunflowers, then report on the numbers of bees that visit the plants in order to gauge the level of pollination activity.

In contrast to these data collection projects, in astronomy research, volunteers apply superior human perceptual capacities to computationally difficult image recognition tasks, providing an important service in data analysis.

[2]http://www.greatsunflower.org/

Galaxy Zoo[3] is such a project, organized by an international team of professional astronomers to classify images of galaxies from the Sloan Digital Sky Survey (Cho & Clery 2009). Volunteers contribute to data reduction through classification tasks performed through a web portal that presents images of galaxies and asks volunteers to make judgements about specific characteristics of the shapes of galaxies. In its first instantiation, *Galaxy Zoo* volunteers classified 750,000 galaxies in record time and the data have been re-incorporated into virtual astronomical observatory tools used by both professional researchers and the public. In its second and third versions, *Galaxy Zoo 2* and *Galaxy Zoo: Hubble*, the project has elicited progressively more complex classification judgments from volunteers, implemented based on the high quality of the results from the simple initial classification. At the start of its third year, *Galaxy Zoo* had classified over 56 million galaxies and counts a growing contributor base of over a quarter million volunteers around the world.

Beyond simply providing image processing services for science, *Galaxy Zoo* participants have made new discoveries, such as Hanny's Voorwerp, an astronomical object of unknown nature (*voorwerp* means "object" in Dutch), which was discovered in 2007 by Hanny van Arkel, a Dutch elementary school teacher. Hanny's Voorwerp is remarkable for its unusual form and color and for emitting more energy than any object previously observed in the universe. The *Galaxy Zoo* researchers will use the Hubble Telescope to examine this new astronomical body. Hanny's Voorwerp demonstrates how profoundly volunteer contributions to scientific research can influence the course of scientific knowledge production.

In addition to innovation, *Galaxy Zoo* volunteers deliver quality; their collective reliability is as good or better than that of professional astronomers. The project's leaders ensure quality by having each image evaluated by multiple volunteers (up to 250 ratings per image), with algorithmic indentification of low-consensus items for professional review. Even without these quality assurance strategies, researchers have found that elementary school children can provide scientifically valid data for species identification, with seventh-graders reporting counts of crab species at 95% accuracy and third-graders correctly identifying animals 80% of the time, an acceptable reliability rate for most ecological studies (Delaney, et al. 2007).

Organizational designs that involve the public are not new to science (e.g., the Audubon Christmas bird count started in 1900), but we are now reaching the point where ubiquitous computing makes broad participation by the public in scientific work a realistic research strategy for an increased variety of scientific research problems. The potential benefits of citizen science are beginning to be realized more widely, particularly when coupled with traditional scientific studies; combined with the growth of technologies that can enable broader participation, this has lead to a swiftly increasing number of projects (Silvertown

[3] http://www.galaxyzoo.org/

2009). Designing such projects for sustainable and scientifically valid outcomes requires a better conceptual understanding of this organizational design, however, which is the goal of this paper.

3 Theory Development Process

There are many strategies for theory development. In the interdisciplinary research that characterizes many phenomenon-oriented studies, contextualizing an existing theoretical framework can provide a strong initial model for study of a new phenomenon or research context. Developing an initial conceptualization requires working out how to transfer what we know from other settings to a new context, given their similarities and differences. For contextualization of a model to provide satisfying results, the initial model must bear some resemblance to an empirically and theoretically informed understanding of the phenomenon.

To develop our model, we first considered a variety of prior settings. At the most basic level, we chose to analyse citizen science projects as a kind of small group, specifically, a work team. We selected the literature on small groups as a starting point because there are a number of functional similarities in the structure of participation, making it reasonable to employ a model drawn from small groups theory as a starting point. For the current study, our conceptual model draws on work in the small group literature, (e.g., Hackman & Morris 1978, Marks, et al. 2001, McGrath & Hollingshead 1994). Although citizen science efforts are typically organized as projects, which are a distinct unit of analysis, their governance and structure are frequently similar to those of an organization, making our conceptual model relevant to organizational design (and vice versa).

