Purposeful Gaming & Socio-Computational Systems: A Citizen Science Design Case

Nathan Prestopnik
Syracuse University
337 Hinds Hall
Syracuse, NY 13244-4100
315-443-2911
napresto@syr.edu

Kevin Crowston
Syracuse University
348 Hinds Hall
Syracuse, NY 13244-4100
315-443-1676
crowston@syr.edu

ABSTRACT

Citizen science is a form of social computation where members of the public are recruited to contribute to scientific investigations. Citizen-science projects often use web-based systems to support collaborative scientific activities, making them a form of computer-supported cooperative work. However, finding ways to attract participants and confirm the veracity of the data they produce are key issues in making such systems successful. We describe a series of web-based tools and games currently under development to support taxonomic classification of organisms in photographs collected by citizen-science projects. In the design science tradition, the systems are purpose-built to test hypotheses about participant motivation and techniques for ensuring data quality. Findings from preliminary evaluation and the design process itself are discussed.

Categories and Subject Descriptors

H.5.3. [**Group and Organization Interfaces**]: Collaborative Computing.

General Terms

Design

Keywords

Citizen-science, socio-computational systems, purposeful gaming motivation, engagement, data quality.

1. INTRODUCTION

Citizen science is a phenomenon where members of the public are recruited to contribute to scientific investigations [1, 2]. Notably successful citizen-science projects include asking participants to help classify astronomical photographs, report bird sightings, count insects in the field, or use spatial reasoning skills to align genomes or fold protein strings. Such activities draw many individuals into a cooperative endeavor toward a common scientific goal. They feature a mix of tasks that can only be performed by people (e.g., making an observation or classifying an image) supported by computational scaffolding to organize these efforts. As such, citizen science often relies on some form of socio-computational system. While citizen science has a long history, such systems are relatively new, providing a variety of open questions of great interest to those who study socio-computational systems, as well as to scientists who may wish to use citizen science approaches to support their own research.

Group'12, October 27–31, 2012, Sanibel Island, FL, USA.

An interesting and sometimes challenging issue for citizen science is that some scientific topics are highly "charismatic" but many others are not. For example, bird watching, astronomy, and conservation all have existing communities of interest and a certain appeal, even for non-enthusiasts. However, important work is also being conducted in areas that attract much less public interest, such as moth, mold, or lichen classification. While enthusiasts exist for virtually all areas of the natural sciences, socio-computational systems rely on attracting large numbers of participants. As a result, the motivations of citizen science participants are important to understand, to attract new participants and retain old ones.

Furthermore, while some citizen scientists are quite expert, many are not and indeed, many may be novices. Therefore, successful projects must develop scientific tasks that can be performed by novices, while still ensuring the interest of those with more experience. Assuring the quality of data produced by the non-expert citizens using these systems is also of concern. The specific interest of this research, therefore, is to explore the relationships that exist between citizen science, socio-computational system design, attraction and retention of participants, and the impact of these on data quality.

Unfortunately, it is difficult to use current, real-world citizen-science projects as vehicles for exploring motivation, participation, users, technology, and data quality. The challenges are practical: citizen science project developers, researchers, and managers have little time available to devote toward research projects not directly related to their specific object of inquiry. Because currently instantiated citizen-science projects are working production systems, it is difficult to adjust project parameters, conduct experiments, issue surveys, interview participants, or otherwise gather information about the citizen science phenomenon. Invasive data collection efforts are likely to be disruptive and may have deleterious impacts on existing participant enthusiasm and data quality. In short, the potential drawbacks of granting complete access to socio-computational researchers outweigh any benefits that might accrue.

On the other hand, low-impact methods of investigation (e.g., interviewing or surveying staff members or researchers, passively gathering information about project websites and systems, etc.) are less likely to produce data required to address motivational and data-quality questions. Studying citizen science without fine control over the systems of interest creates a different problem: artificiality will infect any knowledge generated by such research, as simulations, mock-ups, and de-contextualized inquiry substitute for realistic exploration of actual systems that are highly situated within complex problem spaces.

We address these challenges by developing socio-computational systems explicitly designed to serve a dual purpose: as vehicles for scientific inquiry and as functional and useful systems built and deployed to solve specific, real-world problems. Building systems is not a new approach to research, but the approach has recently been reconceptualized under the name design science. This approach resides in the familiar territory of system design and evaluation, but wraps these well-known activities around a broader research agenda targeted at natural or social-psychological science. The strength of this approach is that complex phenomenon such as socio-computational systems and/or citizen science can be explored in a very realistic manner, while maintaining a great deal of control over the user experience.

The remainder of this paper is divided into three parts. First, a discussion of design science is presented. Second, an ongoing design science project in the socio-computational and citizen science domains is described. This project involves the creation of several games and tools to support an important science task in the biological sciences: species classification. Finally, results from the design process so far and from preliminary evaluations are reported, including discussions of the design science approach as a vehicle for socio-computational systems scholarship.

2. DESIGN SCIENCE

Design science is an approach to scholarly study that couples traditional research methodologies with the development of an IT artifact to address natural science or social-psychological research questions coupled with design-related problems [3-5]. Design science is practiced (mostly without using the term) in many domains, particularly human-computer interaction (HCI) and computer science (CS) more generally. The term and its formal conceptualization come from the field of information systems (IS), where system design is often viewed as atheoretical and so not research. In this setting, rigorous conceptualizations of design as a research tool are necessary to encourage its broader acceptance. However, even in fields where system design is generally embraced, the reconceptualization can be valuable, as the focus on designing useful artifacts often results in inattention to larger research questions. For example, in [6] many HCI evaluation practices are criticized as "usability evaluations" instead of scientific "evaluations for research", what [7] calls the "I did this and it's cool" form of study.

Design science research has two equally important outcomes: 1) a functional IT artifact that helps address a specific, challenging, and practical design problem within a given context, and 2) meaningful scholarly contributions to a field of inquiry. Compared to typical social-science research approaches, the design science approach requires additional components, including interactions with subject-matter experts (SMEs), a situational focus on the context in which a design will be deployed as well as system building and testing. Compared to typical systems research, the approach requires explicit use of theory to guide design decisions and—importantly—an ability to draw more general conclusions about these theories from the experience of building the system.

