# SoCS: Socially intelligent computing to support citizen science

The proposed research project will investigate the capabilities and potential of social computational systems (SoCS) in the context of citizen science, that is, research projects involving "partnerships between volunteers and scientists that answer real-world questions". The projects to be studied include a core of scientists and project leaders coordinating the work of a larger number of distributed volunteer contributors in a variety of scientific disciplines. These projects are a form of social-computational system. Whether it be volunteers playing a role in massive, distributed sensing networks exploring the migration of birds, or applying their unique human perceptual skills to searching the skies, human motivation and performance is fundamental to system performance. However, undertaking science through a social computational system brings unique challenges. To understand and address these challenges, this proposal presents a three-phase study of SoCS to support scientific research, grounded in group theory and rooted empirically in case studies and action research.

The proposed research project has the following three specific goals:

1. developing a practical understanding of the conditions (cognitive, social or cultural) under which SoCS can enable and enhance scientific and education production and innovation in citizen science projects;
2. generating new models of SoCS that support large-scale public participation in scientific research; and
3. developing and testing SoCS that reflect explicit knowledge about human cognitive and social abilities in this setting.

More specifically, the proposal includes case studies of several citizen science projects to establish the nature of the SoCS currently in use, development of SoCS to support different kinds of citizen science projects and evaluation of the impacts of these systems on the outputs and processes of the projects.

*Expected intellectual merits.* The proposed study will make theoretical and practical contributions.

- *Theoretical contributions.* This research will add to our understanding of large-scale SoCS by focusing on the sociotechnical structures and processes involved in production of scientific knowledge in research projects undertaken in partnerships between professionals and volunteers. The study will contribute to theory by refining and validating a conceptual framework of the relationships between organizational structure, work design and computing technologies in use, which will be used to guide systems design and evaluation.
- *Practical contributions.* The research will gather requirements for technologies to support a diverse set of citizen science projects, leading to development and evaluation of specific instances of generic participation platforms for social computational systems. Results will aid scientists and project leaders (the "practitioners" of citizen science) in identifying appropriate social and computational technologies to employ when improving cyberinfrastructure for current projects or launching new ones.

*Expected broader impacts.* The project will benefit society by:

- investigating how involving the public in scientific research through socially intelligent computing systems can advance science directly, in addition to goals of outreach or informal learning;
- generating and disseminating insights directly applicable to improving the design and implementation of SoCS for citizen science projects, thereby improving the available technologies through which the public engages in scientific research;
- determining the conditions under which citizen science projects provide a solution for large-scale data collection, as well as opportunities to leverage public interest in other aspects of scientific knowledge production; and
- creating empirically-based theoretical models of social computing in large-scale contribution systems, which are increasingly used to create public goods.

As well, the project will contribute to the education of doctoral, masters and undergraduate students who will learn about research and system development through their participation in the proposed project.

# SoCS: Socially intelligent computing to support citizen science

Innovations in information and communication technologies (ICTs) are bringing humans and computers together in powerful new ways. Of particular note, the widespread deployment of ICT has enabled new options for distributed collaboration. Phenomena such as free/libre open source software (FLOSS) development, Wikipedia editing and other forms of online interaction (e.g., Adler et al., 2008; Crowston and Howison, 2006b; Crowston and Scozzi, 2008; Forte and Bruckman, 2008; Nov, 2007; Xu et al., 2009) prompt us to consider the potential of social-computational systems (SoCS) for new forms of engagement and collective action. Many of these large-scale social-computational systems are producing increasingly valuable and freely available public goods. Our study will investigate the capabilities and potential of SoCS in the context of citizen science, that is, research projects involving "partnerships between volunteers and scientists that answer real-world questions"[1] (Bonney and Shirk, 2007; Clark and Illman, 2001; Cohn, 2008; Trumbull et al., 2000).

Citizen science takes many forms, but a number of projects have evolved from long-standing programs employing volunteer monitoring for natural resource management (Ballard et al., 2005; Cooper et al., 2007; Firehock and West, 1995). The approach is often employed as a form of education and outreach to promote public understanding of science (Baron, 2004; Bauer et al., 2000; Brossard et al., 2005; Krasny and Bonney, 2005; Osborn et al., 2002). However, citizen science projects are increasingly focused on enabling benefits to the scientific research as well (Baretto et al., 2003; Bonney and LaBranche, 2004; McCaffrey, 2005). 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 (Brewer, 2002; Fore et al., 2001; Trumbull et al., 2000) obtained by a broad range of participants, as well as leading to unexpected insights and innovations (Lee et al., 2006; Seattle Aquarium, 2005).

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 enquiry. 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 biological and environmental 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. By contrast, in projects organized by researchers in astronomy, such as NASA's Clickworkers (Gulick et al., 2007), volunteers apply superior human perceptual capacities to computationally difficult image recognition tasks, providing an important service in data analysis and demonstrating the power of human-computer partnerships to address complex tasks.

This type of public contribution is not new to science (e.g., the Audubon Christmas survey of birds 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 projects (Onsrud and Campbell, 2007; Onsrud et al., 2004). The potential benefits of citizen science are beginning to be realized more widely, particularly when coupled with traditional scientific studies (Dufour and Crisfield, 2008), leading to a rapid increase in the number of projects (the blog `http://citizensci.com/` lists dozens of examples). Bhattacharjee (2005) notes that use of citizen science research methods were previously seen as a barrier to obtaining research funding from NSF, but more recently NSF has started funding such projects based on their scientific merit.

Citizen science projects are similar in some respects to other virtual collaborations such as FLOSS, Wikipedia or scientific collaboratories (Finholt, 2001; Finholt and Olson, 1997) but are united by scientific goals that pose particular constraints on task and system design. First, reliability of data collection is critical to the value of a scientific project, and can not simply be left to the "wisdom of crowds". Second, including volunteers in scientific research projects results in a unique distributed organizational structures, raising new challenges for project leaders to manage. For example, the design of collaboration

---

[1] from Citizen Science Central `http://www.birds.cornell.edu/citscitoolkit/`

tools may tacitly assume that participants have comparable and high levels of skill and will contribute relatively equally. However, equal skill or participation is rarely the case for citizen science volunteers, who may have widely varying levels of skill or knowledge, and contribute at levels differing by orders of magnitude. These factors raise unique concerns for the development of SoCS for citizen science.

The increasing scale of citizen science projects, some of which involve tens of thousands of members of the public in distributed data collection and analysis, suggests a need for additional research on appropriate human-computer partnerships to support this kind of scientific work. To begin, further study is needed to understand how ICTs are currently employed in citizen science projects, and the functional and cognitive demands of the particular research tasks that need to be supported. The current level of computing support varies, from simple data collection to more sophisticated task support. Many projects are very low tech; for example, the Great Sunflower Project[2] is studying pollinator service (bee activity) supported by a simple installation of the Drupal open source content management system (CMS) to manage data collection by seventy-seven thousand volunteers across the continental United States. By contrast, other systems that support citizen science projects are quite sophisticated and carefully designed to address user needs, e.g. the various projects from Cornell Lab of Ornithology (CLO), which currently engage three hundred thousand participants annually. Similarly, projects vary greatly in the level of computation involved, from simple data gathering and aggregation, to various kinds of data checking and validation, to search and retrieval of data, to complex forms of analysis. Data validation is a particularly persistent and multi-faceted challenge for projects collecting observations from thousands of distributed, independent participants. While it can be addressed through either social or computational approaches, there is little research on how to effectively employ combinations of social and computational approach to create a SoCS for data validation. Search and retrieval is likewise a more complex matter from both social and computational perspectives than it initially appears, as user needs and expertise are largely unexplored, but are expected to be unusually diverse in citizen science projects, driven by the diversity of participants.

Beyond documenting the current state, research needs to address the conceptualization and design of computing systems that can support mutually beneficial outcomes from public participation in scientific research. For most scientifically-oriented endeavours, a primary reason to use citizen science is to enable economies of scale. Using ICTs allows unprecedented scalability for some projects. More ambitiously, systems could allow unprecedented kinds of analysis of the data. In particular, there is significant interest from organizations involved in large-scale ICT-enabled citizen science for developing tools to support analysis and inquiry by non-scientists. These tools would be particularly beneficial for science education, and would be employed in both informal and formal learning contexts. However, computational systems to support citizen science must be developed through integrated social and computational research in order to truly deliver on the potential innovation and learning benefits from large-scale public participation in scientific research.

Furthermore, there is need for flexible, generic platforms to support scalable, sustainable citizen science participation, which will reduce resources/time spent on systems development as a whole, while supporting best practices. Developing platforms to support reuse of the technology for additional areas of scientific inquiry maximizes the return on investment for research funders such as the NSF. As an example of a potential model for a citizen science platform, CLO's eBird system has been implemented to support ornithology research; in 2006, volunteers contributed over 4.3 million observations, and participation continues to grow. eBird is frequently mentioned by practitioners as a gold standard for current technologies enabling large-scale distributed citizen science projects, and the technology draws interest from scientists outside of ornithology as a form of social-computational system that they would like to implement for research on other taxa. This platform in particular enables the development of shared data resources for a research community, supporting a variety of socially desirable outcomes for advancing scientific research. However, this is just one example of the many types of citizen science projects, so further research is needed to establish the common problems and opportunities for this mode of social-computational engagement.