We also examined a variety of more specific models developed to describe particular empirical settings. Citizen science projects are similar in some respects to massive virtual collaborations and peer production phenomena such as FLOSS or Wikipedia. These technology-mediated massive virtual collaborations are frequently referred to as forms of crowdsourcing, an ill-defined but now common term which refers to a set of distributed production models that make an open call for contributions from a large, undefined network of people (Howe 2006). Initially introduced as a novel alternative business model, more recent popular use of the term has applied it to any form of collective intelligence that draws on large numbers of participants through the Internet.

While most citizen science does not fully meet the definition of peer production (Benkler 2002), in which the prototypical model is non-hierarchical and self-organizing, they do share other features of this work design. At a conceptual level, citizen science VOs and other lightweight peer production models are often subject to pooled interdependence, in which each incremental piece of work contributes to the whole without being contingent on other parts. This strategy relies on minimally sized and minimally complex work units that are completed by large numbers of contributors (Haythornthwaite 2009), which can make up for the inconsistency of participation and turn-over of contributors to maintain sustainability despite dynamic membership (Butler 2001). These principles of task design are also consistent with the design requirements for most virtual citizen science projects, which must balance volunteer recruitment and retention efforts, and in some cases may need to acquire greater numbers to achieve geographic skill, as well as making up for inconsistencies in participation.

Additional related studies focus on cyberinfrastructure (CI), that is, large-scale ICT to support distributed scientific research, which is also known as e-infrastructure, or in scientific contexts, eScience (Jackson, et al. 2007, Edwards, et al. 2009). Virtual citizen science similarly relies increasingly on ICT to overcome discontinuities inherent in massively distributed work, often with the goal of increasing the scale of participation. The level of CI support for citizen science VOs also varies substantially, from simple data collection to more sophisticated task and social support. Initial studies of CI projects have focused primarily on the same challenges as those of collaboratories, particularly coordination, geographic dispersion and social aspects of science and technology (e.g., Hepsш, et al. 2009, Lee, et al. 2006, Lawrence, et al. 2007). Unlike collaboratories, CI projects were found to have more decentralized leadership with the increasing scale of participation resulting in increasing discontinuities and reliance on ICT to moderate their effects.

However, none of these prior efforts fully address the unique characteristics of the citizen science setting, as the combination of volunteer contributions and scientific goals 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” (Surowiecki 2004, Roman 2009). Including volunteers in scientific research projects results in very different distributed organizational structures than those of scientific collaboratories, raising new challenges for effective collaboration. Current research on CI typically assumes that scientists and related professionals are the primary participants; 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 may have widely varying levels of skill or knowledge and are likely to contribute at levels differing by orders of magnitude. Combined, these factors raise unique concerns for designing systems and organizations to support citizen science. To address these concerns required development of a tailored conceptual model.

Given the similarity of citizen science VOs to other forms of massive virtual collaboration, we drew initially on our prior research on FLOSS teams by starting our work with a conceptual model developed from a review of empirical literature on FLOSS development, which itself extends an earlier framework (Crowston, et al. 2005). To adapt this model for citizen science projects, we examined the literature in related areas—most notably volunteerism and scientific collaboration—to tailor the model for the context of online participation. The original constructs were examined individually and retained, replaced or augmented by other constructs. Synthesizing elements from organization design, sociology and studies of nonprofit management with a framework based in small group theory strengthens our conceptual model for understanding the antecedents of scientific knowledge production through massive virtual collaboration.

Among the adaptations, the more contextually relevant processes of scientific research, data management and volunteer management were substituted for software development practices and firm involvement practices noted in the FLOSS setting. Similarly, the concepts of software implementation and evolution were translated into knowledge and innovation, the desired outputs of scientific knowledge production. The emergent states of roles and commitment were retained for their potential relevance to the context of citizen science VOs. Finally, social processes were included in the individual level of the model as joining and contributing processes, with the expectation that additional social processes will be revealed through empirical research. The result of this process is a conceptual model that is firmly anchored in basic theory, whilst still being adapted to the unique features of this setting.

4 Conceptual Model

In this section we present the conceptual model for further study. Figure 1 shows the initial version of our model.