The problem spaces addressed by design science inquiry are typically complex, sometimes referred to as "wicked" problems because they defy easy or obvious answers [8-10]. Problems suitable for a design science approach include both those that are unsolved and those which offer opportunities for newer or better solutions [3]. However, to be meaningful to researchers outside of the specific problem space, the IT artifact must also become a vehicle for broader natural science or social-psychological inquiry. Theory, design and evaluation are thus interrelated in design science research, coherent pieces of a whole [11] and conducted iteratively [3, 4].

Theory: The word "theory" is used broadly here [12], encompassing the adoption of existing theory as a lens through which to approach design, as well as consultation with experts and review of non-theoretical, project-specific design literature. This stage may also result in the generation of new theory, produced either from literature or from data, and conceptualized either prior to design of the IT artifact, during its development, or after its evaluation. The theory stage may be seen as both a beginning and an end to design science research: theory adopted early will inform design, and new theory will come from it.

Design: Design science research revolves around the design of an IT artifact, where theoretical and practical underpinnings shape a functional system. The designed artifact may ultimately produce new theory, so artifact design must take future evaluation into account. The design scientist must always keep in mind the research questions to be addressed through research evaluation of the artifact.

Evaluation: The evaluation stage is about more than saying "yes this worked," or, "no, this didn't work." It must address the project's broader research questions by validating adopted theory or leading to the generation of new theory. Evaluation is not always an end point for research; evaluation will often suggest ways to improve the artifact (as a system to address the problem space or as a research tool) in its next design iteration.

3. CITIZEN SCIENCE DESIGN CASE

In this section, we describe our socio-computational system project situated in the citizen science domain, with emphasis on our research goals, the problem space, and design parameters.

3.1 Research Goals

Our study addresses two research questions. First, a critical issue in socio-computational system design generally, and citizen science systems in particular, is attracting and retaining enough participants to make achievement of project goals possible. Systems with too little participation will be unlikely to generate meaningful quantities of scientific data.

To address this question, we draw on psychological theories about motivation [e.g. 13]. In [14], three basic motivations for individuals who are engaged in collective on-line activities are suggested: money, love, and glory. For citizen-science projects, offering payment to participants is rarely an option (project resources are typically too low), and most participants do not expect compensation for their efforts. Instead, participants indicate that inherent interest in the subject of scientific inquiry, the relevance of data collection efforts to particular interests or hobbies, the perception that a project will be fun and engaging, an interest in collaborate with experts, altruistic reasons, and hope for broader recognition as reasons for becoming involved in citizen-science projects [15-19]. These reasons match well with the notions of “love” and “glory” as motivators [14]. There has been less scholarly or practical attention paid to how citizen science systems might be designed to motivate participants who do not hold these predominantly intrinsic motivations. As a result, most citizen-science projects rely heavily on participants who have preexisting enthusiasm for the scientific topic of the project, be it astronomy, bird watching, or classifying insects.

In the broader collective computing domains, several models for attracting participation have been deployed. In systems such as von Ahn’s reCAPTCHA [20], which facilitates optical character recognition (OCR) on scanned books, the system is devised as an obstacle between users and their goals; reCAPTCHAs are used to verify that login attempts to web systems are coming from a human user, and to log in, users must use the reCAPTCHA tool. Other systems, such as the ESP game (an image tagging system) [21], Phetch (which produces accessible descriptions of images) [22], or TagATune (where users tag music clips) [23] are designed as games, capitalizing on “love” forms of motivation, and giving people enjoyable activities to undertake while also producing meaningful work almost as a by-product.

Games in particular seem to have great potential as a motivator for participation and as a tool for producing high quality scientific data. However, from a review of citizen science websites [2], it seems that few existing projects use games to motivate participation. Notable exceptions include *Fold It*, which disguises the science of protein string folding as a highly engaging puzzle game, and *Phylo*, where players compare genetic sequences in a colorful and abstract puzzle game. Both capitalize on human spatial reasoning abilities to solve problems that are difficult to automate. The *Fold It* player pages (http://fold.it/portal/players) reveals that more than 300,000 players are contributing to this project; furthermore, *Fold It* recently made headlines for an important AIDS research breakthrough generated by players of the game. Some projects, like *Stardust@Home*, incorporate game-like elements such as leader boards, high scores, or other participation metrics, but do not frame their scientific activities as games per se. Scholarly study of socio-computational games and games for citizen science may produce insights into how different participant groups can be attracted to citizen-science projects and motivated to participate in them.

Our second research question is about techniques for ensuring data quality, a necessary precondition for further scientific use of the data, but difficult for several reasons. First, for many scientific problems there is “ground truth,” i.e. correct answers. Participant opinions are not as inherently valid as they might be in systems designed to produce, for example, image tags for search engines. For data to be scientific, valid, and accepted, the right answers must be produced by participants and confirmed by experts. Second, in many areas of science, specialized knowledge is required to provide data, but few citizen science participants are experts. Furthermore, the effect of systems (especially game-like interactions) on data quality is largely unknown. Therefore, finding methods to turn scientific tasks into things that non-scientists can do well, as well as finding techniques to confirm the validity of participant-provided data, are important research goals. To address these questions, we draw on theories from the problem domain, which we describe next.

3.2 Problem Space

The problem space we address in this design research comes from the biological sciences, particularly entomology, botany, and oceanography. In this domain, experts, enthusiasts, and curious members of the general public routinely collect and upload photographs of different living things. A photograph of an insect, plant, or animal, tagged with the date and location where it was taken, can provide valuable scientific data, e.g., on how urban sprawl impacts local ecosystems or evidence of local, regional, or global climactic shifts. However, to be useful, it is necessary to know what the picture is of, expressed in scientific terms, i.e., the scientific name of the species depicted. Some participants have the necessary knowledge (e.g., avid birders can generally identify particular bird species), but many potential participants do not.

To aid in identification of the species of specimens, biologists have developed taxonomic keys, which identify species from their particular combinations of characteristics, known as character-state combinations (i.e., attributes and values). The specific characters and states vary by taxon, but are broadly similar in structure. For example, a moth character might be its “orbicular spot,” with states including, “absent,” “dark,” “light,” etc. Given sufficient characters and states, it is possible to identify a photographed specimen to a specific family, genus, species, or even sub-species.