---

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

In response to this need, the current proposal presents a three-phase, theory-based study of the sociotechnical aspects of massive collaboration of volunteer participants in scientific research, leading to the design and testing of SoCS for citizen science research projects. The proposed study is grounded in social theory, and rooted empirically in case studies and action research in citizen science projects. The overall aim of the proposed research is to identify key requirements for and to build and evaluate social-computational systems for enabling citizen science projects to involve distributed, diverse volunteers in producing large scale, high quality, valued scientific research in an organizationally sustainable fashion. The proposed research has the following goals:

1. developing a practical understanding of the conditions (cognitive, social or cultural) under which socially intelligent computing can enable and enhance scientific and education production and innovation in citizen science projects;
2. generating new models of social computing to describe large-scale public participation in scientific research; and
3. developing and testing computational systems that reflect explicit knowledge about human cognitive and social abilities in this setting.

More specifically, the proposal includes case studies of several citizen science projects to establish the nature of the SoCS currently in use, development of SoCS to support different kinds of citizen science projects and evaluation of the impacts of these systems on the outputs and processes of the projects. This work will thus build on and extend our ongoing research on citizen science projects, supported by a newly awarded grant from the NSF Virtual Organizations as Sociotechnical Systems program ("the VOSS grant"). That grant is supporting development of a typology of citizen science projects and some initial case studies of projects; the focus on building and evaluating social computational systems distinguishes the current proposal from the VOSS grant.

*Expected intellectual merits.* The proposed study will make the following theoretical and practical contributions.

- *Theoretical contributions.* This research will add to our understanding of large-scale social computing by focusing on the sociotechnical structures and processes involved in production of scientific knowledge in research projects undertaken in partnerships between professionals, volunteers and, potentially, computational systems. The study will contribute to theory by refining and validating a conceptual framework of the relationships between organizational structure, work design and computing technologies in use, which will be used to guide systems design and evaluation.
- *Practical contributions.* The research will gather requirements for technologies to support a diverse set of citizen science projects, leading to development and evaluation of specific instances of generic participation platforms for social-computational systems. Results will aid scientists and project leaders (the "practitioners" of citizen science) in identifying appropriate social and computational technologies to employ when improving systems for current projects or launching new ones.

*Expected broader impacts.* The project will benefit society by:

- investigating how involving the public in scientific research through socially intelligent computing systems can advance science directly, in addition to goals of outreach or informal learning;
- generating and disseminating insights directly applicable to improving the design and implementation of social-computational systems for citizen science projects, thereby improving the available technologies through which the public engages in scientific research;
- determining the conditions under which citizen science projects provide a solution for large-scale data collection, as well as opportunities to leverage public interest in other aspects of scientific knowledge production; and
- creating empirically-based theoretical models of social computing in large-scale contribution systems, which are increasingly used to create public goods.

As well, the project will contribute to the education of doctoral, masters and undergraduate students who will learn about research and system development through their participation in the proposed project.

The remainder of this proposal is organized into five sections. In section 2, we develop a conceptual framework for our study. In section 3, we present the study design, with details of data collection and analysis. In section 4, we present the project management plan. We conclude in section 5 by reviewing the intellectual merits and expected broader impacts and results of prior NSF support.

## 2 Conceptual development

In this section we develop the initial conceptual framework for our study. For this project, we have chosen to analyze citizen science projects as a kind of work team. Guzzo and 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 typically individually responsible for their own tasks. Adopting this perspective provides a useful structure for other theorizing and allows us to draw from the extensive research on work teams, thus providing a strong theoretical starting point for our project.

For the current study, our conceptual framework draws first on work in the group literature (e.g., Hackman and Morris, 1978; Marks et al., 2001; McGrath and Hollingshead, 1994). However, in recognition of the significant differences between citizen science projects and the kinds of groups on which these theories were based, we augment this model by drawing on other theoretical perspectives, incorporating elements from organizational design, sociology (social movements theory in particular) and studies of nonprofit management. The model thus synthesizes concepts and relationships from the literature on organizational design, job design, volunteerism and participation in virtual communities, at both individual (i.e., volunteer, staff member) and organizational/project levels. As well, we explicitly consider the role of technology in the human-technology partnerships in this setting.

Given the similarity of citizen science projects to other forms of massive virtual collaboration such as FLOSS, we draw in particular on our prior research on FLOSS teams. Figure 1 shows the initial version of our framework (slightly modified from the VOSS grant), which is adapted from one we developed from a review of literature on FLOSS development (Crowston et al., Under review), and which extends the framework that was the basis for our earlier NSF-funded research (Crowston et al., 2005a). We note though that most of our research on FLOSS focused on the core development team, while in the citizen science context, we will be equally or more concerned with the volunteers who are peripheral to the core and who may have only limited interactions with one another and with project leaders. These team members may vary greatly not only in their scientific and technical skills and interests, but also in their degree of identification with and contribution to the project, so the importance of these factors will be explored explicitly rather than taken for granted.

We organize our conceptual framework as an input-mediator-output-input (IMOI) model (Ilgen et al., 2005). Inputs are the starting conditions of a team, which includes member characteristics and project/task characteristics (Hackman and Morris, 1978). 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 properties of the team that are typically dynamic in nature and vary as a function of team context, inputs, processes and outcomes" (Marks et al., 2001, p. 357). Outputs are the task and non-task consequence of a team functioning (Martins et al., 2004). For example, outcomes for a citizen science project can include scientific data collected (a task output) as well as the volunteer learning about the science (a non-task output from the point of view of the research itself). Finally, the framework includes feedback loops between outputs and inputs, treating outputs also as inputs to future group processes and emergent states (Ilgen et al., 2005). In the remainder of this section, we briefly describe each of the elements of the framework and relations among them.

## 2.1 Inputs

Inputs are the initial conditions of a project, drawn from the surrounding environment and affected by prior project outputs. We include both individual level characteristics (the volunteers and the project staff) and project/task 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 (e.g., volunteers may range from school children working on a class project to retirees with extensive experience or even formal training in the field). 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 in citizen science projects (Christie, 2004; Lawrence, 2006; Sergent and Sedlacek, 1990).

At the organizational level, we will examine the effects of organizational, task and technology design. Organizational design is a key point of differentiation between citizen science projects 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. The scale of participation is another differentiating feature: while large-scale scientific collaboratories have yielded research papers with as many as 2500 contributing authors, citizen science projects may involve a contributor base that is several orders of magnitude larger. However, the overall structure of these projects seems likely to parallel the "onion" structure that describes many FLOSS projects: a core of highly involved project leaders, surrounded by a larger group of active volunteers and a still larger group of occasional

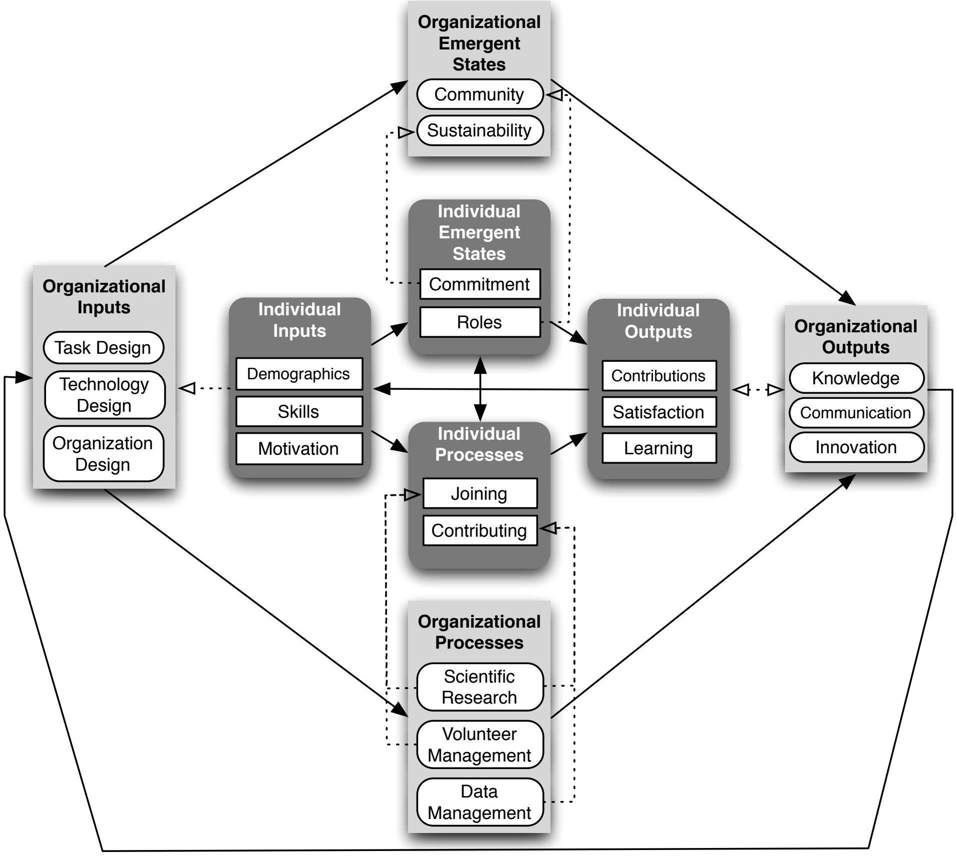

Figure 1: Initial conceptual framework: a multilevel input-mediator-output-input model

contributors (Crowston et al., 2005a). 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 as staff or advisers to the projects, while other participants are lay volunteers.