As noted above, the basic structure is drawn from work on work teams. Guzzo & Dickson (1996), p. 308 defined a work team as "made up of individuals who see themselves and who are seen by others as a social entity, who are interdependent because of the tasks they perform as members of a group, who are embedded in one or more larger social system (e.g., community, or organization) and who perform tasks that affect others (such as customers or coworkers)". A team differs from a community of practice because members have a shared output, whereas in communities of practice members share common practices, but are individually responsible for their own tasks (Wenger 1999). Members of a citizen science project share a goal and social identity and they perform interdependent tasks. Although the individual tasks are typically designed to reduce reciprocal and sequential interdependencies and thereby reduce coordination costs, the collective outcome is strongly affected by pooled interdependence (Thompson, et al. 2003). Even though individual tasks seem independent, the final product comprises the collective contributions and value of each individual contribution is dependent upon others' contributions.

Drawing on further on the small groups literature, we organized our conceptual model as an input-mediator-output-input (IMOI) model (Ilgen, et al. 2005). We selected the IMOI model because it provides a theoretical framework of socially-embedded teams over time and improves on the prior IPO models of work groups by including feedback loops and separating emergent states from processes. The structure of an IMOI model categorizes constituent concepts as inputs, mediators and outputs. Inputs are the starting conditions of a team, which includes member characteristics and project/task characteristics (Hackman & Kaplan 1974). Mediators represent factors that mediate the influence of inputs on outputs and are further divided into two categories: processes and emergent states. Processes represent dynamic interactions among team members as they work on their projects, leading to the outputs. Emergent states are constructs that characterize dynamic team properties, which vary based upon context; they describe the team's cognitive, motivational and affective states, rather than activities and processes. Outputs are the task and non-task consequences of a team functioning. The feedback loop from outputs to inputs treats outputs as inputs to future team processes and emergent states; as a result, not all processes or inputs may be active at any given time, depending on the state of system functioning.

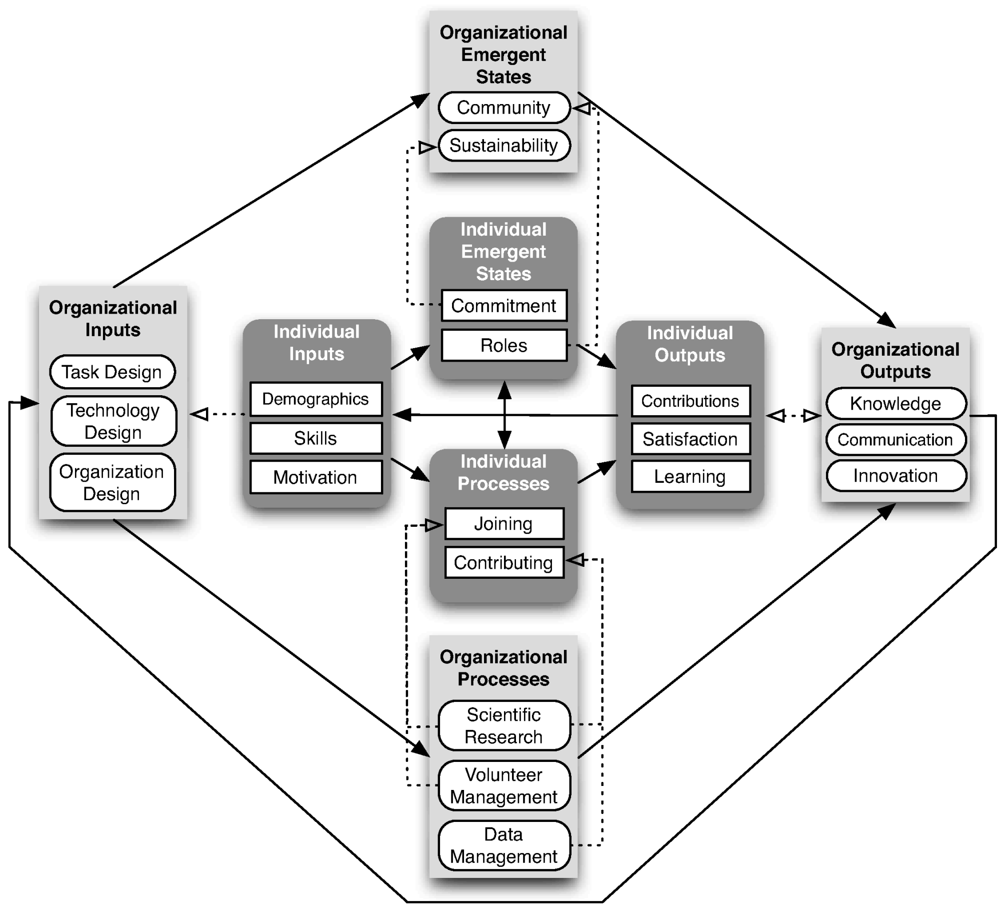

Figure 1: A conceptual model of citizen science virtual organizations.

