Current Developments in Big Data and Sustainability Sciences in Mobile Citizen Science Applications

Nikita S. Panchariya*, Andrew J. DeStefano*, Varsha Nimbagal*, Revathi Ragupathy*, Serkan Yavuz*, Katherine G. Herbert***, Emily Hill**, and Jerry Alan Fails*

*Department of Computer Science
Montclair State University
Montclair, New Jersey, USA
{panchariyan1, destefanoa1, nimbagalv1, ragupathyr1, yavuzs1, herbertk, failsj }@mail.montclair.edu
****To Whom Correspondence Should be Addressed*

**Department of Mathematics and Computer Science
Drew University
Madison, New Jersey, USA
emhill@drew.edu

***Abstract*—Sustainability Sciences Studies is an interdisciplinary approach towards understanding how to develop a culture of conservation. This culture of conservation can be viewed from many different aspects, from the individual person's decisions to larger community's impacts. In all of this, we see large quantities of data in a push-pull relationship with each other, with stakeholders needing to have access to real-time generated data sets and analytics from populations as large as a metropolitan community and be able to respond to as well as disseminate information to these populations. In this paper, we present a survey of the literature in these areas. The pervasive and diverse nature of big data in these fields demonstrates the need for data scientists to collaborate to identify ways to address the common and disparate needs that different projects may have in relation to big data. We also briefly present a platform and implementation the Sustainability Studies Mobile Toolkit (SSMT) and Geotagger that seek to address some of these needs.**

***Keywords* — Big Data, Citizen Science, Sustainability Science, Crowdsourcing, HCI, Gamification.**

I. INTRODUCTION

By the end of 20th century, after the invention of World Wide Web (www) and super computers, the volume of data continued to grow. In 2005, Roger Mougalas from O'Reilly media coined the word 'Big Data' which referred to huge data that becomes difficult to manage using traditional databases [26]. "Big Data" is a term used to describe a massive volume of both structured and unstructured data that is so large that it's difficult to process using traditional database and software techniques. In the Techonomy Conference in 2010, Eric Schmidt stated "there were 5 exabytes of data created by the entire world between the dawn of civilization and 2003. Now the same amount is created every two days" [26]. All the data that we produce today are part of tomorrow's history.

In this paper, we review Big Data in sustainability science as it pertains to citizen science. We present how crowdsourcing and human-computer interaction (HCI) are involved with this research area. Sustainability research and experiments are seeking to overcome challenges in the world today and are searching for ways to modify the social-ecological systems to provide food, water, energy, health and human security in a way that is economically, ecologically and socially reasonable for many generations [17]. In the area of sustainability, researchers are seeking to identify the factors affecting ecological systems. These factors must be identified as in the near future they may irreversibly transform the systems around them. We closely look into how big data which is being gathered can be useful to minimize the adverse effects on the environment. We also looked into Big Data models in sustainability, sustainability evaluation methods, problems of big data models in sustainability and scientific data infrastructure for sustainability science mobile applications. The other topic of our survey is citizen science and crowdsourcing. In citizen science, we look at data transcription, investigation and collection tools such as *Sensr*, and *Biotracker*. We also consider how technology can be used for designing and maintaining cyberinfrastructure. In addition, we survey how HCI can be useful for the citizen scientist to volunteer for a project and also for researchers to attract the volunteers. Here we focus on gamification looking into the interest of the citizen scientist. Finally, we present GeoTagger and our Big Data work in sustainability and citizen science data collection.

II. BACKGROUND ON BIG DATA

A. *Background*

Big data is defined by the complexity and combination of the four V's: Volume, Velocity, Variety and Veracity. Volume is the quantity of data that is generated. In fact, humans create more than a few quintillion bytes of data every day. Velocity is the speed at which the data is ingested, integrated and analyzed for use in real time [19]. Big data is unstructured and lacks any common format. Big data formats can be in image, audio, email, etc. All these vast variety of data has to be ingested and merged with structured data to convert it to useful information. Finally, veracity, or the accuracy of big data, should be verifiable. Big data analytics can give useful information from the vast data variety using complex models.

Citizen science is a branch of science where the volunteers from the public (citizens) collaborate with the researchers for some research [7]. There are many different areas in which citizens participate where large amounts of data is collected by the citizens and analyzed by the scientists. Big Data and citizen science can be seen in action as citizen science is closely related in collection, storage and analysis of large amounts of data. For example, members of the public can participate in data transcription projects which transform physical data to digital data [7]. Some data collection tools like *Sensr* and *Biotracker* can be used by the citizens for the collection of useful data.

Big Data is the future of solving many environmental problems. The more we use Big Data, the faster we can steer our world to sustainability. Sustainability can be looked into various branches like energy, water and agriculture. There is much analysis and interpretation of data in energy but we have not used much data on water resources and agriculture. Big Data analytics on energy usage has helped to reduce the consumption cost leading us to efficient energy usage. Data about water is widely spread among different organizations that include satellite imagery and remote-sensing data [16]. Water data integration will provide the ability to cross spatial scales from local to regional to global, establish a link between science and application, to share information, develop new assessment tools to meet the collective needs and Decision Support Systems (DSS) for analysis and forecasting [16]. Big data can change the methods of agriculture and can help us feed growing number of people during the era of climate change [16].