The second organizational input, task design, encompasses several related concepts, including the research design for the study, the job design for volunteers and researchers and the task design for citizen science protocols. Citizen science as a mode of production is likely suited only to particular types of research, so research designs and protocols must reflect careful consideration of job design and task design (Bonney and Shirk, 2007; Cohn, 2008; Pilz et al., 2005; Prysby and Super, 2006; Wilderman, 2007). Some tasks may be feasible and interesting for volunteers, with proper design, while others may have to be left to paid professional staff. While some researchers have carefully honed research and protocol design configurations for effective data collection by volunteers (de Solla et al., 2005; Fore et al., 2001), it is not clear whether the willingness, interest and diverse skills of the volunteers are fully engaged (Dufour and Crisfield, 2008). There may be room for volunteers to contribute productively to additional aspects of scientific research, within the appropriate enabling structures.

Finally, technology design and use is of particular interest given the potential of socially intelligent computing to support data collection, processing and management in citizen science projects. The increased diffusion of powerful ICTs into consumer products (e.g., PCs, high speed Internet, GPS, digital imaging), increases the potential for widespread deployment of technology enabling new modes of interaction and wider participation. In recognition of the importance of technology design, the Appalachian Trail (AT) MEGA-Transect Project's Citizen Science Working Group report recommends that project partnerships include technologists to help address the potentially substantial data management and information systems challenges (ATC Citizen Science Working Group, 2006), in addition to scientists and educators to address the scientific and educational goals of the project.

When considering how organizational design and task design interact with computing in the context of citizen science projects, the entire research process must be examined. For citizen science projects, concerns over volunteers' ability to use instrumentation and the usability of data reporting forms (and subsequent usability of the data they capture) has prompted careful attention to usability testing of technologies designed for volunteers (ATC Citizen Science Working Group, 2006; Prysby and Super, 2006). However, not all projects have the financial resources or available expertise to engage in this design process, which leads to concerns over the scientific impact of the choices of technologies in use. In particular, preliminary reports from practitioners suggest that data validation, reporting, and analysis tools are routinely overlooked or left for later development. Because reporting and analysis provide means of communication and feedback to volunteers, which is universally considered by project leaders important to volunteer recruitment, motivation, and retention, this practice may undermine the long-term sustainability of projects that are not initially able to invest in development of computational tools that address the social and intellectual needs of the project participants. These are only a few such considerations; understanding the range of interactions between such diverse end users and technologies that support this form of scientific research is important to creating usable, robust systems for collecting usable independent contributions by distributed volunteers (Cataldo et al., 2006).

## 2.2 Processes

In the IMOI model, the inputs described above are conceptualized as affecting the effectiveness of projects through two sets of moderators, processes (described in this section) and emergent states (described in the following section). Processes are the dynamic interactions among group members as they work on their projects, leading to the outputs. Understanding these work practices is key to answering our first question regarding the conditions (cognitive, social or cultural) under which socially intelligent computing can enable and enhance scientific and education production and innovation in citizen science projects.

At the individual level, processes of interest include joining a group, participating and making contributions (Baretto et al., 2003; von Krogh et al., 2003). Individuals can participate in and contribute to a project at a range of levels. A minimal level involves simply providing computing resources (e.g., SETI@home) or serving as a subject for research (e.g., by joining a subject pool for online surveys). Even in these cases, participants may benefit by learning about the scientific goals and methods of the project. A higher level of participation may involve data collection or monitoring (e.g., the Audubon Society's Christmas Bird Count), usually focused on the volunteers' local environment. Similarly, participants might do basic data analysis (NASA's Clickworkers), contributing their human perceptual and knowledge organization capabilities. With training, some volunteers participate at still higher levels, contributing to research through problem formulation or hypothesis testing (Trumbull et al., 2000), assisting in running projects by selecting sites, revising protocols, supervising or training other volunteers, and even by making novel contributions (Prysby and Super, 2006; Seattle Aquarium, 2005).

At the organizational level, processes of interest include the process of scientific research itself, throughout the data life-cycle. A key issue here is the nature of the science being done, the kinds of data and analysis required and the mapping of tasks to different actors, e.g., volunteers or staff. Similarly, the processes employed by the project for data management will have a significant impact on the project outcomes (ATC Citizen Science Working Group, 2006). A particular concern for scientific research are the processes for ensuring data reliability and validity. While some scientists are skeptical about the value of volunteers' data, at the same time, novel methods of validating data through crowdsourcing have emerged. One example is the GalaxyZoo project[3], through which over a million astronomical images have been evaluated, each evaluated independently by several dozen distributed volunteers to provide a reliability check. In addition, volunteers must pass a quiz to verify their ability to adequately identify galaxies before they can participate. The North American Bird Phenology Project[4] is also employing this approach in creating a digital data archives of an enormous historical data set. For this project, virtual volunteers are contributing to the transcription of more than six million scanned handwritten records of migratory bird sightings, collected by over three thousand volunteers between 1880 and 1970.

Finally, a unique aspect of citizen science projects within the larger context of scientific research practice is the applicability of volunteer management practices more often associated with nonprofit organizations, e.g., recruitment, selection, orientation, training, supervision, evaluation, recognition and retention of volunteers (Farmer and Fedor, 1999; Gordon and Babchuk, 1959; Hange et al., 2002; Wilson and Pimm, 1996). In addition to determining conditions under which desirable outcomes can be obtained, a goal of the study is to understand how design requirements for the social and the computational aspects of systems interact, in order to optimize participation and the scientific value of the work (Caruana et al., 2006; Hochachka et al., 2007; Stevenson et al., 2003).

### 2.3 Emergent states

The second set of mediators are emergent states, the dynamic properties of the group that vary as a function of inputs and processes. Past research suggests a number of potentially relevant emergent states, including 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. At the project level, research on other kinds of virtual teams has identified the importance of factors such as trust, cohesion, conflict and morale that affect the feelings of community in the group, and thus its long term sustainability (Ahuja and Carley, 1999; Jarvenpaa and Leidner, 1999; Kasper-Fuehrera and Ashkanasy, 2001; Markus et al., 2000; Maznevski and Chudoba, 2000; Montoya-Weiss et al., 2001; Overdevest et al., 2004). In our work, we will seek to determine the role these factors may play in citizen science projects.

At the individual level, we are particularly interested in the evolution of volunteers through different roles in the group, from initial volunteer through sustained contributor, and potentially to more central roles (Crowston and Fagnot, 2008; Preece and Shneiderman, 2009). A closely related concern is volunteers' level of commitment to the project and how it influences their task performance (Byron and Curtis, 2002; Cnaan and Cascio, 1999). In previous work (Crowston and Fagnot, 2008), we described different motivational factors at play as participants move from curious initial participants to sustained contribu-

---

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

[4] `http://www.pwrc.usgs.gov/bpp/`

tors to meta-contributors, whose efforts help structure and thus ease the contributions of others. We view the decision to make an initial contribution as largely curiosity-driven ("testing the waters"), driven by project visibility and facilitated by the expected ease of joining and participating and the contributor's having available time and some level of expertise, domain interest and self-efficacy. By contrast, we expect that the decision to continue contributing derives from the contributor's feelings of commitment to the project and its goals, the intrinsic motivation of the task, and feedback from the task and other participants. Finally, we suggest that the decision to meta-contribute is driven by a sense of group membership, leading to feelings of obligation to the group, as well as by the intrinsic motivation of the task. In the proposed study, we will explore whether these stages of engagement and the theorized motivations are useful in describing the participation of contributors to citizen science projects.

In the IMOI model, processes and emergent states are conceptualized as moderating the relation between inputs and outputs of the project. At the individual level, the input elements of organizational design, task design and technology design affect motivation and participation of distributed volunteers, thus affecting the outputs (Bussell and Forbes, 2002; Ilgen and Hollenbeck, 1991; Lawrence, 2006; Sproull et al., 2005; Wasko et al., 2004). For example Bussell and Forbes (2002) describe a variety of ways in which people can volunteer and note the importance of carefully designed processes to retain volunteers, both indicating the importance of task design. At the project level, the inputs transform the means of production of scientific knowledge (Langlois and Garzarelli, 2008), shaping the demand for supporting social-computational systems (Murray and Harrison, 2002; Onsrud et al., 2004) and potentially transforming the organizational design.

## 2.4 Outputs

Finally, outputs represent task and non-task consequences of a functioning group (Martins et al., 2004) that lead to the project's effectiveness. At the individual level, task outputs for a scientific project are contributions, often raw or processed data although other contributions are possible depending on the project. Important measures of this output (and thus, the overall effectiveness of the project) include the quantity and especially the quality of the data, analysis and findings. In addition to the individual-level outputs, a citizen science project 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. For example, a new astronomical body, now called *Hanny's Voorwerp*, was discovered by a Dutch elementary school teacher volunteering with the GalaxyZoo project (Cho and Clery, 2009). Finally, at the societal level, the success of a project may affect public participation in and perception of science (Clark and Illman, 2001), create informal learning opportunities (Krasny and Bonney, 2005; Lewenstein et al., 1998), and provide the mechanisms for knowledge production at an unprecedented pace and scale (Bhattacharjee, 2005; Dhondt et al., 2005).

To further conceptualize project effectiveness, we draw on Hackman's (1987) model of group effectiveness. In addition to task completion, Hackman also considers as an output satisfaction of group members' individual needs, which includes aspects such as individual learning and personal satisfaction. These measures of effectiveness relate closely to the educational mission of many citizen science projects. Finally, Hackman includes the importance of the continued ability of the group to work together, speaking to the sustainability of the project, in both the task and the social structure of the group. In other words, a project is not effective if it achieves its goal once but drives away participants in the process.