In the remainder of this section, we present each of the elements of the model in more detail.

4.1 Inputs

Inputs are the starting conditions of a project, including both individual-level characteristics and project-level characteristics. At the individual level, staff and volunteers come to the project with diverse demographics, levels of skill and motivations for participation that affect their individual contributions to the project. While demographics and skills will vary among volunteers involved in different projects, both practical reports and academic theory suggest a number of common motivators for volunteerism, which may have differential effects on individual experiences and performance (Lawrence 2006, Cnaan & Cascio 1999).

At the organizational level, we examine the effects of organizational, task and cyberinfrastructure technology design. Organization design is a key point of differentiation between citizen science VOs and other scientific collaboratories. The configuration and geographical distribution of participants can vary widely, as can the size of the core research group, which can range from a single PI with a research assistant or two to an interorganizational network of governmental agencies, scientific researchers and nonprofit organizations, each with different interests to fulfill 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 (Crowston, et al. 2006). One important difference in citizen science projects is that there are often formal status differences that separate these groups, e.g., most core participants likely have graduate training and formal roles in the projects, whilst other participants are lay volunteers.

The second organizational input, "task design", encompasses several related concepts. Task design in this context includes the research project design for the study, the job design for volunteers and researchers and the task design for citizen science protocols. Citizen science research designs and protocols must reflect careful consideration of job design and task design (Cohn 2008). Some tasks may be feasible and interesting for volunteers, with proper design, whilst others may have to be reserved for paid professional staff. Organization design theories link individual-level inputs and outputs (motivation and performance) to the task design (Ilgen & Hollenbeck 1991), as do theories of volunteerism (Pearce 1993, Wilson 2000).

Finally, technology design and use is of particular interest given the potential of cyberinfrastructure to support citizen science VOs, in particular with data management (Chin & Lansing 2004). Best practices guides recommend that project partnerships include a scientist and an educator to address the scientific and educational goals of the project and a technologist to address potentially substantial data management and information systems challenges (Bonney & LaBranche 2004). When considering how organization design and task design interact with technology in the context of scientific VOs, understanding the range of interactions between such diverse end users and the technologies that support participation is important to creating usable, robust cyberinfrastructure systems for collecting useful contributions from distributed volunteers (Luther, et al. 2009).

At the individual level, similarities to peer production models provoke the perennial questions about motivation to participate. While participation in peer production is generally expected to be motivated by self-interest, virtual citizen science projects appear more altruistic on the surface. In practice, this perception seems partially true; for example, participants in the *Galaxy Zoo* project report multiple motivations that reflect both altruism and self-interest (Raddick, et al. 2009b). In discussing virtual citizen science practices, Raddick:2009b also emphasize the potential social benefits arising from progressive levels of participant engagement in citizen science. This general model of progressive engagement is echoed elsewhere (e.g., Preece & Shneiderman 2009, Fischer 2002) and core-periphery models of voluntary participation, much like those seen in research on more traditional work groups (e.g., Cummings & Cross 2003), are a consistent feature across a number of these domains, such as open source (e.g., Crowston, et al. 2006).

4.2 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 can vary widely, including data collection, reduction and analysis tasks. Understanding these work practices is the first key to designing and supporting technological and social arrangements that support intellectual production and innovation in virtual citizen science projects.

At the organizational level, the processes of interest include that of scientific research itself. The nature of the research and discipline has an important influence on the kinds of data and analysis required and the mapping of tasks to different actors, e.g., volunteers or professional staff. Similarly, data management processes have a significant impact on project outcomes. A particular concern is the challenge facing interorganizational 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, selection, orientation, training, supervision, evaluation, recognition and retention of volunteers (Pearce 1993).