When looking at sustainability science research, there are many overlaps where citizen science can be of aid and is even imperative to collect appropriate data to measure the success for initiatives. First, as illustrated above, citizen scientists can enable rapid, diverse data collection and aid in the translation of that data into appropriate digital forms. However, it is also important to note that the key concept that distinguishes sustainability sciences research from other environmentally related problems is the concept that it affects a change in behavior. The area of sustainability science looks at interested stakeholders in a particular sustainability consideration and reviews the sociological motivators for the behaviors involved. In many sustainability situations, the interested stakeholders can be linked to the general public or the citizen science community. Therefore, to both aid in data collection as well as assess the impact of sustainability initiatives, collecting data about citizen scientists in suitability contexts is imperative. When we consider that this can involve metropolitan size communities, with varying needs, pushing and pulling data from sources in real-time while generating new data that must may need to be funneled to other key decision makers, for example in a hurricane or earth quake situation, these problems become highly connected to Big Data. In the following sections we highlight some of the key big data projects in sustainability and citizen science, discuss crowdsourcing and HCI and their impacts on this area and then describe our project GeoTagger.

B. *Big Data, Sustainability, Citizen Science and HCI*

When considering sustainability sciences, purposes, any mobile application development would need to consider the issues of Big Data, Citizen Science and HCI. Sustainability sciences looks to impact decision making of various stakeholders in the environment to effect, hopefully, positive change. Therefore, any mobile application development would require collection of large sets of data, in real time, as well as need to visualize such data in meaningful, concise ways. Moreover, it would also need to, besides "push" pertinent data to the user, collect data from users and visualize this data to that there can be continuous engagement of users as data evolves, such as in a weather-related crisis.

Crowdsourcing is about collaborating of citizens in scientific research in order to address the real world problems. The primary goal of crowdsourcing is "to gather data/resources from non expert contributors in order to support scientific investigation".[28] One of the example of crowdsourcing is Wikipedia, online encyclopedia which is most probably used as a greatest source of information. But all this information here is crowdsourced, a huge amount of data is being collected in this encyclopedia which closely relates to a big data issue. The organization of this information properly and presenting it to the users in a way that every individual can easily understand is related to the HCI section. When HCI is introduced in crowdsourcing and citizen science projects, the cost of user participation reduces significantly as the projects are completely virtual.

HCI plays an important role in these crowdsourcing issues by introducing the concepts to the users and taking care of the user interfacing. One of the key aspects of sustainability science understands the impact on the decision maker and one of the possible methods to achieve it is through crowdsourcing technology. In order for the users to generate and use the data for sustainability purposes crowdsourcing technologies can be used and for the citizen scientist to be able to understand the project we need to bridge the gap between technology and the users where HCI comes into picture. This can be illustrated by a scenario of hurricane sandy, where there are various sustainability issues like hurricane evacuation and sheltering, human and infrastructure systems for hurricane evacuation, etc [30]. According to the United States census bureau there are 19.9 million people living in the New York metro area [29]. Therefore when we talk about crowdsourcing in sustainability one such application could be thought of where we have 19.9 million concurrent users. If there can be an application which is user friendly and crowdsourced, these users can support in the further research that how can such situations be handled in a better way.

III. BIG DATA IN CITIZEN SCIENCE

A. *Data Transcription in Citizen Science*

The *Old Weather* project (mentioned above) uses volunteers to visit the website and transcribe a number of handwritten pages from shiplog books predominantly on weather observations, occasionally interspersed with battle action, movements of personnel etc. The authors in this work want to offer a new design recommendation, which would recognize varying levels of engagement of users with the project. The new design uses gamification as the platform so that users will find it interesting to engage in the project with more enthusiasm.

How the gamification is employed: Rankings were added to the system to recognize the level of contribution made by each volunteer. This was the form of encouragement for participation. The three levels of gamification were as follows, in the first level, volunteer joins the ship as a caudate and after achieving 30 weather observations, they are promoted to Lieutenant (2^{nd} level). The top transcribers in the ship complete to be a captain (level 3). The results of the survey comprised of positive aspects like validation, tracking personal projects, competition. There were some negative aspects too, like demotivation, stress, distrust, quantity over quality and mainly trivializing research objective. [7]

B. Data Collection in Citizen Science

There are different tools that can be used for collection of large amounts of data from volunteers in citizen science projects. *Sensr* is an authoring environment for building mobile data collection and management tools, where people do not need extensive programming skills. Using *Sensr* involves combining a visual programming environment with a mobile application. *Sensr* has been implemented as both a Website and a mobile application. The Website form describes the campaign and drag and drop methods can be used to create the interface for collecting data. The mobile application comprises three tabs; My Campaign, for subscription of the volunteers and to explore through the list of campaigns; My Data, for storing the reported data; and Settings, for personal information. The authors used three case studies to investigate their design considerations. The case studies include air quality monitoring, watershed monitoring, and local parks conservation of neighborhood diesel emissions from the local participants. People shared their experiences with photos and videos. Managers who monitored the case studies observed that the interaction with the system is not a problem. The issues raised in these case studies are data quality as the data is generated from the public and there is low support for localization. Pre-existing complexity of the mobile applications also had negative effect in collecting the data [15].