A key point of the IMOI model is that outputs themselves become future inputs to the process. Positive personal outcomes can lead to increased motivation for future participation, and individual learning can increase a member's capability to work on additional tasks. At the project level, learning may lead to innovation in research approach, resulting in changes in the task design and thus the group processes. Positive project outputs may lead to increased interest among practitioners in applying this mode of research, as well as increased visibility for the project, helping to recruit and retain additional volunteers.

In summary, synthesizing elements of prior research on groups with contextually relevant theory provides a solid theoretical grounding for studying the organization of large numbers of virtual volunteers for scientific research and the context, structure and function of citizen science projects. This model is being applied in the VOSS grant, and we anticipate that it will evolve as it is tested in that project, allowing the current project to benefit from that learning. The refined model will then be applied to analyze the nature of the citizen science projects and suggest important factors to consider in the design of social-computational systems for these settings.

# 3 Research design

In this section, we discuss the design of the study. We address the overall research strategy, methodological integration, concepts to be examined, details of the proposed data collection and analysis techniques. Our overall plan is to use our framework to guide an initial round of exploratory research to identify the factors and relations that are most important for socially intelligent computing to support citizen science. In subsequent phases, the framework and findings from the case studies will be used to guide the design and evaluation of potentially beneficial systems to support these projects.

We envision our entire research project as having three overlapping phases. Each phase will last roughly a year, though the transition between these phases will be gradual, as the start of one phase overlaps the completion of the previous stage. The timeline for the research is shown in Figure 2, and the synthesis of methods in Figure 3. In the first phase, we will conduct in-depth case studies of citizen science projects to understand the conditions (cognitive, social or cultural) under which socially intelligent computing can enable and enhance citizen science projects. In the second phase, we will design and deploy systems to support two or three citizen science projects, which will be monitored and maintained for the duration of the project. In the third phase, we will collect user feedback from both volunteers and citizen science project managers and conduct other evaluations to assess the impact of the technologies on the projects' outcomes. These phases will be described in more detail in this section. For each phase, we describe the goals of the phase and data collection and analysis approaches as appropriate.

## 3.1 Phase One: Detailed Case Studies and Stakeholder Involvement

In the first phase, roughly year one, our focus will be development of detailed comparative case studies exploring the constructs from our conceptual framework and their relations for a small number of projects. The depth of enquiry for these case studies will necessarily restrict the number of cases so we expect to select two or three citizen science projects for this phase of research (in addition to the one or two planned as part of the VOSS grant). Theoretical sampling of cases will be guided by the taxonomy that is being developed as part of the VOSS grant, and will reflect such factors as scientific field of enquiry, integration of citizen scientists as research collaborators, project scale and scope, overall effectiveness and use of ICT. We are particularly interested in the applicability of participatory modes of research for social science and "little science" research (de Solla Price, 1963).

As a number of these variables are interdependent, we will select cases that exemplify common organizational characteristics. For example, case selection might include a small project with one PI and a few students using open source software to collect data for a single study (e.g., The Great Sunflower Project); a project run by a large research group with multiple projects that employ centralized systems to engage public participation at a national level over a period of years (e.g., eBird at the Cornell Laboratory of Ornithology); and an inter-organizational network of researchers, governmental agencies, and public sector groups with federated information systems supporting the collection of multiple types of

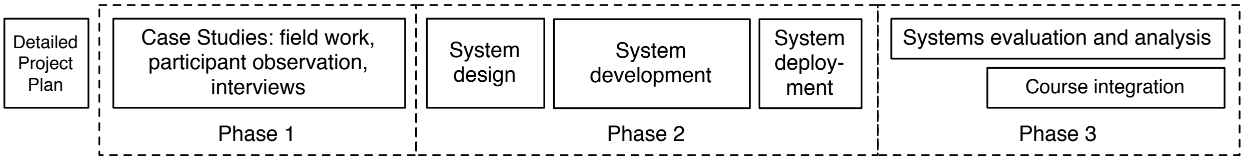

Figure 2: Timeline of proposed study activities.

environmental monitoring data for long-term ecological research and natural resource management (e.g., the Northeast Phenology Monitoring Pilot).

The case study methodology will be similar to that developed for the VOSS grant but extended to provide insight on possible technology designs. For each case, we will examine the organizational, task and current technology designs (if any) as inputs, the individual and project level processes and emergent states, and the outputs and effectiveness of the project, both for individual participants and overall. The case study protocol will be aimed at understanding the technological and social arrangements that support production, and the social and technological barriers to and enablers of participation in the selected citizen science projects. The insights drawn from these in-depth case studies will contribute to further refinements of our conceptual framework, through which we will identify specific mechanisms employed in task design to support research quality and ongoing participation, as well as current and potential forms of computational support.

We are currently conducting a pilot case study with one citizen science project which has multiple location-based implementations of a single protocol. The Northeast Phenology Monitoring Pilot collaboration[5] has agreed to serve as a case site during their 2009–2010 citizen science program pilot to develop standardized protocols and regionally coordinated monitoring. Several other potential partners have been identified and approached regarding this study; all are interested in participating. Letters from selected partners confirming this interest are included in the supplementary documentation. In developing our methodology, we will also explore the potential of applying citizen science approaches for collecting data on a broader range of projects than we can address directly.

### 3.1.1 Phase 1 Data Collection

To explore the constructs identified in the conceptual development section of this proposal (Figure 1), we will collect and analyze a range of data. Data collection specifics for case studies will vary based on the organizational characteristics, but each case will include documentation of data and volunteer management practices, research and protocol designs, participant observation and interviews with researchers, project managers, and citizen scientist volunteers. Development of our conceptual framework for citizen

---

[5] Partners include the USA National Phenology Network, the National Parks Service Inventory and Monitoring Northeast Temperate Network, Appalachian Mountain Club, Appalachian Trail Conservancy, The Wildlife Society, and several National Parks; `http://usanpn.org/?q=nps`

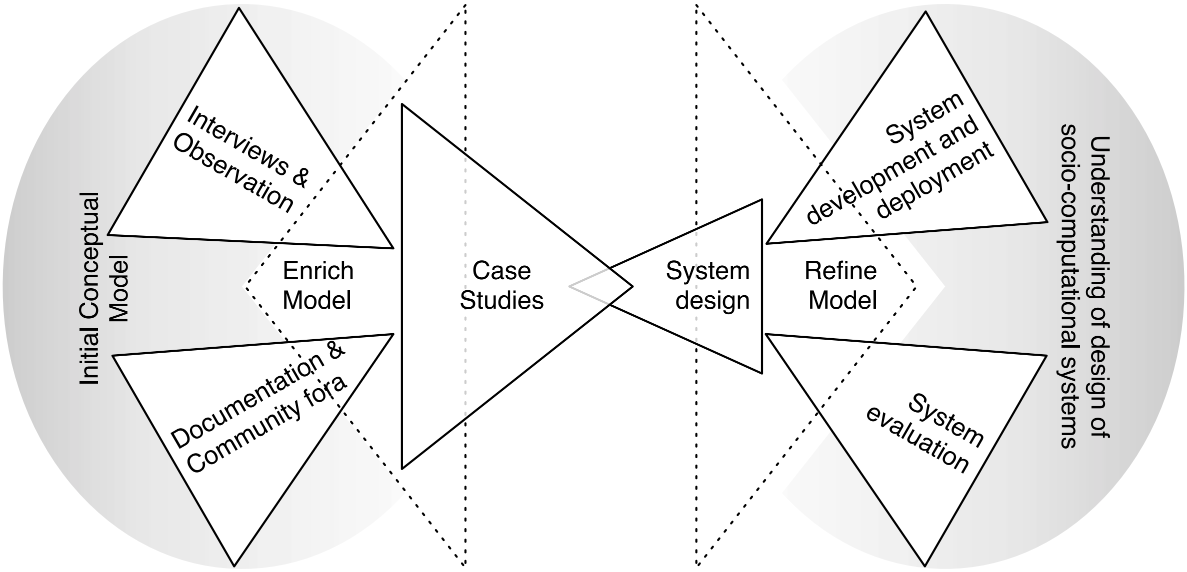

Figure 3: Planned cycle of data collection, analysis, and conceptual development.

science projects requires integration of a variety of data collected through multiple methods. To accomplish this, we are planning an iterative process of data collection and conceptual framework development process, as shown in Figure 3. Each stage of data collection will contribute to ongoing analysis and will inform the subsequent research efforts. In the remainder of this section, we will briefly review each kind of data and our plans for data collection.

*Interviews and observation.* Interviews will be conducted face-to-face when feasible (e.g., for local sites and during observational periods) and by phone in other cases. Initial interviews will be conducted with project staff and volunteers to develop a deeper understanding of the structural relationships affecting research and participation outcomes and to uncover issues and challenges faced by researchers managing citizen science projects. Where possible, we will also observe the functioning of the projects to explore the experiences and perspectives of both the research group members and the citizen science volunteers. Funding has been requested to support travel to the project sites to enable observation of the core research group at work. For all case studies, the researchers will engage in participant observation as a citizen science contributor. The independent and distributed nature of participation in many citizen science projects is particularly convenient for participant observation. Observation and interviews will be employed to document task design, focusing in particular on volunteer and data management activities and research planning. This approach will provide insight into organizational structure and context through participation in such activities as research group meetings and volunteer training events.