4.3 Emergent States

Emergent states are dynamic properties of the group that vary as a function of inputs and processes; past research suggests a number of potentially relevant emergent states. These 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 (Lee, et al. 2006). At the level of the project, research on other kinds of VOs has identified the importance of factors such as trust, cohesion, conflict and morale that affect the sense of group community and thus its long-term sustainability (Markus, et al. 2000).

At the individual level, the evolution of volunteers through different roles in the group, from new volunteer through sustained contributor and potentially to more central roles, is relevant to organization design. A related concern is volunteers' level of commitment to the project and how it influences their task performance (Cnaan & Cascio 1999). Understanding how these factors affect the social and technological barriers to and enablers of participation is important for effective cyberinfrastructure and organization designs.

Processes and emergent states are conceptualized as moderating the relation between inputs and outputs of the project in the IMOI model. At the individual level, the input elements of organization, task and technology design affect motivation and participation of distributed volunteers (Lawrence 2006, Sproull & Kiesler 2005). At the project level, they may transform the means of production of scientific knowledge, shaping the demand for supporting cyberinfrastructure and potentially transforming the organization design.

4.4 Outputs

Finally, outputs represent task and non-task consequences of a functioning group, signaling effectiveness. At the individual level, the task outputs are contributions, often raw or processed data, although other contributions may be possible. In addition to the individual-level outputs, a citizen science VO will have outputs at the project level, such as the scientific knowledge created from the data. Innovative findings, processes and tools can also emerge from involving the public in scientific research, with the potential for dramatic discoveries like Hanny's Voorwerp.

Hackman's (1987) model of group effectiveness 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 the sustainability of the project's goals and social structure. In other words, a VO is not effective if it achieves a goal but drives away participants in the process.

An important feature of the IMOI model is that outputs themselves become future inputs to the dynamic processes. Positive personal outcomes can lead to increased motivation for future participation and individual learning can increase a member's ability to contribute. At the project level, learning may lead to innovation in research approach, resulting in changes to the task design and group processes. Positive project outputs may lead to increased interest among practitioners in engaging the public in research and increased visibility for the project, helping to recruit and retain additional volunteers. At the societal level, the success of a project may affect public participation in and perception of science, create informal learning opportunities and enable knowledge production at an unprecedented pace and scale (Trumbull, et al. 2000, Cohn 2008).

4.5 Relationships Between Levels

To capture the interplay between dynamics at the level of individual members and the overall project, in this model, we nested two IMOI models. The organizational and individual levels are differentiated in order to better guide organization design and engineering goals while recognizing that many of the individual-level characteristics that are relevant to decision-making are outside of the direct control of project organizers. For example, the member characteristics identified as inputs in the FLOSS model are separated from those factors that are more readily influenced by the top-down styles of organizing that dominates current citizen science practices.

In the model, the individual level and organizational level IMOI models are connected with dashed lines, representing potentially interesting connections between individual participation and organizational operation. These connections are drawn from the literature on organizational behavior and volunteerism to better reflect the phenomenon, as the organizational behavior literature makes assumptions about the relationships between individuals and organizations which must be validated against findings from studies of volunteer management for transfer to a context of voluntary participation.

For example, the joining process at the individual level—typically an online registration—are affected by the volunteer management processes and scientific research processes occurring at the organizational level. Although surveying a sample of citizen science projects confirms that relatively few projects have volunteer screening processes that would prevent individual participation based on member characteristics, an example of this relationship can be seen in another astronomy citizen science project, NASA's *Stardust@home*[4]. In order to ensure quality outcomes, would-be participants must pass a relatively challenging test to demonstrate adequate skill in using the "Virtual Microscope" to locate particles collected from outer space before online registration is even permitted. The link between these joining and research processes, as well as other connections shown in the IMOI model, makes explicit our expectations about the broader context of citizen science VOs. These relationships can be examined

[4]http://stardustathome.ssl.berkeley.edu/

more closely through empirical studies to verify or disconfirm the usefulness of these links and concepts, providing insight for revisions and further model development.

5 Discussion and Future Work

The IMOI model presented in this paper provides an example of using an existing theoretical framework as a template for a contextualized conceptual model suitable for guiding research in a novel context. In this section, we discuss the exploratory research it guides.