A challenging aspect of this problem is that researchers working within the same biological or ecological disciplines do not necessarily agree upon taxonomic keys. In fact, many researchers develop their own key variations to support their own specific research endeavors. Keys are therefore typically written for expert users, and are often complex, highly variable, and difficult to translate into a form that will be suitable for use in a socio-computational system, where expert understanding of characters, states, and taxonomic identification cannot be assumed.

A second challenge is that even with an established key, some characters and states are beyond the ability of most members of the general public to identify without training (e.g., the previous “orbicular spot” example). Others require true expert knowledge to apply (for example, classifying species by their sex organs). In some cases, especially for sub-species, true identifications cannot be made without access to specialized equipment; for example, some species are distinguishable only through their genetic makeup. This means that an IT artifact designed to support the classification task will be unlikely to effectively support both extremely knowledgeable users and extremely novice users; experts will require advanced tools with great flexibility, while novices may require simplified systems that have expert knowledge pre-built into them. In both cases, a web-based classification system will only be able to support some kinds of characters and states, while others will be impossible.

3.3 Design Parameters

To explore the motivations of citizen science participants and address the challenge of species classification in the biological sciences, a series of IT artifacts were designed and implemented. IT artifacts were designed and developed by a team of 21 professionals and students with varied technical and artistic expertise. Thirteen of the developers were hired on the project as either part- or full-time employees or volunteers. The remaining developers participated through their coursework (i.e. developing systems or components of systems for a class). Because this research is supported by a large and diverse group of developers, an ambitious program of design and development was organized, including five components that address specific aspects of the problem space, enabling exploration of our research questions.

3.3.1 *Artifact 1: Citizen Sort*

The IT artifacts hosted on the *Citizen Sort* website (http://www.citizensort.org) include both tools and games, organized along a continuum from "tool-like" to "game-like." Arranging the systems in this manner allows for comparative evaluations of participant motivation with regard to tools, games, and IT artifacts that fall somewhere in between. In addition, this arrangement allows researchers to manipulate specific website elements to either direct participants to tools or games or allow participants to self-sort based on their individual interests.

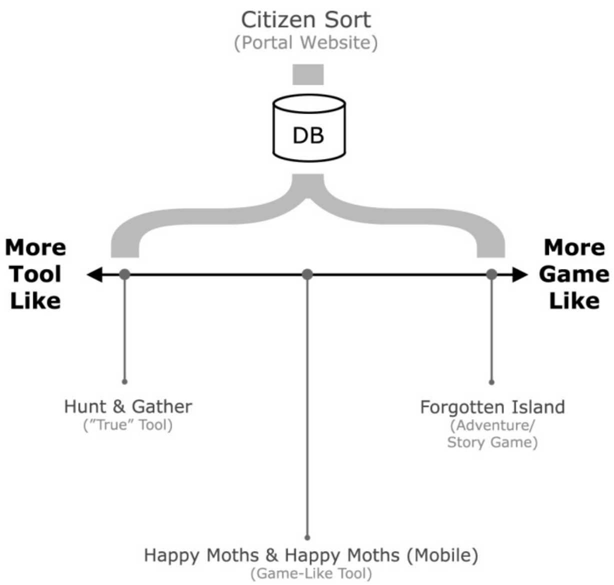

Figure 1. The *Citizen Sort* research design shows a theorized continuum from very tool-like instantiations to very game-like instantiations. Different user groups are hypothesized to be motivated by artifacts in different places on this continuum based on their personal goals, expectations, and interests vis-a-vis citizen science.

Four of the major artifacts of this design effort are organized around a fifth, a portal website (*Citizen Sort*) designed to direct participants to a variety of tools and games for biological classification. The portal website controls global functionality, including features like user-account management, administrative management of tools and games, content management of the website itself, dissemination of project data, and management of subsidiary projects. A centralized database ties all IT artifacts in this project tightly together.

3.3.2 *Artifact 2: Hunt & Gather Tool*

Hunt & Gather is a "true" tool, designed without additional motivational elements (see [24] for a discussion of motivators vs. satisfiers in web applications). *Hunt & Gather* lets users create characters and states for themselves, tag large numbers of photos with those characters and states, and let other knowledgeable individuals work with the characters, states, and photos on a per project basis.

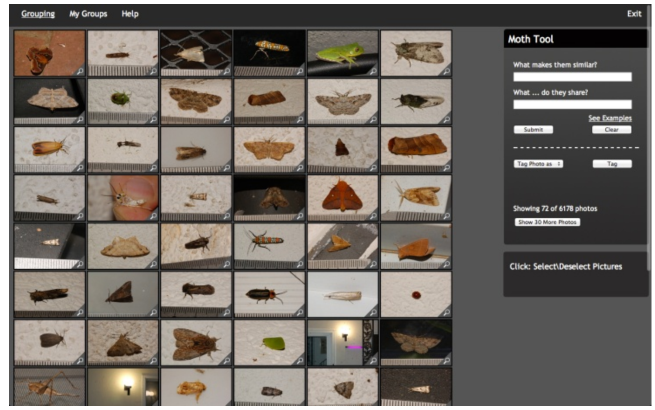

Figure 2. *Hunt & Gather* classification tool. Users can set up a collection of photos and work together to develop a taxonomy of characters and states.

Hunt & Gather will allow socio-computational researchers to explore the motivations of users who are attracted to citizen science tools, rather than games; it is hypothesized that these users will be experts or enthusiasts. Furthermore, characters and states created by novices or enthusiasts can be compared to characters and states generated by professional scientists. *Hunt & Gather* will help explore how good non-expert users are at producing characters and states that might be useful to experts in the biological sciences.