Biotracker is a gamified mobile application designed to collect data for plant phenology for project Budburst. This project involves the collection of data related to perennial bud beginning and falling leaves, which. is useful for scientists studying tree allergies and global climate change. *Biotracker* primarily started using badges that are placed on the profiles. Scientists were given badges according to their activity. When a user reaches the #1 spot, he/she is given a leader badge. Before evaluating the *Biotracker* app, students were trained in downloading, installing the app and setting up a large number of Floracaches in a nearby classroom. Students go outside for 30 min to play with the application. Once they are done collecting data, they were asked 15 questions online about the quality and assessment of *Biotracker.* Most students are motivated by the fun of using *Biotracker*, whereas. The rest of the students were motivated by earning badges and completing activities. Some participants were not interested in the plants but they enjoyed using the *Biotracker* app. Contribution to science was not the major motivating factor for those who use the *Biotracker* application. Contributing to the public is also not a motivating factor as *Biotracker* does not explain how it is useful for the public. Community involvement is a significant motivation factor. Some participants felt competition between friends is a motivation factor but not competition with strangers. Even though there are many positive motivational factors for using *Biotracker*, the authors cannot conclude that the *Biotracker* application can engage participants for extended periods of time. [1]

C. Technology in Citizen Science

In the article "Technology and Work Practices in citizen science", this study presents the role of computational techniques in the citizen science projects. To design and maintain the cyberinfrastructure for the projects, the main purpose of this study is generating insights and examining the relation between information and communication technologies and work practices. The study was conducted on three citizen science projects involving independent volunteers to monitor plant species and report data regarding them through online forms. After the data was collected from the interviews, it was analyzed by the inductive analysis method. The three projects were as follows:

Parks: This is an inter-organizational partnership developing program for long term monitoring of plants and animals. Natural resource management was the primary goal. Volunteers monitored the life-cycles of plants and animals in different National Parks.

Mountains: This project was managed for the north-eastern mountain ranges for plant monitoring. Education and outreach was the focus. Volunteers reported the reproductive stages of targeted plant species.

Gardens: This project was about monitoring plant-pollinator relationships by a single academic researcher. Volunteers reported number of times bees visit the flower during 15 minutes.

The observation for all three projects was as follows: each of them used ICT along with online data entry and marketing. More communication tasks for participant support were managed by the Gardens. Identity management was one of the functions managed by the Mountains while prioritizing outreach and member development. While providing access to many distributed partners Parks operated virtually and also gave public access to shared data [27].

D. Investigation of Tools in Citizen Science

In the article *Testing a Grassroots Citizen Science Venture Using Open Design, "the Bee Lab Project*", the authors present the investigation of citizen science tools used for the Open Design to beekeeping so that participants can be motivated to gather reciprocal data. It explains the issues and opportunities for designing citizen science tools which can be used for solving individual problems.

The Bee Lab Citizen Science Project's research objective is to create tools for Citizen Science using Open Design plans. These created tools should be used to find out the pitfalls and opportunities to collect data that would be useful for community bee health insights.

In the research method hive monitoring kits and supporting assembly were designed which can be used by lay users in the work shop. In the second work shop, they have tested the effectiveness of the kits and the willingness of collecting data by beekeepers. For each kit, 10 beekeepers were recruited by British Beekeepers Association (BBKA) network. Participants were trained and tested for the assembled kits and designs. Day activities were recorded and later the kits were verified by technicians that they are working properly. Then feedback was collected if they can be sold for the retailers. Lab kits provided are designed for the mixed audience as there are participants with mixed skills. In workshops, total weight, the weight of hive feeders and hive internal temperatures measured using the designed kits.

Kits are intended to help beekeepers so that they do not over inspect the hives because it might affect the bee health. Sometimes, the wireless connectivity and GPS could cause problems for honey bees as the location of hive is sensitive. These kits record the bee hive activity periodically (hourly by default, but user editable). In the workshops, participants were also trained for assemble and disassemble the kits. Combining OD and CS tools, improve problem solving techniques which helps the community better. Kit assembly provides integrated validation and assembly feedback. Overall, transparency for users reciprocates helping the society using these tools. [22]

E. *Online Citizen Science Research*

One of the key concerns with involving citizen scientists in projects is their continued enthusiasm for the work for the life of the research project. In *Designing for Dabblers and Deterring Drop-Outs in Citizen Science* , the authors discuss how online citizen science projects have large proportions of participants who contribute in small quantities [8]. The members of public are recruited to collaborate and contribute in the scientific investigations which are called a citizen science projects. These projects are important since the participants are used for making an observation or classifying an image. The successful citizen science projects include where the participants are asked to classify the astronomical photographs, report bird sightings, count insects in the field etc. The citizen sciences often rely on the socio-computational system which is clearly explained in this paper [25].

In the previously mentioned *Old Weather* project, it is important to know how "the dabbler" contributors differ from the committed volunteers. The collaboration of scientists partnering with volunteers who are contributing in small quantities for the research of citizen science using the latest technologies like web and mobile technologies for these projects need a dedicated research work [8].

The *Old Weather* project is used as an example where both virtual citizen science and humanities project has benefits. The task is to transcribe mainly the weather observation and also any events such as war or any personnel changes recorded in the historical ships log book. The weather observations from the log book help the scientist for climate modeling. The participants often slow down after the initial flurry of activity and these can jeopardize the project sustainability but also the scalability of the citizen science [8].