This conceptual model is currently being employed as part of an exploratory study. The goal of the study is to develop a refined model for further research, and to guide cyberinfrastructure and organization design for citizen science VOs. The larger study design involves developing a typology of citizen science projects, which provides input to refinement of the conceptual model. The typology will provide a resource for multiple practitioner audiences and is guiding case selection for in-depth studies to validate and further refine the conceptual model.

For the initial stage of research, semi-structured interviews with citizen science project organizers focused on understanding the inputs and processes at the project and individual levels. Therefore, interview questions inquired into how organizers became involved in their current projects and how they were involved with the scientific, volunteer and data management aspects of the project. The participants' responses confirmed the relevance of the context-specific organizational inputs and processes and touched on several other constructs identified in our model.

Participant observation provided triangulation for the interview responses and fieldwork also demonstrated the relative influence of some factors in different contexts of activity. However, inductive content analysis coding also yielded additional themes to consider. Some of these are more specific types of inputs, outputs, or processes; for example, the relevant scientific research processes can refer to study development, pilot testing, data collection, analysis, or quality control, among other processes. Each of these processes is differently impacted by the inclusion of the public in scientific research. Most other emergent themes relate to the physical contexts and cultural environments in which the project operates (Wiggins 2010), which are characteristics that don't fit neatly into the current model, suggesting a point for future revision of the model.

The conceptual model described here has provided direction for initial exploratory research and suggested directions for future work to further develop and validate the model. In particular, the preliminary findings suggest that adopting the project rather than the organization as the unit of analysis will make the model more consistent with empirical observation without imposing the assumptions about organizational arrangements that are implicit in the current conceptual model. Empirical research found that the standard organizational forms can be overly simplistic or otherwise inadequate as a basis for understanding organizing in citizen science VOs (Wiggins 2010). The project level of group interaction is distinct from those of small work groups and organizations (Grudin 1994), which has implications for organization design efforts. Project teams and communities of practice can be distinguished by their goal orientation among other features (Wenger 1999), but empirical observation of citizen science VOs to date indicates a hybrid "community of purpose" might better describe many projects, with characteristics of both a project team and a community of practice or interest. Revisions based on the initial empirical findings will strengthen the model by reconceptualizing structures and processes to better accommodate the wide variety of ways that projects organize their activities.

6 Conclusion

In summary, synthesizing elements of prior research on small groups with contextually relevant concepts provides a theoretical foundation for studying the organization of large numbers of virtual volunteers for scientific research. Several differences in settings between prior work and the citizen science VO context suggest the need to both validate the applicability of this body of theory and search for possible extensions.

This paper contributes a multi-level design-oriented conceptual model for organizing investigation into the impact of design on scientific research by distributed volunteers in collaboration with scientists. The discussion also describes the process of model development through contextualization of an existing theoretical framework. Finally, the conceptual model complements the prior research on scientific collaboratories and cyberinfrastructure projects by developing a theoretical foundation to understanding the design of virtual organizations that involve the public in scientific research.

7 Acknowledgements

This work was partially supported by U.S. NSF grant 09-43049, "VOSS: Theory and design of virtual organizations for citizen science".

References

C. Baretto, et al. (2003). 'A Model for Integrating the Public into Scientific Research'. *Journal of Geoscience Education* **50**(1):71–75.

Y. Benkler (2002). 'Coase's Penguin, or, Linux and the Nature of the Firm.'. *Yale Law Journal* **112**(3):367–445.

R. Bonney & M. LaBranche (2004). 'Citizen Science: Involving the Public in Research'. *ASTC Dimensions* p. 13.

R. Bonney & J. L. Shirk (2007). 'Citizen Science Central'. *Connect* pp. 8–10.

D. Brossard, et al. (2005). 'Scientific Knowledge and Attitude Change: The Impact of a Citizen Science Project'. *International Journal of Science Education* **27**(9):1099–1121.

B. Butler (2001). 'Membership size, communication activity, and sustainability: A resource-based model of online social structures'. *Information Systems Research* **12**(4):346–362.

G. Chin, Jr. & C. Lansing (2004). 'Capturing and supporting contexts for scientific data sharing via the biological sciences collaboratory'. In *Proc. CSCW 2004*, pp. 409–418.