*Project documentation.* We will collect several forms of documentation for each case study, including documents created for citizen science volunteers and those created for organizational use. The documents will include volunteer recruitment and educational materials, research design and protocol documents, instructions for volunteers, observation reporting forms and web interfaces and published findings. Additional project documentation may include meeting minutes or other project management materials (e.g., organizational charts, planning calendars, etc.) as well as data management policies and procedures. Public-facing and internal documents are evidence of volunteer management and data management practices and provide a basis for evaluating the work structure of citizen science research projects.

*Online community fora.* In some cases, projects communicate with volunteer participants via online discussion board or email listservs. The contents of these fora provide rich data from participants relevant to the problems they encounter, the insights they develop, the kinds of resources and contributions they can make, the volunteer roles that evolve and the way they interpret and carry out research protocols. Online communities' data also provide evidence of volunteer management practices in citizen science projects, allowing us to consider the role of organizational design for supporting large-scale volunteer contribution via online community management. This type of mechanism to support participation may be internally managed by project staff or may be delegated to a reliable group of "super volunteers" with proven expertise, as in many other online community contexts (Powazek, 2001). The choice of these or other solutions to address citizen science project community management is likely to be a function of organization and task design.

### 3.1.2 Phase 1 Analysis

In the first phase of the analysis, we will build on the results from our prior work and our in-process detailed case studies to refine our conceptual framework. In particular, we expect to further investigate the following concepts.

*Organizational design.* As we develop a more complete understanding of the variations in organizational design characteristics for citizen science projects, our investigation will focus on identifying the antecedents to these organizational design choices. This analytical effort will contribute to developing generalized use case scenarios to document the conditions under which citizen science projects provide a good solution to meeting research goals. A particular concern is the extent to which projects exhibit properties of communities in their organizational structure.

*Task design.* We expect that task design strategies will vary based on organizational design. Our analysis will extend the use case scenarios for organizational design by matching the organizational structures from our taxonomy development to the most common task design solutions. A particular concern is how participants are motivated to participate in the projects. This analysis intends to provide a set of heuristics for researchers planning, managing and evaluating citizen science projects to employ as they design and redesign research projects for this mode of engagement.

*Participant roles.* We plan to document citizen science project roles using several approaches. First, we will look for evidence of explicit formal roles for staff and volunteers during our taxonomy development. Second, we will discuss project organizational structure as well as formal and informal roles of volunteers with project managers during interviews. Finally, we will look for behavioural evidence of informal roles in online communities for citizen science projects. This analysis of informal and structural roles should provide a useful tool for task design within citizen science projects, as the ability to effectively leverage the skills and interests of a diverse pool of contributors can make an enormous impact on the scalability of the projects, ongoing participation, and the quality of work.

*Technology in use.* Finally, we will document the current state of technology support and the role these systems play in the project. We will consider in particular how the functions of the project are divided between the human and systems actors and the nature of the affordances provided by the technology. An interesting aspect of this analysis is the value systems embedded in the systems. Most citizen science projects are heavily dependent on the contributions of volunteers, making volunteer motivation particularly important to project success.

### 3.2 Phase Two: System Development

For the second phase, roughly year two, we will test our understanding developed from the case studies through an action research and design science element, in which undergraduate and master students enrolled in existing information management courses will develop and test prototypes of systems to support selected citizen science projects, using design insights from our case studies and based on our conceptual framework. Based on the model, we argue that to be useful, systems for citizen science projects need to be designed taking into account the organizational structure of the projects and the task design and as well, the demographics, skills and motivations of the participants (the inputs to our model). Systems should help support a variety of project processes, such as volunteer management, in addition to the core research and data management tasks and the individual level functions of joining and contributing. Evaluation of systems should examine both individual and project level outcomes, and go beyond a focus on research outputs to include individual advancements and community sustainability.

For example, the Northeast Phenology Monitoring Pilot Study plans to collect phenology observations (i.e., times of cyclic natural phenomena) from hikers on the Appalachian National Scenic Trail (AT), using paper forms submitted at drop-boxes along the trail where electronic communications are extremely limited. At other locations, such as the Boston Harbor Islands National Recreation Area, small groups of volunteers will collect data in the company of park staff. Though the challenging maritime environment involves exceptional logistical constraints and field conditions would make any researcher hesitate to bring a laptop to the site, the wireless communication infrastructure is excellent and volunteers are eager to experience an otherwise inaccessible ecosystem. At Marsh-Billings-Rockefeller, phenology observations will be collected by school groups for which data entry could be integrated into the curricular materials that are being developed. All of these variations in implementing the partnership's standardized protocol are the direct result of contextual factors and resource availability, which made paper data collection the most practical initial step for testing the monitoring protocol.

A relatively simple information system for this project could involve a web-based reporting site, which would need adequate flexibility to support collection and integration of the same data by different types of participants in very different organizational and physical contexts. For example, data might be entered by observers, researchers, park staff, other park volunteers or schoolchildren. As well, even a simple system for this kind of project will need to include ways to check the reliability of contributed data, particularly given the potential combination of expert, trained and inexpert users. Cornell's eBird system, for example, has bounds by time, location and species for exceptional reports that require additional verification, though that verification is done manually. A more ambitious project might use the camera and GPS capabilities of smart phones to upload geotagged pictures of events or observations to a system that enables volunteer analysis, e.g., of species and stage of development, similar to what is done for the GalaxyZoo project. This sort of technology has been specifically mentioned as a desirable social-computational system by project leaders, but it remains to be seen whether volunteer participants would make use of the technology and whether it could supplant other modes of participation.

The particular goal of this effort is to develop generic frameworks for systems to support various types of citizen science projects by identifying generalizable tool kits that can be used in a variety of projects, rather than point solutions for a single project. In this respect, we hope to parallel developments in other kinds of massive voluntary collaborations. For example, in FLOSS development, tools such as version control systems and bug trackers have been found to be useful for nearly all projects. Similarly, wikis have emerged as a generic tool to support many instances of distributed text development. We hope to design and develop generally useful technologies to support broad classes of citizen science projects. The characterization of the classes of projects and their needs will lead to description of new models of social computing that support large-scale public participation in scientific research.

*System design and development.* Systems to enhance citizen science research quality and participation will be generated as deliverables for project-based information management courses at the Syracuse University iSchool and possibly with other partners. Organizations that serve as project sites for these courses will receive the completed design recommendations and have opportunity to provide feedback on prototypes. We anticipate building systems by integrating already existing systems with a small amount of custom-developed code. For example, a system might be built on a content management system like Drupal that supports authentication of users, editing and dissemination of news articles, etc. On this foundation, a custom module might be developed to support the specific kind of data collection and validation needed for a class of citizen science projects. Geographical data might involve a Google Maps mashup as well as processing of GPS data.

In addition to the benefits to the citizen science projects, this system design process will provide a practice-based formal learning opportunity for undergraduate and masters students. Planning of curricular integration will begin during the first year, in partnership with the course instructors, with system design and development to be done in the Fall of the second year. Courses targeted for involvement include Systems Analysis (taught by the PI), Science Data Management, Human-Computer Interaction and Website Design. Students in these courses regularly design and implement small scale systems for real customers and faculty teaching them are experienced in managing this kind of development, making us confident of being able to create a working system, based on the planned research to establish requirements for systems that can generalize across projects. We can also draw on a program currently offered by the iSchool that employs masters students to work on development projects. Finally, we will explore involvement of students enrolled in the iSchool's Cyberinfrastructure Facilitators, eScience Librarianship, and Digital Libraries certificate programs.

A competition will be held to select a few (two or three) of the most promising designs for deployment in an active citizen science project towards the end of the second year. Implemented designs will be instrumented for evaluation; for example, web site usage data would be retained and analyzed for browser-based systems, and the evaluation of the design will be instrumented to enable comparison to prior technologies wherever possible. Our project plan includes training the research groups to maintain the technology to ensure sustainability. Rather than relying on students in courses to provide training and initial support, we plan to hire a small number of students for these roles. The seasonal nature of data collection in most citizen science projects allows for a natural cycle of implementation and evaluation.

### 3.3 Phase Three: System Evaluation

Finally, in phase 3, roughly year 3, we will assess the impact of the deployed systems on the processes, emergent states and outputs identified in our research model (with updates from our findings in the case studies). The goal of this assessment is to examine their influence on relevant project outputs, e.g., the quantity and quality of data collected and participation. Appropriate measures of participation will vary by design solution but might include greater frequencies of participation and increased levels of repeat contribution. Our analysis will also examine the evidence for informal learning benefits to volunteers, as well as the formal learning benefits of students involved in the system development. Finally, we will examine mediating variables, such as the dimensions of volunteer motivation identified in our framework. This aspect of our research will examine organizational impact of systems and the ways it transforms the work of scientific research. Common mechanisms for adjusting organizational structure to meet the changing demands of conducting citizen science research will be documented, as will best practices for the citizen science project types. Data for the evaluation will be obtained from observation of usage (supported by appropriate instrumentation of the tools, taking into account possible privacy concerns of participants), surveys of and interviews with participants and observation of the project practices. Numerical data will be analyzed statistically (e.g., comparing the number and quality of contributions before and after systems implementation); other data will be analyzed qualitatively (e.g., participants' impressions of systems functions).