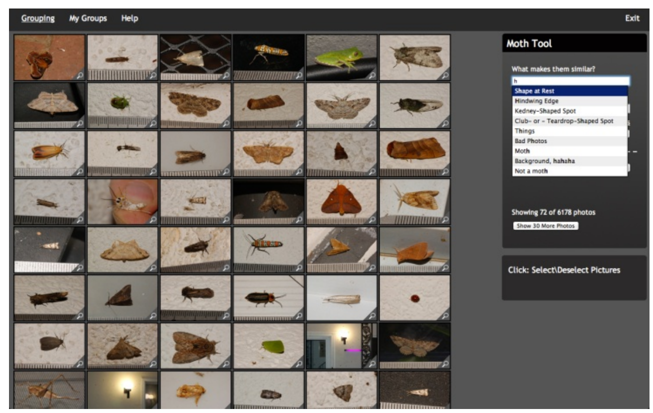

Figure 3. *Hunt & Gather* allows a group of users to collectively develop and refine taxonomies for a given collection of photos. Drop-down suggestions keep users informed about the characters and states that have already been defined. Pre-set choices for bad photos can be used to filter unwanted images from the collection permanently.

3.3.3 *Artifact 3: Happy Moths*

Happy Moths (designed to be renamed for each new instantiation: *Happy Sharks*, *Happy Plants, etc.*) is a "game-like tool," in that it offers tool-like functionality but structured as a game. Participants are presented with a set of ten photographs of some organism (in *Happy Moths*, pictures of moths) and then asked to identify the various character-states of each. One difference between *Happy Moths* and *Hunt & Gather* is that the design aims to increase participant motivation by providing a score (per round and overall) giving feedback on performance. *Happy Moths* players are scored based on how well their classification decisions match those of a previously classified-photo that is seeded into the game (the "Happy Moth"). Because players will not know which photo is the Happy Moth until the end of each game, they need to do well on all photos to ensure a high score.

Figure 4. ***Happy Moths*** **setup screen, where photos can be pre-sorted as bad images or not an example of the specimen of interest.**

Figure 5. ***Happy Moths*** **game round, where players are asked to answer a question (character) by dragging a photo to the appropriate answer (states).**

A second difference is that *Happy Moths* is built around characters and states established by professional scientists as a useful taxonomic key. *Happy Moths* is a more controlled experience for users, and may ultimately produce more reliable data when used by novices or enthusiasts with limited classification experience. As well, the quality of a player's performance on the Happy Moth can be taken as evidence of their data quality, and agreement among classifications performed by different users on the same photo can be used as an indicator of data validity.

Happy Moths also includes a mobile version, developed as an HTML5 mobile app and deployable on a variety of mobile devices. The mobile version of the game is very similar to the web-based version of *Happy Moths* (both systems contain the same logic and draw upon the same API and database). *Happy Moths (Mobile)* will introduce mobile technology as a variable in comparative evaluation studies; it will be useful in exploring whether mobile technologies make this game seem more or less game-like to users and whether ubiquitous access will help attract participants. It can also be used to collect data about where, how, and by whom the mobile version of the game might be used, and it will be possible to compare the quality of data produced by players of the two versions of the game.

Figure 6. ***Happy Moths*** **scores page, where players are provided feedback on their performance, and rewarded for correctly classifying the hidden "Happy Moth."**

3.3.4 Artifact 5: Forgotten Island

Finally, an important goal of this research is to explore the full range of the "tool-like" to "game-like" continuum. Few citizen-science projects attempt to leverage the power of storytelling or fantasy in games to motivate users. In [25-27], these elements and others are noted as key motivators in educational games; it is hypothesized that such motivators will hold true in citizen science games as well. To explore this hypothesis, as well as to generate insight into the kinds of users who might be attracted by such a game, the fifth IT artifact in this design-science project is a point-and-click adventure game called *Forgotten Island*.

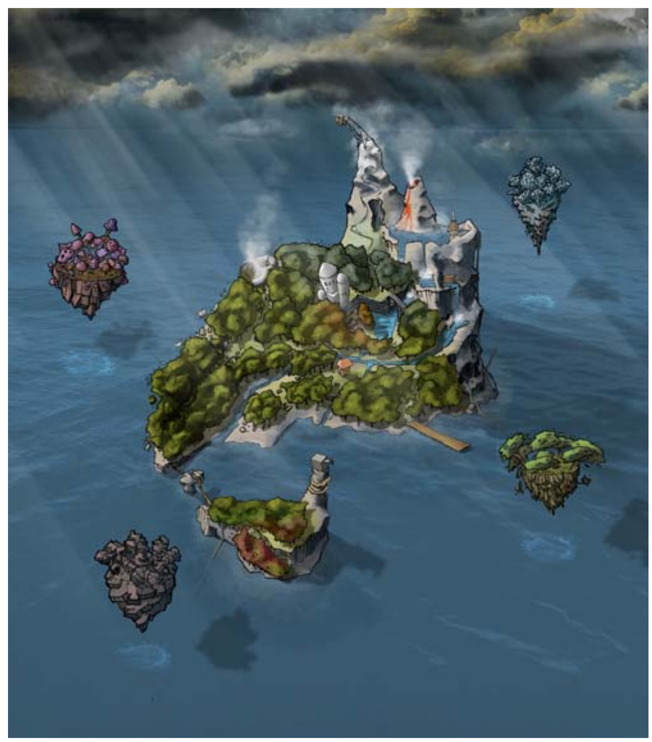

Figure 7. ***Forgotten Island's*** **game world is a mysterious island that the player explores and rebuilds while undertaking citizen science classification task. The game world was deliberately designed in a hand-drawn style to accentuate that exploring will be a fun, engaging, and whimsical experience for the player.**

Figure 8. The game world is made more mysterious and detailed through immersive, explorable locations. Unlocking these locations and advancing the story requires in-game tools and equipment that can only be acquired by undertaking the classification activity to earn game resources.

Figure 9. In *Forgotten Island*, the game story motivates the classification task. Story elements are conveyed to players through a comic book style interface.

Figure 10. The classification task itself is similar to *Happy Moths*. Key differences include that players in *Forgotten Island* classify one photo at a time to preserve a better balance between game experience and science experience, and that classifications earn the player endogenous rewards (game money) instead of feedback on their scientific contributions and point-based scores in competition with other players.

Forgotten Island is story driven, featuring an island to explore and a mystery to unravel. Players still classify insects, plants, or animals as in *Happy Moths*, but the classification task is motivated by the story and designed to fit into the background texture of the game. Players use classification as a way to earn game money that can be used to purchase equipment or items to progress the fantasy story.