IV. BIG DATA IN SUSTAINIBLITY SCIENCE

We presently lack the data, analysis, models and theories for a predicable future and therefore a new kind of integrated science is needed to cultivate insight of social-ecological systems. In order to achieve this there are at least two key needs which must co-evolve: place-based comparative long-term theory-driven research and the observation systems needed to support this research [17]. Big Data analytics can provide useful information from a vast variety of data using complex models. Computer models can be developed using existing data and with it we can predict the future of sustainability. The factors that affect our sustainable environment must be identified before we can change those factors to transform their effects in the future. The solution for a sustainable future is hidden inside big data and therefore new project proposals are being developed to reach this objective.

A. *Sustainability*

Researches delve into sustainability performing a collection of experiments and studies that require continuous variation, evaluation and education. In the paper "Measuring and Monitoring Progress Toward Sustainability", the authors identify approaches the relevance of this topic wherein experiments are conducted to reveal how sustainability will impact this generation and those to follow. This article demonstrates how sustainability can be achieved to create a resilient social-ecological system. They demonstrate it is imperative that we establish a capacity to create and execute policies for social-ecological systems, predict consequences and calculate outcomes. In the following sections, we highlight some studies that highlight the importance of these aspects. These studies demonstrate the enormous advantages that can be obtained from adaptive design and implementation of projects for conservation, development and sustainability. The challenge of "Managing Human-Environmental Systems for Sustainability" involves the flow of resources into and from use and information on stocks, rates and tradeoffs [17]. Ideally, the resource levels would be known, their changes would be monitored and the approaches to the limits of the resource would then be quantified The *SAM4SN Model* was used to observe at a macro-level how awareness can spread in a community. Communities shared some smart metering functions [4] Smart meters are a new kind of energy meter. They are a replacement for existing meters which send electronic meter readings to your energy supplier automatically. The entities of the models, also known as the agents, relate to the people involved in the consumption of one limited or critical resource. They provide real-time feedback on energy usage and its cost.

B. *Big Data Models in Sustainability*

FuturICT project is one of six proposals within the European Commission's Flagship Initiative to develop new Information and Communication Technologies (ICTs) such as

Planetary Nervous System, to enable collective awareness, Living Earth Simulator, to explore the side effects of opportunities of human decisions, Global Participatory Platform for social, economic and political participation, an 'Open Data Platform', to foster the creativity of people and new business opportunities, a 'Trustable Web', to support safer privacy-respecting information exchange, as well as value-sensitive ICT to promote responsible interaction [12]. The *FuturICT* project will create an Innovation Accelerator (IA) to identify inventions and innovations early on, to distil valuable information from a flood of information, to find the best experts for projects and to fuel distributed knowledge generation by 'Crowd-sourcing' [12]. This IA will also catalyze the integration of project activities in one single platform.

C. Sustainability Evaluation Methods

Further, the "Managing Human-Environmental Systems for Sustainability" was challenged by analyzing a variety of methods and providing the results. A seven-day space flight was used as an example of what happens when the stock and user rate are known [17]. The future use can be estimated and the end of the flight established. The outcome was that as long as the total projected use does not exceed the stock, adequate sustainability is maintained. In the Spaceship Earth example the stock was not so well quantified, the general significance was known, which meant that stock was no longer a fixed value, but its amount may have had potential to be altered and rates of use could be altered as well. Thus, the study successfully examined the layers of management and sustainability. Also studied was the relationship between macro policy and landscape level processes forms the contest within which forest commons systems are situated. A forest system and a human system and when connected constitute the forest commons system. A set of three outcomes pertaining to the forest system (carbon and diversity) and livelihoods that pertain to the human system was displayed [A10].

An Agent Based Exploratory Model (ABM) called the Spread of Awareness Model for Social Norm (SAM4SN was the method implemented to simulate actions and interactions of autonomous agents [4]. It was used to study individual and collective behavioral changes toward environmental sustainability using *ICT* (Information Communications Technology) -based services. *The SAM4SN Model* simulated the micro-behaviors of individuals regarding limited or critical resource consumption. It described the mechanism of social interaction and its effect on environmental collective behavior, aiming to avoid a rebound effect. Other types of methods can be developed to create a collective awareness of the impact that human actions have on our world [12]. We can establish a framework for assessing changes in social-ecological systems, using metrics and indicators that can be collected and compared across a range of cases.

D. Basic Problems of Big Data Models in Sustainability

We have witnessed the changing earth in the past few decades, changes beyond our comprehension. There are numerous unknown factors contributing to these changes and we will be able to identify these unknown factors if we can grasp the complex data collected. We need to build data models that can provide a predictive analysis and a prescriptive solution. These complex models should be affordable and also furnish more meaningful output for the management decisions. The important factor that affects the development of models and its understanding is the complexity of data and the different factors involved in the data and its analysis. While data analysis provides an abundant amount of information, it lacks clarity.

Hence, we need to continue conducting research and performing experiments on sustainability in our socio-ecological systems to demonstrate the enormous impact on the outcome. Since there is immeasurable potential to improve our ability to foretell the effects of policy interventions on/of human actions such as; greenhouse gas emissions, agriculture and forestry practices, nutrient mobilization, on ecosystem service and of ecosystem services on livelihoods, health, energy and food security many industries are moving towards sustainability. This is a difficult management decision executed by observing the complex result from the model [16].

E. Impact of Sustainability of our Socio-Ecological Systems

Accordingly, managing for sustainability confronts two critically contested concepts simultaneously; management and sustainability [A10]. This paper explores forms of management for sustainability that rest on other means to achieve outcomes than better knowledge. The observations in Managing Human-Environmental Systems for Sustainability impact the needs to be filled if management-through-calculation is to occur effectively in human-environmental systems. It is important to remember, human actions and interventions are only a part of the processes that influence the behavior of human environmental systems.