## 4 Management plan

Based on preliminary assessment of the effort required, we are requesting funding for one graduate student and summer support for the PI (1 summer month). The PI will work during the summer on project management and research design, and supervise the graduate student during the academic year. He will take particular responsibility for case selection, overall project design and report writing. The graduate student will support the PI in case selection, theory development, and definition of constructs and variables. She will have primary responsibility for data collection and analysis, under the oversight of the PI. The budget includes travel support for case study interviews and observation of research group meetings. Volunteer participant observation work will be done independently, in the citizen scientist mode. We plan to involve colleagues teaching project-based courses to oversee the system development, though the PI and student will also be involved. In addition, we have budgeted for 8 students (4 masters and 4 undergraduate) working 10 hours/week to assist in technology deployment and project support. We anticipate deploying pairs of students (a masters student and undergraduate) to support each citizen science project, with other employed making any needed fixes or enhancements to the systems and doing system maintenance. During Phase III, students will also be employed in collecting and analyzing usage data from the projects to support the evaluation. An initial project activity will be the development of a more detailed timeline against which progress will be measured. The budget includes support for these activities.

## 5 Conclusion

Through the three phases of the study, described above, we will develop a better understanding of the growing phenomenon of citizen science projects, some of which now involve tens of thousands of volunteers in scientific research. We posed the following research goals:

1. developing a practical understanding of the conditions under which socially intelligent computing can enable and enhance scientific and education production and innovation in citizen science projects;
2. developing new models of social computing; and
3. developing and testing systems reflecting explicit knowledge about people's cognitive and social abilities for this setting.

The proposed study will address these goals using theory from group research and data from detailed case studies, and system development and evaluation. More specifically, we will carry out detailed case studies of a few citizen science projects and pilot test computational technologies to support effective citizen science projects, based on contextual designs by undergraduate and masters students. The proposed research will thus examine the spectrum of ways that citizens can meaningfully contribute to scientific projects while maintaining the integrity of the research and will advance our understanding of the factors that affect the effectiveness of such projects.

The project will benefit society by investigating how involving the public in scientific research through socially intelligent computing systems can advance science directly, in addition to goals of outreach or informal learning; generating and disseminating insights directly applicable to improving the design and implementation of social-computational systems for citizen science projects, thereby improving the available technologies through which the public engages in scientific research; determining the conditions under which citizen science projects provide a solution for large-scale data collection, as well as opportunities to leverage public interest in other aspects of scientific knowledge production; and creating empirically-based theoretical models of social computing in large-scale contribution systems, which are increasingly used to create public goods. As well, the project will contribute to the education of doctoral, masters and undergraduate students who will learn about research and system development through their participation in the proposed project.

To ensure that our study has a significant impact, we plan to broadly disseminate results through journal and conferences publications, and on our Web pages. We also plan to disseminate results directly to, and invite contributions from, interested practitioners. Our collaborative relationships, particularly with the CLO and USA National Phenology Network, will provide further avenues for dissemination to the broader practitioner community. Courses involving system design will also be incorporated into the curricula of the Syracuse University School of Information Studies. The design and development of prototypes in year 2 of the proposed project will provide class projects for these students. Finally, the project will promote teaching, training and learning by students involved in the research project, providing the opportunity to develop skills in data collection and analysis. In addition, curricular integration with existing information management coursework will provide undergraduates and masters students with hands-on experience in systems development.

## 5.1 Results from prior funding

The PI for this grant, Crowston, has been funded by several NSF grants within the past 48 months. The most relevant grant, newly awarded, is 09–43049, “VOSS: Theory and design of virtual organizations for citizen science” ($150,000 for 2 years). This grant will support development of a taxonomy of citizen science projects, some initial case studies of one or two citizen science projects and a workshop of scientists leading citizen science projects to discuss best practices. The current proposal is designed to complement and build on this project by deepening the case studies and employing the taxonomy to guide development and testing of socially intelligent computing systems for public participation in scientific research.

Other relevant grants HSD 05–27457 ($684,882, 2005–2008, with R. Heckman, E. Liddy and N. McCracken), *Investigating the Dynamics of Free/Libre Open Source Software Development Teams*, IIS 04–14468 ($327,026, 2004–2006) and SGER IIS 03–41475 ($12,052, 2003–2004), both entitled *Effective work practices for Open Source Software development*. These grants have supported a study of the evolution of effective work practices for virtual groups, specifically, for FLOSS projects. Findings from these grants included a taxonomy of success measures for FLOSS projects (Crowston et al., 2006a) and implementation of novel measures of success (Wiggins et al., 2009), evidence about the structure of projects (Crowston and Howison, 2005, 2006a) and descriptions of key practices, e.g., for decision making (Heckman et al., 2007a; Li et al., 2008), leadership (Heckman et al., 2007b), motivation (Crowston and Fagnot, 2008) and group maintenance (Scialdone et al., 2008, 2009). Overall, this work has resulted in eight journal papers (including *IEEE Software* (Crowston and Howison, 2006b), *Software Process—Improvement and Practice* (Crowston et al., 2006a), *IEEE Transactions on Professional Communications* (Crowston et al., 2007a) and *IEE Proceedings Software* (Crowston and Scozzi, 2002), among others (Crowston and Howison, 2005, 2006a; Crowston and Scozzi, 2008; Crowston et al., 2007b; Howison et al., 2006a); a book chapter (Crowston, 2008); and multiple conference papers (Annabi et al., 2006; Crowston and Fagnot, 2008; Crowston et al., 2003, 2005a,b,c,d, 2006b, 2008; Heckman et al., 2006, 2007a,b; Howison and Crowston, 2004; Howison et al., 2006b, 2008; Li et al., 2006, 2008; Scialdone et al., 2008, 2009; Scozzi et al., 2008; Wiggins et al., 2008a,b, 2009). These grants have supported a total of six PhD students; several others have been involved in specific aspects of the projects.

# References

B. Thomas Adler, Luca de Alfaro, Ian Pye, and Vishwanath Raman. Measuring author contributions to the Wikipedia. In *Proceedings of WikiSym*, Porto, Portugal, 8–10 September 2008.

Manju K. Ahuja and Kathleen M. Carley. Network structure in virtual organizations. *Organization Science*, 10(6):741–757, 1999.

Hala Annabi, Kevin Crowston, and Robert Heckman. Group learning in the early years of Apache web server. In *Proceedings of the IFIP 2nd International Conference on Open Source Software*, pages 77–90, Lake Como, Italy, 8–9 June 2006.

ATC Citizen Science Working Group. Citizen science: Overview of the basics. Technical report, September 2006. URL `http://sain.utk.edu/apptrail_documents/Citizen_Science_Report_20070629.pdf`.

Heidi L. Ballard, David Pilz, Eric T. Jones, and Christy Getz. *Training Curriculum for Scientists and Managers: Broadening Participation in Biological Monitoring*. Institute for Culture and Ecology, Corvalis, OR, 2005.

Claudia Baretto, David E. Fastovsky, and Peter M. Sheehan. A model for integrating the public into scientific research. *Journal of Geoscience Education*, 50(1):71–75, 2003.

Jeremy Hugh Baron. What should the citizen know about science? *Journal of the Royal Society of Medicine*, 97(1):50, JAN 2004.

Martin W. Bauer, Kristina Petkova, and Pepka Boyadjieva. Public knowledge of and attitudes to science: Alternative measures that may end the “science war”. *Science Technology Human Values*, 25(1):30–51, 2000.

Yudhijit Bhattacharjee. Ornithology: Citizen scientists supplement work of Cornell researchers. *Science*, 308(5727):1402–1403, 3 June 2005.

Cornelia Boldyreff, Kevin Crowston, Björn Lundell, and Tony Wasserman, editors. *Proceedings of The Fifth International Conference on Open Source Systems*. Springer, Skövde, Sweden, 3–6 June 2009.

Rick Bonney and Melinda S. LaBranche. Citizen science: Involving the public in research. *ASTC Dimensions*, page 13, May/June 2004.

Rick Bonney and Jennifer L. Shirk. Citizen science central. *Connect*, pages 8–10, March 2007.

Carol Brewer. Outreach and partnership programs for conservation education where endangered species conservation and research occur. *Conservation Biology*, 16(1):4–6, February 2002.

Dominique Brossard, Bruce Lewenstein, and Rick Bonney. Scientific knowledge and attitude change: The impact of a citizen science project. *International Journal of Science Education*, 27(9):1099–1121, 2005.

Helen Bussell and Deborah Forbes. Understanding the volunteer market: The what, where, who and why of volunteering. *International Journal of Nonprofit and Voluntary Sector Marketing*, 7(3), 2002.

Ian Byron and Allan Curtis. Maintaining volunteer commitment to local watershed initiatives. *Environmental management*, 30(1):59–67, 2002.

Rich Caruana, Mohamed Elhawary, Art Munson, Mirek Riedewald, Daria Sorokina, Daniel Fink, Wesley M. Hochachka, and Steve Kelling. Mining citizen science data to predict prevalence of wild bird species. In *Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 909–915. ACM New York, NY, USA, 2006.

Marcelo Cataldo, Patrick A. Wagstrom, James D. Herbsleb, and Kathleen M. Carley. Identification of coordination requirements: Implications for the design of collaboration and awareness tools. In *Proceedings of the 20th anniversary conference on Computer-supported cooperative work (CSCW '06)*, pages 353–362, Banff, Alberta, Canada, 2006. ACM.

Adrian Cho and Daniel Clery. International year of astronomy: Astronomy hits the big time. *Science*, 323(5912):332, 2009.

J. Christie. Volunteer attitudes and motivations: Research findings and their application for sustainable community involvement programs in natural resource management. In *Effective Sustainability Education: What Works? Why? Where Next? Linking Research and Practice*, Sydney, Australia, 18–20 February 2004. URL `http://www.environment.nsw.gov.au/resources/cee/christie.pdf`.