Forgotten Island allows us to explore how endogenous reward systems can motivate players to participate in a scientific collaboration. It will also help explore how established taxonomies of motivational game features for learning [e.g. 25, 26, 27] might apply to non-educational games. Two additional and conflicting hypotheses will be evaluated: 1) that a fantasy adventure game will improve scientific data quality because players will be immersed in the game experience, motivated, and willing to provide high quality data, or 2) that a fantasy adventure game will reduce data quality because players will be more interested in progressing the story than in doing science, and will be willing to "cheat" on the science task to get ahead in the game.

4. EVALUATION METHOD

Prior to starting system development, background research was conducted in the form of literature review, analysis of ongoing citizen science project systems, and SME interviews. Ten SME interviews with nine scientists and developers who are currently undertaking citizen-science projects were conducted. This phase of the project informed research questions and planning for the IT artifacts to be developed, and is reported in more detail elsewhere. As design progressed, additional SMEs were consulted, including naturalists with expertise in classification. Consultation with experts is ongoing, shifting between formal, interview-style consultation and informal participatory-research approaches [28].

This research is in now the design stage, with limited formal evaluation so far. In design science, however, design activities are a central aspect of research and are a vehicle for producing new knowledge. Accordingly, we have developed an evaluation strategy that includes some evaluation activities that take place during design. Individual developers working on the project have been asked to periodically review the games and tools where they have had a central development role, as well as games and tools where they have not been as directly involved. These reviews focus on the artifact itself, rather than individual work practices. Currently, 31 reviews have been collected on three of the five projects (*Happy Moths*, *Hunt & Gather*, and *Forgotten Island*).

In addition, formal focus group evaluation sessions have been conducted, targeted at two different versions of *Happy* Moths. An early focus group session brought four expert entomologists together codify their knowledge of the classification task and to collect their impressions of an early prototype of the *Happy Moths* game. Results from this session resulted in several changes to the game. Participants in a second set of focus groups were students as a large university located in the northeastern United States. Five participants were recruited from an outdoor club and environmental conservation courses, while three were recruited from the university's School of Information Studies. These groups were classified as "nature" and "gamer" participants respectively. Participants were asked to play a new version of *Happy Moths* and provide their opinions on two different visual designs: a "gamer" version designed to look more like a video game, with no naturalistic visual motifs other than the classification photos themselves, and a "nature" version designed to appear more tool-like while showcasing a variety of nature imagery and content.

5. PRELIMINARY RESULTS & DISCUSSION

5.1 Participant Groups

During the first *Happy Moths* focus group session, SMEs helped to define three groups of potential participants who will be important for this research: 1) experts (professional scientists), 2) enthusiasts (individuals with intrinsic interest in science and/or the particular topic of a citizen science project), and 3) gamers (ordinary citizens with no particular interest in citizen science, but at least some interest in online games or entertainment). Because it may be difficult for some projects to attract enough expert and enthusiast users to be viable (especially those lacking "charismatic science" that is inherently interesting to many people), the gamer user group is of particular interest. The gamer group is hypothesized to be much larger than the enthusiast or expert groups, making it a potentially valuable source of participants. However, the gamer group, by definition, is composed of individuals who have virtually no knowledge of scientific classification. Finding ways to make the classification task enjoyable and, critically, understandable to these users will be an important outcome. One way of addressing this challenge, used in *Happy Moths*, is to have SMEs generate character questions and state answers that make sense to laypeople. So, for example, *Happy Moths* asks about simpler character-state combinations such as color or shape, and avoids complex questions about "discal spots," orbicular spots," "reniform spots," etc. In many cases, technical language has also been simplified to help lay users understand characters and states without the need for extensive training. In the *Happy Moths* focus group, SMEs had conflicting opinions about these approaches; some agreed that simplifying the tasks and language would be beneficial and still produce good data, while others felt that more technical nomenclature should be preserved as a learning opportunity for participants.

This disagreement raises another point about the differences between users: systems that motivate gamers may actually be de-motivating to enthusiasts and vice-versa. In the first *Happy Moths* focus group session, researchers suggested that systems designed to appeal to gamers (e.g., *Forgotten Island*) have a high likelihood of alienating enthusiasts. Enthusiasts are seeking opportunities to explore their passions and interests, while gamers are seeking entertainment. Over the course of design and evaluation so far, it has emerged that as a system focuses more on entertainment, it imposes increasing obstacles on enthusiasts who seek rapid access to their hobby of choice. For example, *Forgotten Island* paces the classification task and requires players to explore a variety of locations, collect items, and undertake many other story-driven activities besides classification. For an enthusiast interested in classification, these extra activities may be perceived as annoying wastes of time, rather than as engaging or fun. Similarly, SMEs frequently suggest that players will be more engaged and motivated if they learn something about science, but it is not clear that gamers will be similarly motivated.

5.2 The Role of Iteration

The purpose of taking each project in this design science study through several design iterations is threefold: each iteration 1) improves the IT artifact's ability to address the problem space, 2) produces new research findings, and 3) helps to eliminate poor system design as a confounding factor for research.

In the case of *Citizen Sort*, many specific design decisions have been discussed with the project's SMEs, particularly the decision-making that went into the *Happy Moths* game, which has (because it encapsulates the core classification task) received the most formal evaluation to date. Many design decisions have been upheld, while a few have been questioned (e.g., the visual style of *Happy Moths*, where expert reviewers suggested that a more "natural" or "nature-themed" design would better appeal to enthusiast users). In some cases, design decisions have been rejected outright. In the first iteration of *Happy Moths*, music was included, but focus group SMEs and the developers themselves unanimously rejected the choice to include music after testing it in several different settings. Now finishing its third and final iteration prior to public release, *Happy Moths* has no music and a streamlined game mechanic that is expected to be more fun and less distracting for players.