Consequently, the Agent Based Exploratory Model (SAM4SN) had an enormous impact on environmental awareness by identifying emerging patterns and scenarios leading to a reduction of the consumption of energy and water [4]. It also promoted a reduction in the consumption of energy and water, promoting awareness on its overuse and a better understanding of behavioral mechanism leading to sustainability. The enormous body of work on big data and sustainability highlights the difficulties of defining it but also the necessity of achieving it.

V. BIGDATA IN CROWDSOURCING

Crowdsourcing in general looks at paradigms for solving problems by collecting data, interactions, and other types of inputs from a wide varied of users and sources through a distributed mechanism. In looking at both the decision-making aspects of sustainability sciences as well as the techniques needed to engage users in large-scale citizen science projects, crowdsourcing becomes one mechanism by which data can be collected for such projects. With mobile devices and applications these crowdsourcing activities, especially under varying purposes, can results in huge quantities of data coming into the project quickly, needing near instantaneous analysis and dissemination to participants. We review some popular crowdsourcing tools to illustrate what is capable with the technology.

A. Crowdsourcing Evaluation Tools

MM-EVAl is an evaluation procedure and the tool used for crowdsourcing at a micro-level multimedia annotation. The research problems such as multimedia event recognition, video retrieval and classification, and human behavior analysis for which annotating multimedia content has become very important as there should be a precise start and end of an event which is done at a micro-level. The introduction of a web based annotation tool such as OCTAB (Online Crowdsourcing Tool for Annotations of Behaviors) which gives precise and convenient multimedia behavior annotations from Amazon Mechanical Turk interface. The annotation is evaluated by analyzing the majority of votes in the crowd-sourced repeated annotation by using crowdsourcing. The methodology used is comparing the performance of the participants in 3 categories like experts, crowd-sourced unique and majority. The methodology used is an experiment online where the behavior to annotate based on eye, facial expression, one's head moment and verbal. The experiment is based on online environment where the tasks are design for the crowdsourcing participants and analyzed it according to unique, expert and majority classifications [21].

B. Impact of Socio-Computational System Design

Design science is the method where the traditional research is combined with the development of an IT to address natural science or a social-psychological research questions with design-related problem. With design science research there are two important outcomes the functional IT artifact that helps in addressing specific challenge and practical design problem within context and the second outcome is meaningful scholarly contributions to the field. The design science requires interaction with the subject matter, system building and testing. It discusses the theory, design and also the evaluation to improve the tool furthermore. The design science is applied to the citizen science design case to study the impact of socio-computational system design to get meaningful results. This paper explores the factors like the design parameters where the design is made more tools like or games, here the participant chooses their interest to tools or games and evaluate the motivation of the participant. The evaluation method used in this paper is a literature review and conducted interviews with the scientists and the developers who are currently undertaking the citizen science projects. The citizen science and socio-computational games are often gamifying which is developed by existing task. The approach of game-taskification explores scientific content through unique themes and a story during play etc. engages the connection between the participant in the form of motivation and also entertainment and science. [25].

VI. HCI IN BIG DATA

"Human-computer interaction is a discipline concerned with the design, evaluation and implementation of interactive computing systems for human use and with the study of major phenomena surrounding them" [14]. By this definition, HCI can and most probably is used in every field of computing. In *What HCI can do for Citizen Science*, indicates that using such design methodologies when it comes to citizen science, a relatively new endeavor for HCI researchers, has been shown to benefit both citizens and scientists. Using crowdsourcing, the essence of citizen science, is a great way to obtain or analyze large amounts of data, a feat that scientists would be otherwise unable to accomplish.

However, the varying degrees of interest, education, and access, among other factors, can greatly impact the amount and quality of the work done. Equally important is how the citizens and scientists can collect and analyze the data. Therefore, the design of these interfaces must be of paramount importance.

A. Concerns in Gamification in Citizen Science Projects

The nature of citizen science projects often means that participants are performing some sort of tedious or boring activity. Also, the participants are volunteers, which compound the problem of motivating individuals. Gamification, or "the use of game design elements in non-game contexts" [7], is often used in such situations to inspire and retain participation. Gamification is a commonly used practice in citizen science and sustainable HCI in general. A good example of gamification applied to citizen science is a project called Old Weather, reviewed in the article, "I want to be a Captain! I want to be a Captain!": Gamification in the Old Weather citizen science project, which asks users to transcribe weather logs from old sailing ships. Due to the fact that the logs are handwritten and hard to read, compounded by the fact that the material is not particularly interesting and it is a tedious task, Old Weather added a ranking system. All participants join a ship and start as a Cadet, after 30 transcriptions the participant is promoted to Lieutenant. The participant with the most transcriptions will be the Captain. However, unlike the other ranks, the rank of Captain can be lost if someone else has more transcriptions for the same ship. In 2013, Old Weather had over one million transcriptions. Since the tasks in Old Weather are not particularly interesting, to have such a number of transcriptions is a good sign that gamification was indeed working. However, gamification is not a complete answer.