Fiona Clark and Deborah L. Illman. Dimensions of civic science: Introductory essay. *Science communication*, 23(1):5, 2001.

Ram A. Cnaan and Toni A. Cascio. Performance and commitment: Issues in management of volunteers in human service organizations. *Journal of Social Service Research*, 24:1–38, 1999.

Jeffrey P. Cohn. Citizen science: Can volunteers do real research? *BioScience*, 58(3):192–107, 2008.

Caren B. Cooper, Janis Dickinson, Tina Phillips, and Rick Bonney. Citizen science as a tool for conservation in residential ecosystems. *Ecology and Society*, 12(2), 2007.

Kevin Crowston. The bug fixing process in proprietary and free/libre open source software: A coordination theory analysis. In Varun Grover and M. Lynne Markus, editors, *Business Process Transformation*. M. E. Sharpe, Armonk, NY, 2008.

Kevin Crowston and Isabelle Fagnot. The motivational arc of massive virtual collaboration. In *Proceedings of the IFIP WG 9.5 Working Conference on Virtuality and Society: Massive Virtual Communities*, Lüneberg, Germany, 1–2 July 2008.

Kevin Crowston and James Howison. The social structure of free and open source software development. *First Monday*, 10(2), 2005.

Kevin Crowston and James Howison. Hierarchy and centralization in free and open source software team communications. *Knowledge, Technology & Policy*, 18(4):65–85, 2006a.

Kevin Crowston and James Howison. Assessing the health of open source communities. *IEEE Computer*, 39(5):89–91, May 2006b.

Kevin Crowston and Barbara Scozzi. Open source software projects as virtual organizations: Competency rallying for software development. *IEE Proceedings Software*, 149(1):3–17, 2002.

Kevin Crowston and Barbara Scozzi. Coordination practices within free/libre open source software development teams: The bug fixing process. *Journal of Database Management*, 19(2):1–30, 2008.

Kevin Crowston, Hala Annabi, and James Howison. Defining open source software project success. In *Proceedings of the 24th International Conference on Information Systems*, Seattle, WA, 2003.

Kevin Crowston, Hala Annabi, James Howison, and Chengetai Masango. 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, 2005a.

Kevin Crowston, Robert Heckman, Hala Annabi, and Chengetai Masango. A structurational perspective on leadership in free/libre open source software teams. In *Proceedings of the 1st Conference on Open Source Systems (OSS)*, Genova, Italy, 2005b.

Kevin Crowston, James Howison, Chengetai Masango, and U. Yeliz Eseryel. Face-to-face interactions in self-organizing distributed teams. Paper presented at the *Academy of Management Conference*, Honolulu, HI, 2005c.

Kevin Crowston, Kangning Wei, Qing Li, U. Yeliz Eseryel, and James Howison. Coordination of free/libre open source software development. In *Proceedings of the International Conference on Information Systems*, Las Vegas, NV, USA, 2005d.

Kevin Crowston, James Howison, and Hala Annabi. Information systems success in free and open source software development: Theory and measures. *Software Process–Improvement and Practice*, 11(2): 123–148, 2006a.

Kevin Crowston, Kangning Wei, Qing Li, and James Howison. Core and periphery in free/libre and open source software team communications. In *Proceedings of the 39th Hawai'i International Conference on System System*, Kaua'i, Hawai'i, 2006b.

Kevin Crowston, James Howison, Chengetai Masango, and U. Yeliz Eseryel. The role of face-to-face meetings in technology-supported self-organizing distributed teams. *IEEE Transactions on Professional Communications*, 50(3):185–203, 2007a.

Kevin Crowston, Kangning Wei, Qing Li, U. Yeliz Eseryel, and James Howison. Self-organization of teams in free/libre open source software development. *Information and Software Technology Journal, Special issue on Understanding the Social Side of Software Engineering, Qualitative Software Engineering Research*, 49:564–575, 2007b.

Kevin Crowston, James Howison, and Andrea Wiggins. Opportunities for eScience research on free/libre open source software. In *Proceedings of the Oxford e-Research Conference*, Oxford, England, 11-13 September 2008.

Kevin Crowston, Kangning Wei, James Howison, and Andrea Wiggins. Free/libre open source software: What we know and what we do not know. *ACM Computing Surveys*, Under review.

Shane R. de Solla, Kim J. Fernie, Glenn C. Barrett, and Christine A. Bishop. Population trends and calling phenology of anuran populations surveyed in ontario estimated using acoustic surveys. *Biodiversity and Conservation*, 2005.

Derek J. de Solla Price. *Little Science, Big Science*. Columbia University, New York, 1963.

André A. Dhondt, Sonia Altizer, Evan G. Cooch, Andrew K. Davis, Andrew Dobson, Melanie J. L. Driscoll, Barry K. Hartup, Dana M. Hawley, , Wesley M. Hochachka, Parviez R. Hosseini, Christopher S. Jennelle, George V. Kollias, David H. Ley, Elliott C. H. Swarthout, and Keila V. Sydenstricker. Dynamics of a novel pathogen in an avian host: Mycoplasmal conjunctivitis in house finches. *Acta Tropica*, 94(1):77–93, Apr 2005.

Caroline Dufour and Elizabeth Crisfield, editors. *The Appalachian Trail MEGA-Transect*, Harpers Ferry, WV, 2008. Appalachian Trail Conservancy.

Steven M. Farmer and Donald B. Fedor. Volunteer participation and withdrawal. *Nonprofit Management and Leadership*, 9(4), 1999.

Thomas A. Finholt. Collaboratories. In B. Cronin, editor, *Annual Review of Information Science and Technology*, volume 36. Information Today, Medford, NJ, 2001.

Thomas A. Finholt and Gary M. Olson. From laboratories to collaboratories: A new organizational form for scientific collaboration. *Psychological Science*, 8(1):28–35, 1997.

Karen Firehock and Jay West. A brief history of volunteer biological water monitoring using macroinvertebrates. *Journal of the North American Benthological Society*, 14(1):197–202, 1995.

Leska S. Fore, Kit Paulsen, and Kate O'Laughlin. Assessing the performance of volunteers in monitoring streams. *Freshwater Biology*, 46:109–123, 2001.

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

C. Wayne Gordon and Nicholas Babchuk. A typology of voluntary associations. *American Sociological Review*, 24(1):22–29, 1959.

Virginia C. Gulick, Glenn Deardorff, Bob Kanefsky, and Alexandra Davatzes. Student and public participation in acquiring and analyzing HiRISE images. In *Proceedings of the Lunar and Planetary Science XXXVIII*, March 2007. URL `http://adsabs.harvard.edu/abs/2007LPI....38.2248G`.

Richard A. Guzzo and Marcus W. Dickson. Teams in organizations: Recent research on performance and effectiveness. *Annual Review of Psychology*, 47:307–338, 1996.

J. Richard Hackman. The design of work teams. In Jay W. Lorsch, editor, *The Handbook of Organizational Behavior*, pages 315–342. Prentice-Hall, Englewood Cliffs, NJ, 1987.

J. Richard Hackman and Charles G. Morris. Group tasks, group interaction process, and group performance effectiveness: A review and proposed integration. In L. Berkowitz, editor, *Group Processes*, volume 8 of *Advances in Experimental Social Psychology*, pages 45–99. Academic Press, New York, 1978.

Jacqueline S. Hange, Brenda S. Seevers, and Dawn VanLeeuwen. 4-H youth development extension agents' attitudes towards volunteer management competencies. In *Proceedings of the 29th National Agricultural Education Research Conference*, Las Vegas, NV, USA, 11 December 2002. URL `http://aaae.okstate.edu/proceedings/2002/NAERC/4-H%20Hange-Seevers-VanLeeuwen.pdf`.

Robert Heckman, Kevin Crowston, Qing Li, Eileen E. Allen, Yeliz Eseryel, James Howison, and Kangning Wei. Emergent decision-making practices in technology-supported self-organizing distributed teams. In *Proceedings of the International Conference on Information Systems (ICIS)*, Milwaukee, WI, 10–13 Dec, 2006.

Robert Heckman, Kevin Crowston, U. Yeliz Eseryel, James Howison, Eileen Allen, and Qing Li. Emergent decision-making practices in free/libre open source software (FLOSS) development teams. In *Proceedings of the 3rd International Conference on Open Source Software*, Limerick, Ireland, 2007a.

Robert Heckman, Kevin Crowston, and Nora Misiolek. A structurational perspective on leadership in virtual teams. In Kevin Crowston and Sandra Seiber, editors, *Proceedings of the IFIP Working Group 8.2/9.5 Working Conference on Virtuality and Virtualization*, pages 151–168, Portland, OR, 2007b. Springer.

Wesley M. Hochachka, Rich Caruana, Daniel Fink, Art Munson, Mirek Riedewald, Daria Sorokina, and Steve Kelling. Data-mining discovery of pattern and process in ecological systems. *Journal of Wildlife Management*, 71(7):2427–2437, 2007.

James Howison and Kevin Crowston. The perils and pitfalls of mining SourceForge. Paper presented at the *Workshop on Mining Software Repositories, 26th International Conference on Software Engineering*, Edinburgh, Scotland, 2004.

James Howison, Megan Conklin, and Kevin Crowston. FLOSSmole: A collaborative repository for FLOSS research data and analyses. *International Journal of Information Technology and Web Engineering*, 1(3):17–26, 2006a.