A. Cho & D. Clery (2009). 'International year of astronomy: Astronomy Hits the Big Time'. *Science* **323**(5912):332.

F. Clark & D. Illman (2001). 'Dimensions of civic science: Introductory essay'. *Science communication* **23**(1):5.

R. A. Cnaan & T. A. Cascio (1999). 'Performance and commitment: Issues in management of volunteers in human service organizations'. *Journal of Social Service Research* **24**:1–38.

J. P. Cohn (2008). 'Citizen Science: Can Volunteers Do Real Research? '. *BioScience* **58**(3):192–107.

C. B. Cooper, et al. (2007). 'Citizen Science as a Tool for Conservation in Residential Ecosystems'. *Ecology and Society* **12**(2).

K. Crowston, et al. (2005). 'Effective work practices for FLOSS development: A model and propositions'. In *Proceedings of the 38th Hawai'i International Conference on System Sciences*, Big Island, Hawai'i.

K. Crowston & B. Scozzi (2008). 'Coordination practices within Free/Libre Open Source Software development teams: The bug fixing process'. *Journal of Database Management* **19**(2):1–30.

K. Crowston, et al. (2006). 'Core and periphery in Free/Libre and Open Source software team communications'. In *Hawai'i International Conference on System Sciences*, vol. 39, p. 118. IEEE.

J. Cummings & R. Cross (2003). 'Structural properties of work groups and their consequences for performance'. *Social Networks* **25**(3):197–210.

D. G. Delaney, et al. (2007). 'Marine invasive species: validation of citizen science and implications for national monitoring networks'. *Biological Invasions* pp. 1573–1464.

P. Edwards, et al. (2009). 'Introduction: An Agenda for Infrastructure Studies'. *Journal of the Association for Information Systems* **10**(5):364–374.

T. Finholt (2002). 'Collaboratories'. *Ann. Rev. Info. Sci. Tech.* **36**(1).

K. Firehock & J. West (1995). 'A Brief History of Volunteer Biological Water Monitoring Using Macroinvertebrates'. *Journal of the North American Benthological Society* **14**(1):197–202.

G. Fischer (2002). 'Beyond 'couch potatoes': From consumers to designers and active contributors'. *First Monday* **7**(12-2).

L. Fore, et al. (2001). 'Assessing the performance of volunteers in monitoring streams'. *Freshwater Biology* **46**:109–123.

A. Forte & A. Bruckman (2008). 'Scaling consensus: Increasing decentralization in Wikipedia governance'. In *Proceedings of the 41st Hawai'i International Conference on System Sciences*, Big Island of Hawai'i. IEEE.

P. A. Freeman, et al. (2005). 'Cyberinfrastructure for science and engineering: Promises and challenges'. *Proceedings of the IEEE* **93**(3):682–691.

J. Grudin (1994). 'Groupware and social dynamics: eight challenges for developers'. *Communications of the ACM* **37**(1):92–105.

R. A. Guzzo & M. W. Dickson (1996). 'Teams in organizations: Recent research on performance and effectiveness'. *Annual Review of Psychology* **47**:307–338.

J. Hackman (1987). 'The design of work teams'. *Handbook of organizational behavior* **315**:342.

J. R. Hackman & R. Kaplan (1974). 'Interventions into group process: An approach to improving the effectiveness of groups'. *Decision Sciences* **5**:459–480.

J. R. Hackman & C. G. Morris (1978). 'Group tasks, group interaction process, and group performance effectiveness: A review and proposed integration'. In L. Berkowitz (ed.), *Group Processes*, vol. 8 of *Advances in Experimental Social Psychology*, pp. 45–99. Academic Press, New York.

C. Haythornthwaite (2009). 'Crowds and Communities: Light and Heavyweight Models of Peer Production'. In *Proceedings of the Hawai'i International conference on Systems Sciences, 2009*.

V. Hepsш, et al. (2009). 'Ecologies of e-Infrastructures'. *Journal of the AIS* **10**(5):430–446.

T. Hey & A. E. Trefethen (2005). 'Cyberinfrastructure for e-science'. *Science* **308**(5723):817–821.

J. Howe (2006). 'The rise of crowdsourcing'. *Wired Magazine* **14**(6):1–4.

D. R. Ilgen & J. R. Hollenbeck (1991). 'The structure of work: Job design and roles'. *Handbook of Industrial and Organizational Psychology* **2**:165–207.