Gamification is often used in citizen science projects to motivate and retain interest and participation. However, the gamification aspects are usually focused on competition. Because of this, it is possible to actually persuade participants to stop participating because of the competitive tone. Users of Old Weather were de-motivated when they could not keep up with the highest scorers, acquiring and retaining the rank of Captain was stressful, participants felt that the quality of the work may suffer (another problem with crowdsourcing such as citizen science) since people were simply racing through logs to be Captain, some also viewed competition as opposite from the actual objectives of the project. Although some participants appreciated and enjoyed the competitive nature of the project, many gamification schemes focus only on competition and end up alienating a portion of their population.

B. Alternatives

Having noticed this trend of gamification in citizen science the article *Competing or aiming to be average?,* addresses these trends. Specifically, it looks at normification as a means of engaging digital volunteers looks more at the use of competition as a deterrent to initial or continued participation. It also talks about normification, or fitting in with others in a community as a method of eliciting participation [24]. While competitive gamification may be useful for motivating some individuals, it is often the case that citizen science and crowdsourcing in general attempt to attract as large an audience as possible, meaning that a larger number of occasional participants may be better than a few highly active ones. Turning to psychology, there are several theories indicating that a person is influenced to do or not do certain things based on what others around him or her are doing or what opinions those people might develop. The use of social norms has been used before in CSCW and HCI research. The results of several studies indicate that, in a social setting, the actions and behaviors of people who have a competitive nature would be likely to drift towards the norm, especially if they are below the norm [24]. This is significant as a social environment may provide the benefits of competition without its disadvantages. While norms and normification have been used and studied in CSCW research, norms have not received much use in actual crowdsourcing projects.

A citizen science project called Close the Door, which tasked participants with using a mobile app to determine if shops in their city kept their doors open or closed in cold weather, had some interesting results. When referring to competition amongst other participants, the gamification concept of competition had the usual effects of motivating some and de-motivating others. Some participants competed with themselves; essentially trying to continually do better than they had before while not caring about the progress of other participants. When examining normification, the study found that some participants were spurred on by looking at the leaderboard. These participants would either: try to stay with the pack; or, if they were relatively low scorers, would at least attempt to use the app, even if they were not really concerned with comparing scores [24].

Depending on the needs of the project (large number of average performing volunteers versus small number of high performing volunteers), a combination of normification and gamification techniques can be better than simply gamification.

C. Positive Aspects of Gamification Techniques

Gamification is not all bad. Although gamification, specifically competitive gamification, can be a de-motivating factor for some citizen scientists, gamification is still a beneficial method for engaging citizen scientists. Adding to the traditional definition of gamification (see [7]), the article, Using gamification to inspire new citizen science volunteers, points to gamification as a means to support the user's act of creating content and something that stays true to the intrinsic goals of the system. Essentially, gamification should be used with the idea in mind that gamification should enhance interest in the “domain of interest” [1]. People who use such gamified apps do so because it is: enjoyable, leads to social interactions, can cause the formation of long term relationships, provides a feeling of membership in a community, can lead to the discovery of new things, a way to compete, or a way to better oneself [1]. There is also much overlap between the motivations of citizen scientists and people who enjoy playing games, leading to the logical conclusion that gamification can work for citizen science.

The participants of a study outlined in the article were what are called millennials, or Americans born after the year 1980. These millennials are very likely to use technology, including video games, and often find technology to be positive and beneficial. A gamified citizen science app called *Biotracker*, that provided information on plants to Project Budburst, was used by the participants. The participants later reviewed their experiences with the app.

Many participants who were likely to use the app noted that fun was a major factor. Although the participants were not more motivated to use the app because of an existing interest in plants, the participants were motivated to gain knowledge about plants. Contributing to science and the public good were not strong motivating factors, indicating that personal enjoyment is. Also, community involvement and socialization were strong factors. Some gamification factors were considered significant, yet for participants who were interested in participating in other citizen science projects, the gamification factors were unimportant.

The results of this experiment indicates that gamification can be used to motivate millennials to participate in citizen science projects, even if the participants are not typically interested in citizen science. A fun app that provides social interaction and the ability to learn about that which is being observed are major factors contributing to a successful gamified citizen science app.

While gamification has been used extensively in citizen science applications to encourage and retain volunteers, it has also been shown that some aspects of gamification, namely competition, can put off potential participants. Also, individuals who are already interested in citizen science, as opposed to those who are drawn to the gamification aspects, often find that gamification can be a distraction at best. In a study of a gamified mobile app called Floracaching, outlined in *Gamifying Citizen Science: A study of two user groups*, users were separated into two groups: gamer participants and nature participants. In the app, floracaches, similar to geocaches, are specific sites where a plant is located. Users visit a floracache and check in, where they can then provide either a picture or text information about the plants located there.

Both user groups enjoyed aspects of gamification, such as a "treasure hunt feeling" [2] and visiting unique places. Interestingly, both nature and gamer participants were motivated to learn about the plants. However, it is not completely clear if the interest by gamers was due to the fun nature of the app, or if the gamer participants were simply open to learning about nature anyway. This is a different result from other such experiments as gamers, or those who are not particularly interested in citizen science, tend to use the application in a purely game-like manner and are not particularly interested in learning on the whole.

Another interesting fact is the nature participants were interested in the competitive part of the app: the leaderboard. However, these participants would appreciate competition for accuracy, such as most accurately identified plants, instead of simply "getting the fastest" [2]. This is also interesting as it runs contradictory to other findings. The article mentions that the fifty eight total participants evaluated prototypes of the app on university campuses. It does not, however, indicate the age range of the participants (were they actually college students?) So, these results may be skewed somewhat if the participants are college students, as they would be considered millennials (see [1]) who are much more apt to use technology and play video games, which are most often competitive.