James Howison, Keisuke Inoue, and Kevin Crowston. Social dynamics of FLOSS team communications. In *Proceedings of the Second International Conference on Open Source Systems*, Como, Italy, 2006b.

James Howison, Andrea Wiggins, and Kevin Crowston. eResearch workflows for studying free and open source software development. In *Proceedings of the Fourth International Conference on Open Source Software (IFIP 2.13)*, Milan, Italy, 7-10 September 2008.

Daniel R. Ilgen and John R. Hollenbeck. The structure of work: Job design and roles. *Handbook of Industrial and Organizational Psychology*, 2:165–207, 1991.

Daniel R. Ilgen, John R. Hollenbeck, Michael Johnson, and Dustin Jundt. Teams in organizations: From input-process-output models to IMOI models. *Annual Review of Psychology*, 56:517–543, 2005.

Sirkka L. Jarvenpaa and Dorothy E. Leidner. Communication and trust in global virtual teams. *Organization Science*, 10(6):791–815, 1999.

Eva C. Kasper-Fuehrera and Neil M. Ashkanasy. Communicating trustworthiness and building trust in interorganizational virtual organizations. *Journal of Management*, 27(3):235, 2001.

Marianne Krasny and Rick Bonney. Environmental education through citizen science and participatory action research. In Edward Arnold Johnson and Michael Mappin, editors, *Environmental Education and Advocacy*, chapter 13. Cambridge University Press, New York, 2005.

Richard N. Langlois and Giampaolo Garzarelli. Of hackers and hairdressers: Modularity and the organizational economics of open-source collaboration. *Industry & Innovation*, 15(2):125–143, 2008.

Anna Lawrence. 'No personal motive?' Volunteers, biodiversity, and the false dichotomies of participation. *Ethics, Place & Environment: A Journal of Philosophy & Geography*, 9(3):279–298, 2006.

Charlotte P. Lee, Paul Dourish, and Gloria Mark. The human infrastructure of cyberinfrastructure. In *Proceedings of the Conference on Computer-supported Cooperative Work (CSCW)*, pages 483–492, New York, NY, USA, 2006. ACM.

Bruce Lewenstein, Rick Bonney, and Dominique Brossard. Citizen science: Measuring scientific knowledge and attitudes in a new type of public outreach project. Technical Report #98-1, Texas A&M University, Center for Science & Technology Policy & Ethics, 1998.

Qing Li, Kevin Crowston, Robert Heckman, and James Howison. Language and power in self-organizing distributed teams. Paper presented at the *OCIS Division, Academy of Management Conference*, Atlanta, GA, 2006.

Qing Li, Robert Heckman, Kevin Crowston, James Howison, Eileen E. Allen, and U. Yeliz Eseryel. Decision-making paths in technology-supported self-organizing distributed teams. In *Proceedings of the International Conference on Information Systems*, Paris, France, 14-17 December 2008.

Mechelle A. Marks, John E. Mathieu, and Stephen J. Zaccaro. A temporally based framework and taxonomy of team processes. *Academy of Management Review*, 26(3):356–376, 2001.

M. Lynne. Markus, Brook Manville, and Carole E. Agres. What makes a virtual organization work? *Sloan Management Review*, 42(1):13–26, 2000.

Luis L. Martins, Lucy L. Gilson, and M. Travis Maynard. Virtual teams: What do we know and where do we go from here? *Journal of Management*, 30(6):805–835, 2004.

Martha L. Maznevski and Katherine M. Chudoba. Bridging space over time: Global virtual team dynamics and effectiveness. *Organization Science*, 11(5):473–492, 2000.

Rachel E. McCaffrey. Using citizen science in urban bird studies. *Urban Habitats*, 3(1):70–86, December 2005 2005.

Joseph E. McGrath and Andrea B. Hollingshead. *Groups Interacting With Technology*. Sage, Thousand Oaks, CA, 1994.

Mitzi M. Montoya-Weiss, Anne P. Massey, and Michael Song. Getting it together: Temporal coordination and conflict management in global virtual teams. *Academy of Management Journal*, 44(6):1251–1262, 2001.

Vic Murray and Yvonne Harrison. The impact of information and communications technology on volunteer management. Technical report, Canadian Centre for Philanthropy, University of Victoria, Victoria, Canada, 2002. URL `http://nonprofitscan.imaginecanada.ca/files/en/iyv/Murray_SR1_English_Web.pdf`.

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

Harlan Onsrud and James Campbell. Big opportunities in access to "Small Science" data. *Data Science Journal*, 6, 2007.

Harlan Onsrud, Gilberto Camara, James Campbell, and Narnindi Sharad Chakravarthy. Public commons of geographic data: Research and development challenges. In *III International Conference on Geographical Information Science (GIScience)*, pages 223–238, Washington, DC, 2004. Springer.

Dawn Alexandra Osborn, John S. Pearse, and Christy A. Roe. Monitoring rocky intertidal shorelines: A role for the public in resource management. In Orville T. Magoon, Hugh Converse, Brian Baird, Beth Jines, and Melissa Miller-Henson, editors, *California and the World Ocean: Revisiting and Revising California's Ocean Agenda*, pages 624–636, Reston, VA, 27–30 October 2002. ASCE.

Christine Overdevest, Cailin Huyck Orr, and Kristine Stepenuck. Volunteer stream monitoring and local participation in natural resource issues. *Research in Human Ecology*, 11(2):177–185, 2004.

David Pilz, Heidi L. Ballard, and Eric T. Jones. *Broadening Participation in Biological Monitoring: Guidelines for Scientists and Managers*. Institute for Culture and Ecology, Corvalis, OR, 2005.

Derek M. Powazek. *Design for community: The art of connecting real people in virtual places*. New Riders Publishing, Indianapolis, IN, 2001.

Jennifer Preece and Ben Shneiderman. The reader-to-leader framework: Motivating technology-mediated social participation. *AIS Transactions on Human-Computer Interaction*, 1(1), 2009.

Michelle Prysby and Paul E. Super. Programming: Citizen science. The director's guide to best practices, Association of Nature Center Administrators, 2006.

Michael J. Scialdone, Na (Lina) Li, James Howison, Robert Heckman, and Kevin Crowston. Group maintenance in technology-supported distributed teams. In *Best Paper Proceedings, Academy of Management Annual Meeting*, Anaheim, CA, 8–13 August 2008.

Michael J. Scialdone, Robert Heckman, and Kevin Crowston. Group maintenance behaviours of core and peripheral members of free/libre open source software teams. In Boldyreff et al. (2009).

Barbara Scozzi, Kevin Crowston, U. Yeliz Eseryel, and Qing Li. Shared mental models among open source software developers. In *Proceedings of the 41st Hawai'i International Conference on System Sciences*, Big Island, Hawai'i, 2008. URL `http://doi.ieeecomputersociety.org/10.1109/HICSS.2008.391`.

Seattle Aquarium. "Citizen science" science report. Technical report, Seattle Aquarium, 2005.

Marie T. Sergent and William E. Sedlacek. Volunteer motivations across student organizations: A test of person-environment fit theory. *Journal of College Student Development*, 31(3):255–261, 1990.

Lee Sproull, Caryn A. Conley, and Jae Yun Moon. Prosocial behavior on the net. In *The Social Net: Understanding Human Behavior in Cyberspace*, chapter 6, pages 139–161. Oxford University Press, 2005.

R. D. Stevenson, William A. Haber, and Robert A. Morris. Electronic field guides and user communities in the eco-informatics revolution. *Conservation Ecology*, 7(1):3, 2003.

Deborah J. Trumbull, Rick Bonney, Derek Bascom, and Anna Cabral. Thinking scientifically during participation in a citizen-science project. *Science Education*, 84(2):265–275, 2000.

Georg von Krogh, Sebastian Spaeth, and Karim R. Lakhani. Community, joining, and specialization in open source software innovation: A case study. *Research Policy*, 32(7):1217–1241, 2003.

Molly McLure Wasko, Samer Faraj, and Robin Teigland. Collective action and knowledge contribution in electronic networks of practice. *Journal of the Association for Information Systems*, 5(11-12):493–513, 2004.

Andrea Wiggins, James Howison, and Kevin Crowston. Replication of FLOSS research as eResearch. In *Proceedings of the Oxford e-Research Conference*, Oxford, England, 11–13 September 2008a. URL `http://www.oii.ox.ac.uk/microsites/eresearch08/papers.cfm`.

Andrea Wiggins, James Howison, and Kevin Crowston. Social dynamics of FLOSS team communication across channels. In *Proceedings of the Fourth International Conference on Open Source Software (IFIP 2.13)*, Milan, Italy, 7-10 September 2008b.

Andrea Wiggins, James Howison, and Kevin Crowston. Measuring potential user interest and active user base in FLOSS projects. In Boldyreff et al. (2009).

Candie C. Wilderman. Models of community science: Design lessons from the field. In Cathy McEver, Rick Bonney, Janis Dickinson, Steve Kelling, Ken Rosenberg, and Jennifer L. Shirk, editors, *Proceedings of the Citizen Science Toolkit Conference*, Ithaca, NY, 2007. Cornell Laboratory of Ornithology. URL `http://www.citizenscience.org`.

Aubrey Wilson and Gina Pimm. The tyranny of the volunteer: The care and feeding of voluntary workforces. *Management Decision*, 34(4):24–40, 1996.

Bo Xu, Donald R. Jones, and Bingjia Shao. Volunteers' involvement in online community based software development. *Information & Management*, 46(3):151 – 158, 2009.