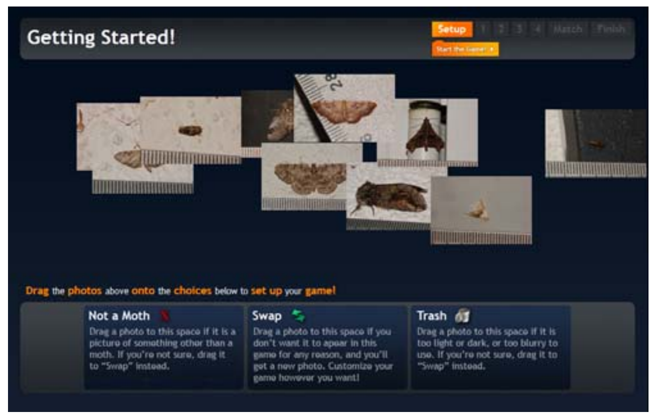

Figure 11. The "gamer" version of *Happy Moths* (version 2).

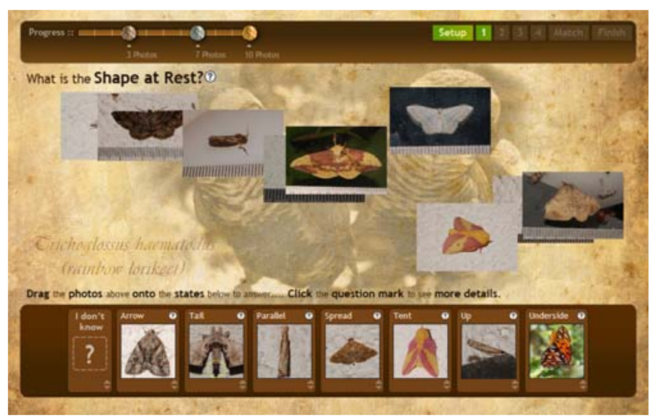

Figure 12. The "nature" version of *Happy Moths* (version 2).

The second round of *Happy Moths* focus groups were specifically designed to test another issue which came out of preliminary evaluation: the visual design of the game. Early visual workups used a contrasting color scheme of dark blue and bright yellow-orange. However, SMEs who are participants in our design partnership, as well as early focus group SMEs, suggested that this "gamer" design "did not emphasize nature enough." Both groups of SMEs were composed of professional naturalists with enthusiasm for science and the outdoors; the blue-orange color scheme didn't speak to them as players. To explore this issue in more detail, second round focus group participants were asked to play one of two versions of the game: a version using the blue-orange color scheme, and a functionally identical version using a nature-themed color scheme and nature imagery.

These focus groups revealed a mix of opinions on visual design. Both gamer and nature participants stated that each design was well conceived and visually attractive. In general, nature participants preferred the nature version of the game, while gamer participants preferred the gamer version. There was agreement that the nature version better supported the science task thematically, while the gamer version's contrasting color scheme made it easier to see and interact with the game space. Participants also agreed that the visual design was a less important issue than usability and how fun and engaging the game's scoring mechanic would be. These findings were very similar to opinions supplied by the development team during individual formal reviews.

We designed the third and final version of *Happy Moths* (see section 3.3.3) drawing extensively upon all of the previously collected evaluation data: focus groups (SME, gamer, and nature), participatory interactions with partner SMEs, developer reviews, and our own prior design experience (for a discussion of design precedent and design as an experience-based activity, see [29, 30]). Great attention was paid to the many perspectives espoused over the course of previous evaluations and design activities, resulting in a greatly polished, more engaging, and more usable game. More extensive evaluation of *Happy Moths* will take place over several months as the game is published online and played by participants as a live citizen science project. This "evaluation for research" [6] will help us to address our deeper research questions on socio-computational system design, motivation, and data quality. The evaluations conducted so far have already helped direct us toward several possible areas of interest.

5.3 Task Gamification vs. Game Taskification

Socio-computational and citizen science games are often developed by "gamifying" an existing task. The *Happy Moths* game adopts this approach, taking a classification task and adding game elements to it: a game-like visual design, scores for doing well, achievements for long-term involvement, leader boards, and high scores to promote competition between players.

The *Citizen Sort* project explores an alternative model, referred to here as "game taskification." In this approach, the typical model of turning tasks into simple games is inverted; rather, the designer starts with the game, rather than the task, designing an interactive entertainment experience and drawing upon well-understood commercial game design principles [e.g. 31]. Rather than simply re-conceptualizing a given task as a game-like activity (i.e. giving players game points for classifying a photo), the game designer must conceptualize the task as just one element or mechanic to be part of a larger (possibly *much* larger) game world. To be effective at generating data, the task must be incorporated in a way that makes it critical to progress through the game, but it need not be the focus of the player experience as it would be in a gamified task. For example, the scientific task might become a way of earning game money, a tool to power up one's character, a lock-picking puzzle, or a host of other possibilities.

The game-taskification approach opens up dramatic possibilities for purposeful games: exploring scientific content through unique themes and stories during play, building unexpected and exciting connections between entertainment and science, or engaging large segments of the population who may not be motivated by gamified tasks alone. However, the game taskification approach is rarely pursued in citizen science or socio-computational system design. Our design and evaluation process for *Forgotten Island* helps to explore one possible reason why not.

Simply put, turning a task into an enjoyable game is a complex endeavor. Developing a fantasy/story game like *Forgotten Island*, which "seduces" players into doing real science [32, 33] without foregrounding the task itself, is an exponentially larger effort than simply implementing the task. By placing the scientific task into the background of a fantasy game, developers are suddenly confronted with a host of new design requirements that are unrelated to the central rationale for designing the game (i.e., in the citizen science domain, collecting scientific data). These include developing a story and writing a script, creating locations, producing concept and final artwork, designing characters, envisioning and producing a compelling sound design, composing a musical score, programming complex functionality such as path finding or AI algorithms, planning and implementing puzzles, and more. In short, the research scientist must take on the role of game director, a role for which few are prepared.

Evaluations of *Forgotten Island*, which includes all of the above elements as design requirements, have underscored these challenges. During individual reviews, developers were asked to make an assessment of how complete they felt an evaluated system was. Developer reviews of *Happy Moths* and *Forgotten Island* conducted at approximately the same time during the development cycle (both in November, 2011) showed significantly different averages for this estimate. *Happy Moths* was evaluated to be 85.9% complete, while *Forgotten Island* was seen as being only 17.5% complete. This contrast in the remaining time to complete each game seems starker with additional information: at the time of review, *Forgotten Island* was still in its first iteration while *Happy Moths* was finishing its second design iteration and moving into its third (i.e., less complete than most developers assessed, but still much more polished than *Forgotten Island*). Currently, *Happy Moths* is nearing completion on its third and final version, while *Forgotten Island* is nearing completion of its first (and due to time and budget constraints, final) iteration.