D. R. Ilgen, et al. (2005). 'Teams in organizations: From Input-Process-Output Models to IMOI Models'. *Annual Review of Psychology* **56**:517–543.

S. Jackson, et al. (2007). 'Understanding infrastructure: history, heuristics, and cyberinfrastructure policy'. *First Monday* **12**(6).

M. Krasny & R. Bonney (2005). *Environmental Education Through Citizen Science and Participatory Action Research*, chap. Ch. 13. Environmental education and advocacy : changing perspectives of ecology and education. Cambridge University Press, New York.

A. Lawrence (2006). "No Personal Motive? ' Volunteers, Biodiversity, and the False Dichotomies of Participation'. *Ethics, Place & Environment: A Journal of Philosophy & Geography* **9**(3):279–298.

K. Lawrence, et al. (2007). 'Warm Fronts and High Pressure Systems: Overcoming Geographic Dispersion in a Meteorological Cyberinfrastructure Project'. In *Hawaii International Conference on Systems Sciences*, vol. 40, p. 675. IEEE.

C. P. Lee, et al. (2006). 'The human infrastructure of cyberinfrastructure'. In *Proceedings of the Conference on Computer-supported Cooperative Work (CSCW)*, pp. 483–492, New York, NY, USA. ACM.

K. Luther, et al. (2009). 'Pathfinder: an online collaboration environment for citizen scientists'. In *Proceedings of the Conference on Human Factors in Computing Systems(CHI)*, pp. 239–248.

M. A. Marks, et al. (2001). 'A Temporally Based Framework and Taxonomy of Team Processes'. *Academy of Management Review* **26**(3):356–376.

M. Markus, et al. (2000). 'What makes a virtual organization work? '. *Sloan Management Review* **42**(1):13–26.

J. E. McGrath & A. B. Hollingshead (1994). *Groups Interacting With Technology*. Sage, Thousand Oaks, CA.

O. Nov (2007). 'What motivates Wikipedians? '. *Communication of the ACM* **50**(11):60–64.

J. Pearce (1993). *Volunteers: The organizational behavior of unpaid workers*. Routledge.

J. Preece & B. Shneiderman (2009). ‘The Reader-to-Leader Framework: Motivating Technology-Mediated Social Participation’. *AIS Trans. on Hum.-Comp. Interact.* **1**(1):13–32.

M. Raddick, et al. (2009a). ‘Citizen Science: Status and Research Directions for the Coming Decade’. In *AGB Stars and Related Phenomenastro2010: The Astronomy and Astrophysics Decadal Survey*, vol. 2010, p. 46P.

M. Raddick, et al. (2009b). ‘Galaxy Zoo: Exploring the Motivations of Citizen Science Volunteers’. *Arxiv preprint arXiv:0909.2925* .

D. Roman (2009). ‘Crowdsourcing and the question of expertise’. *Communications of the ACM* **52**(12):12.

J. Silvertown (2009). ‘A new dawn for citizen science’. *Trends in Ecology & Evolution* **24**:467–471.

L. Sproull & S. Kiesler (2005). ‘Public volunteer work on the Internet’. *Transforming Enterprise: The Economic and Social Implications of Information Technology* p. 361.

J. Surowiecki (2004). *The wisdom of crowds: Why the many are smarter than the few and how collective wisdom shapes business, economies, societies, and nations*. Doubleday Books.

J. Thompson, et al. (2003). *Organizations in action: Social science bases of administrative theory*. Transaction Pub.

D. J. Trumbull, et al. (2000). ‘Thinking scientifically during participation in a citizen-science project’. *Science Education* **84**(2):265–275.

E. Wenger (1999). *Communities of practice: Learning, meaning, and identity*. Cambridge University Press.

A. Wiggins (2010). ‘Organizing from the Middle Out: Citizen Science in the National Parks’. In *Proceedings of iConference 2010*.

J. Wilson (2000). ‘Volunteering’. *Annual Review of Sociology* **26**:215–240.