The experiment also produced a third important result, that social interaction supported by the app would be positive for all users. This would be particularly helpful for gamers since they could communicate with individuals possessing more domain knowledge when completing somewhat more complicated tasks. While gamers believe that the app should also add more value, such as teaching their kids proper data collection methods, all groups walked away with positive experiences [2].

D. *Other Concerns*

In citizen science, a major question is how can citizens submit information to scientists when they may be located in completely different geographic areas [6]? The Internet has answered this question nicely. Yet the Internet is an ever changing entity, and new methods and tools of development mean that it possesses a wide range of customization. HCI 2.0? Usability meets Web 2.0, points out these benefits, but also some disadvantages that are rapidly becoming apparent.

The web is a place that is built by its users. With so much customization and with so many users, how can a website conform to the usability needs of all the potential visitors? This problem is very important for citizen science in particular, as the work citizens are doing may not be the most exciting. Creating interfaces and interactions that are viable for a range of users, while being easy enough for citizens to enter data while minimizing the time taken, yet still conveying everything that is needed by scientists seems to be a monumental task.

The popularity and importance of the web is not a fad or phase, it will likely be popular for the foreseeable future [6]. Understanding how to create effective methods of interaction for the web, and in particular citizen science, is a major issue for HCI.

While the Internet is crucially important to citizen science projects, there is a considerable amount of under-utilized technology that can be of great benefit. The secondary contribution of the paper, "*Sensr*: Evaluating a flexible framework for authoring mobile data-collection tools for citizen science," a literary review of design considerations for citizen science projects, has some interesting facts regarding this matter. In citizen science, it is often difficult for the citizens to participate in person with the scientists and the Internet is often used to facilitate such participation. While many citizen science projects make use of the Internet, may fail to make use of the mobile device as a collection mechanism. As of September of 2012, forty five percent of American adults and sixty six percent of young adults own a smartphone [15]. In a review of three hundred and forty citizen science projects, only eleven percent had a mobile application. Fifty three percent used websites for primary data submission, but those relied on other web services such as Survey Monkey [15]. Location meta-data was commonly used by these projects and images with text or data was also important, yet obtaining this data can be tedious if the citizen needs to carry his or her laptop around to collect said data. The article concludes by stating that mobile applications could improve participation in citizen science, in particular for younger individuals.

While citizen science is about interacting with nature, that does not mean technology cannot play an important role. Without the Internet, it would be very difficult for citizen science projects to succeed today. Citizen science has embraced technology to help further its goals, but technology is always changing, and citizen science needs to change with it. Adopting and designing for smartphones is the next big step to make collecting data easier than ever before. Looking at new ways to motivate and sustain participation, engaging a broader audience, etc.; all of these questions need to be answered.

E. *Looking Towards the Future*

The best way to forge ahead is to look back. An overview of the main themes of sustainable HCI, Mapping the landscape of sustainable HCI looks at what is currently established and what potential or emerging issues are of importance to this field. The hope is that by examining the different aspects and perspectives of sustainable HCI, the groundwork for future advances can be laid.

A primary area of focus in sustainable HCI is persuasive technology [5]. In this case, the goal is to influence users to change their behavior by either explicitly or implicitly providing information to the user that shows sustainable or unsustainable behavior. Ambient awareness is very similar to persuasion [5]. It again focuses on sustainable behavior, but in this case it is less directed (hence ambient). However, it seems that these two genres are identical in all but name. Sustainable interaction design attempts to rethink the use of HCI methods in regards to sustainability [5]. The predominant aim here is to make the actual design process more sustainable, and to reduce pollution and wasted resources. The work in this area tends to be more theoretical and critical. Formative user studies look at how users view sustainability instead of how the designers view it [5]. The body of work tends to look at social and cultural constraints that limit a person's ability to change his or her behavior. The final genre of work is pervasive and participatory sensing, commonly shortened to citizen science [5]. In this case, non-experts collect data which is sent to scientists and researchers who evaluate the data. The data tends to be more diverse than what could have been collected by the scientists themselves because of the potential for significant participation.

Before diving into potential future issues, the paper makes a good point that sustainable HCI has redundancies that hinder its advancement. The field needs to "take stock" [5] of the current research and investigate what questions are answered and what are raised by the research. Future issues include using the expertise of other fields, such as engineering, to come up with solutions as the two fields often work towards similar solutions yet are unaware that the other is doing so. The need for debate between different approaches is also important, as there are plenty of topics to debate and debate may reveal other issues that can be tackled. Also, looking to different disciplines can provide better or further analysis of existing solutions [5].

VII. GEOTAGGER: A COLLABORATIVE AND PARTICIPATORY ENVIRONMENTAL INQUIRY SYSTEM

In the previous sections, we have featured how Big Data is related to sustainability science, citizen science, crowdsourcing and the implications of these when we consider them in the HCI context. In this section, we introduce GeoTagger, which incorporates all of these concepts to address solutions for bringing populations with varying backgrounds into an engaged interaction on citizen science and sustainability sciences activities.