Given the challenges of development, game taskification may or may not be as realistic an approach for designing citizen science games as better-understood methods of task gamification. A host of extra creative design activities can lead to longer development times and many more required resources. However, these costs may be worthwhile to incur if the end result is a game that is widely popular among the general public and produces a high number of classifications from each player.

The game economy of *Forgotten Island* rewards players with in-game currency for each classification that they complete, but also requires them to spend this money on items that are required to progress the story and finish the game. This makes it possible to balance the game economy by varying the reward amounts and item prices so that players must complete a specified base number of classifications in order to win. In its current balance ($50 reward per classification vs. between $250 and $750 cost for various items), *Forgotten Island* can be completed by a player who undertakes 183 classifications over the course of the game (assuming the player makes no classification mistakes and is also perfectly efficient in their purchases). *Happy Moths*, which adopts the gamified task approach, requires far fewer classifications for each "win" (either 5 or 10 classifications, depending on the number of photos in the game). So a player would need to play between 19 and 37 full games of *Happy Moths* in order to achieve the same number of classifications as one game of *Forgotten*

Island. If the story and fantasy elements of *Forgotten Island* engage more players and engage them for longer than *Happy Moths* (as we expect; experiencing an interactive story should be a more compelling reason for many players to undertake classification than inherent interest in nature or science), then *Forgotten Island* may eventually seem a better investment despite the lengthy development cycle.

Even if *Forgotten Island* itself fails to produce very many classifications, it is valuable as a demonstration of the taskified game approach's potential to produce scientific tools that are also commercial entertainment products. One future possibility is to develop and release games like *Forgotten Island* for profit, supporting scientific research (as well as game development activities) through sales of the game. *Forgotten Island* itself will not follow this commercial model; it is a research prototype, developed without commercialization specifically in mind. However, as a model for the game taskification approach, it will be a useful vehicle for exploring purposeful games as commercial entertainment products, unique methods for developing such games, non-enthusiast motivations for participating in citizen science, and the impact that taskified games have on scientific data quality.

A third approach to purposeful game design is not part of our current *Citizen Sort* project, but bears mentioning because it offers interesting possibilities for future study. This approach is to turn a scientific or socio-computational task into a form of payment for play. Many casual games have successfully adopted a model where micro-payments unlock game items, new content, new game mechanics, or new levels of play. Substituting classification for cash payment could be an effective way to reward users for their help and attract gamers to a project, and this is one possible future direction for our research.

5.4 Friction

One complexity of the design science approach is the friction that generates through competition between problem space, research goals, and feasibility to develop the IT artifact. These factors each require tradeoffs among the others. In the *Citizen Sort* project, SMEs want to take ownership of a suite of games and tools to support a citizen classification effort. Their primary goal is that these should produce large amounts of very high quality data. Virtually all other considerations are secondary. From a socio-computational research perspective, however, the interest is in how different kinds of games or tools can motivate different kinds of users and produce different qualities of data. It matters less that each individual tool or game produce the best quality data or attract the right kind of users, than that each game or tool helps generate useful knowledge about the research questions of interest. This means that games like *Happy Moths* or *Forgotten Island* could produce extremely poor classification data but still be a research success in providing evidence of cheating effects or problems with the fantasy/story approach. This outcome would, of course, be considered a failure by SMEs.

In [11], the need for multi-disciplinary expertise as well as expert developers on a design science project is noted, the better to adequately address both the problem space and research goals. Galison [34] describes how such collaborations can be difficult when friction between the varying goals of different interested parties develops. Galison describes the idea of "trading zones" [34] to accommodate the needs of various collaborators through a negotiating process. *Citizen Sort's* project manager takes a central role in these negotiations, coordinating various groups of SMEs and developers, ensuring that natural science and information science requirements are balanced, and verifying that the project scope is feasible for the development team. Our design efforts have validated "trading zone" efforts on this project, with research goals and the problem space largely complementing rather than conflicting with each other.

6. CONCLUSION

Design science is an approach to scientific inquiry where research goals are pursued through the development of an IT artifact positioned to address a real-world problem. This approach has many strengths, including the ability to tightly control research efforts while still enacting them within realistic use contexts. In addition, evaluation of design science efforts can address numerous research questions.

One constraint of design science is the friction that can develop between research goals, the problem space, and system feasibility. While good project management and careful attention to both researcher and stakeholder needs can mitigate these effects, friction is virtually impossible to eliminate entirely. Nonetheless, as the *Citizen Sort* project demonstrates, design science can be a valuable approach to exploring design issues in citizen science, purposeful gaming, and socio-computational system design.

7. ACKNOWLEDGEMENTS

The authors would like to thank the development team for their efforts on this project: Nathan Brown, Chris Duarte, Susan Furest, Yang Liu, Supriya Mane, Nitin Mule, Gongying Pu, Trupti Rane, Jimit Shah, Sheila Sicilia, Jessica Smith, Dania Souid, Peiyuan Sun, Xueqing Xuan, Shu Zhang, and Zhiruo Zhao. The authors would also like to thank the following for their partnership and assistance in *Citizen Sort's* design and evaluation efforts so far: Anne Bowser, Jennifer Hammock, Nancy Lowe, John Pickering, Jennifer Preece, Dana Rotman, and Andrea Wiggins. This work was partially supported by the US National Science Foundation under grant SOCS 09–68470.

8. REFERENCES

[1] Cohn, J.P., *Citizen Science: Can Volunteers Do Real Research?* BioScience, 2008. **58**(3): p. 192-107.

[2] Wiggins, A. and K. Crowston. *From Conservation to Crowdsourcing: A Typology of Citizen Science*. in *44th Hawaii International Conference on System Sciences*. 2011. Kauai, Hawaii.

[3] Hevner, A.R., S.T. March, J. Park, and S. Ram, *Design Science in Information Systems Research.* MIS Quarterly, 2004. **28**(1): p. 75-105.

[4] March, S.T. and G.F. Smith, *Design and natural science research on information technology.* Decision Support Systems, 1995. **15**(4): p. 251-266.

[5] Peffers, K., T. Tuunanen, M. Rothenberger, and S. Chatterjee, *A Design Science Research Methodology for Information Systems Research.* Journal of Management Information Systems, 2007. **24**(3): p. 45-77.

[6] Dix, A., *Human–computer interaction: A stable discipline, a nascent science, and the growth of the long tail.* Interacting with Computers, 2010. **22**: p. 13-27.

[7] Ellis, G. and A. Dix, *An explorative analysis of user evaluation studies in information visualisation*, in

Proceedings of the 2006 AVI workshop on Beyond time and errors: novel evaluation methods for information visualization. 2006, ACM: Venice, Italy.

[8] Brooks, F.P., Jr., *No Silver Bullet: Essence and Accidents of Software Engineering.* IEEE Computer, 1987. **20**(4): p. 10-19.

[9] Brooks, F.P., Jr., *The Computer Scientist as Toolsmith II.* Communications of the ACM, 1996. **39**(3): p. 61-68.

[10] Rittel, H.J. and M.M. Webber, *Planning Problems Are Wicked Problems*, in *Developments in Design Methodology*, N. Cross, Editor. 1984, John Wiley & Sons: New York.

[11] Prestopnik, N., *Theory, Design and Evaluation – (Don't Just) Pick Any Two.* AIS Transactions on Human-Computer Interaction, 2010. **2**(3): p. 167-177.

[12] Gregor, S., *The Nature of Theory in Information Systems.* MIS Quarterly, 2006. **30**(3).

[13] Crowston, K. and I. Fagnot. *The motivational arc of massive virtual collaboration.* in *IFIP WG 9.5, Working Conference on Virtuality and Society: Massive Virtual Communities*. 2008. Lüneberg, Germany.

[14] Malone, T.W., R. Laubacher, and C.N. Dellarocas, *Harnessing Crowds: Mapping the Genome of Collective Intelligence*, in *MIT Sloan Research Paper No. 4732-09*. 2009.

[15] Bradford, B.M. and G.D. Israel, *Evaluating Volunteer Motivation for Sea Turtle Conservation in Florida*. 2004, University of Florida, Agriculture Education and Communication Department, Institute of Agriculture and Food Sciences: Gainesville, FL. p. 372.

[16] King, K. and C.V. Lynch, *The Motivation of Volunteers in the nature Conservancy - Ohio Chapter, a Non-Profit Environmental Organization.* Journal of Volunteer Administration, 1998. **16**(5).

[17] Raddick, M.J., G. Bracey, K. Carney, G. Gyuk, K. Borne, J. Wallin, et al., *Citizen science: status and research directions for the coming decade*, in *AGB Stars and Related Phenomenastro 2010: The Astronomy and Astrophysics Decadal Survey*. 2009. p. 46P.

[18] Raddick, M.J., G. Bracey, P.L. Gay, C.J. Lintott, P. Murray, K. Schawinski, et al., *Galaxy Zoo: Exploring the Motivations of Citizen Science Volunteers.* Astronomy Education Review, 2010. **9**(1): p. 010103-18.

[19] Wiggins, A. and K. Crowston, *Developing a conceptual model of virtual organizations for citizen science.* International Journal of Organizational Design and Engineering, 2010. **1**(1/2): p. 148-162.

[20] von Ahn, L., *Human computation*, in *Proceedings of the 46th Annual Design Automation Conference*. 2009, ACM: San Francisco, California.

[21] von Ahn, L., *Human computation*, in *Proceedings of the 4th international conference on Knowledge capture*. 2007, ACM: Whistler, BC, Canada.

[22] von Ahn, L., G. Shiry, K. Mihir, L. Ruoran, and B. Manuel, *Improving accessibility of the web with a computer game*, in *Proceedings of the SIGCHI conference on Human Factors in computing systems*. 2006, ACM: Montreal, Quebec, Canada.

[23] Law, E. and L. von Ahn, *Input-agreement: a new mechanism for collecting data using human computation games*, in *Proceedings of the 27th international conference on Human factors in computing systems*. 2009, ACM: Boston, MA, USA.

[24] Zhang, P. and G.M. von Dran, *Satisfiers and dissatisfiers: A two-factor model for Website design and evaluation.* Journal of the American Society for Information Science and Technology, 2000. **51**(14): p. 1253.

[25] Malone, T., W., *What makes things fun to learn? heuristics for designing instructional computer games*, in *Proceedings of the 3rd ACM SIGSMALL symposium and the first SIGPC symposium on Small systems*. 1980, ACM: Palo Alto, California, United States.

[26] Malone, T.W., *Heuristics for designing enjoyable user interfaces: Lessons from computer games*, in *Proceedings of the 1982 conference on Human factors in computing systems*. 1982, ACM: Gaithersburg, Maryland, United States.

[27] Malone, T.W. and M. Lepper, *Making learning fun: A taxonomy of intrinsic motivations for learning*. Aptitude, learning, and instruction: Vol. 3. Cognitive and affective process analyses, ed. R. Snow and M. Fair. 1987, Hills-Dale, NJ: Erlbaum.

[28] DeWalt, K.M. and B.R. DeWalt, *Participant Observation: A Guide for Fieldworkers*. 2002, Walnut Creek, CA: AltaMira Press. 287.

[29] Lawson, B., *Schemata, gambits and precedent: some factors in design expertise.* Design Studies, 2004. **25**(5): p. 443-457.

[30] Lawson, B., *How Designers Think*. 4 ed. 2005, Oxford, UK: Architectural Press.

[31] Schell, J., *The Art of Game Design: A Book of Lenses.* 2008, Burlington, MA: Elsevier, Inc. 483.

[32] Jafarinaimi, N., *Exploring the character of participation in social media: the case of Google Image Labeler*, in *Proceedings of the 2012 iConference.* 2012, ACM: Toronto, Ontario, Canada.

[33] von Ahn, L. and L. Dabbish, *Designing games with a purpose.* Commun. ACM, 2008. **51**(8): p. 58-67.

[34] Galison, P., *Image and Logic: A Material Culture of Microphysics.* 1997, Chicago: The University of Chicago Press.