A. *The Sustainability Studies Mobile Toolkit Framework*

The rapid growth of sustainability science rose due to the ecological and climate issues recently. There might be a noticeable shift in the weather which can cause severe storms according to the different weather models. The field of sustainability puts attention decision making process, data collection, integration, etc. which might help the ecological concerns. Sustainability Studies Mobile Toolkit (SSMT) is a project that addresses the needs to develop facilities for collection, analysis and decision making of data for ecological, environmental and biological scientist, where they can collaborate on sustainability issues primarily through mobile or tangible devices. In this project, the visualization of data can help the scientist to design and display analytics and understand their work and its implications better via a tangible interface device. It also gives an option for a citizen scientist with average understanding to the field to get involved if the researcher is willing to do so. Some of the challenges that the ecological collaborators faced to get used to the technology where as follows: For the collaborative exploration of the experiment and a zero-time learning curve, the technology is expected to be flexible and easy enough to understand by a non-technical individual. Scientist should be able to control the distribution of data and provide security to the data; reliable data collection can make it happen. One of the challenges faced was also that the scientific data may have different form; there could be a possibility of the difference of opinion between the scientist and the interested stakeholders [13].

B. *GeoTagger*

The technology is very important aspect and large impact in the present day society not only with adults but also children. About 66% of children aged 8-18 old in United States have cell phones and also spend 7.5 hours interact with the entertainment media every day. With the estimated number the alarming increasing; there is a decrease in interacting with the nature [3,9,11,20,23]. We need a technology which encourages children to come together and also explore the surroundings outdoor. The technology should be able to bring children together to share, analyze and create with their friends or their peers. Now-a-days the use of technology has been taken to whole new level. The user who use the computer are at an age group which varies from children to adults for various activities which involves games, social interaction, school or work related assignments.

Coupled with the ideas in the SSMT and work from [9], the GeoTagger collaboratory environment was created [10]. The GeoTagger mobile application is approach to bridge this gap especially the children in between the nature and technology. It differentiates itself from various citizen science approaches with limited data collection and analysis; providing social interactions about the observations they made. Using GeoTagger the participants can tag their interested observations about their surroundings and real world things and view what others are interested in, followed by a conversation about their interests. This can be very useful for a specific purpose, like collecting data about any specific plants or anything else in the habitat, treasure hunts, etc. as the tags can be configured. One key aspect too of this tool is that since its design is highly collaborator with special attention paid to the HCI aspects, it enables users, especially children, to participate in the field. Therefore, it presents a flexible mobile application that can allow interactive data collection and share activities to occur in real time, with little restriction on the age appropriateness or education level of the user. The different activities supported by GeoTagger are actively engaging the users in collaborative, constructive and generative activities. For children especially, it works in concert with activities that help the social and cognitive development of the children, engaging children through peer discussions, competitions and collaborations which benefit the development and learning for them. It provides a platform where a mobile, cloud based system can support environmental enquiry and collaborative learning.

The design process initially involved children and adults work together iteratively as design partners. There was lot of brain storming sessions over past year which includes low-tech prototyping, in field explorations with paper and medium fidelity prototypes, and sticky note activities. The goal was very clear that the tags should be fun and should share tags with friends, peers and groups. Geotagger involves the following aspects:

System Overview: The application of the system includes profiles of the users and tags adventures and groups. The important aspect of the system is the social interaction with the friends or collaborators.

Profiles: To be able to use the application the user has to create an account. Each user has to provide minimal information to login into the account.

Tags: To use this feature the user has to login to the application with the user name. This feature enables the user to add a name, description, GPS location and also the location or the facets that can be searched. Tags can be edited or deleted only by the user. Friends or collaborators can only comment on the tagged items.

Friends or collaborators: The user can view and comment on their friends' tag items which highlights the social interaction of the application[10].

The project has come a long way from forming a basis to the system to motivate the users to use the application especially children. The Geotagger is now available – at least in part – on several platforms including the web, and Android and iOS mobile devices. Future work includes making sure the Geotagger data and inquiry-based platform is available and fully functional in the three main platforms: web and mobile (iOS and Android). Future work will include more in-depth analysis of its effectiveness and usage patterns.

VIII. CONCLUSION

The wave of big data has just begun and it will have a great impact on everything in this universe. The power of Big Data is yet to be uncovered. However in this growing era of Big Data it is necessary to think that this data can be applied for the environmental, ecological and biological issues which can be useful to the mankind. The GeoTagger work contributes to the major areas of Sustainability Science and citizen science. Our survey points towards how efficiently the technology can be used for the ecological researches and also the citizen scientist to make their work easier. We closely look into how to transcribe, collect and analyze the Big Data in these fields for the development of some technology which can be useful. GeoTagger is one of the examples mentioned which implements collaborative learning and environmental correlation with the technology for the children. The science which is enhanced technically has the capacity to expand in engaging the people in the real world problems and issues and inspire them to be a part to find a solution for the same. Looking towards the future we hope to develop tools for some of the sustainability issues of the environment and consider bridging the gap between the science and technology.

IX. ACKNOWLEDGEMENTS

We would like to acknowledge the PSE&G Institute for Sustainability Science for funding the SSMT and GeoTagger Research. We would also like to recognize Drs. Jennifer Bragger, Paul A.X. Bologna, Jennifer Krummins and Kirsten Monsen, Karina VR Schafer, and Dirk Vanderklein for their feedback and participation.

X. REFERENCES

[1] A. Bowser, et al. "Using gamification to inspire new citizen science volunteers," in *Proceedings of the First International Conference on Gameful Design, Research, and Applications*, 2013, pp. 18-25.

[2] A. Bowser, D. Hansen, J. Preece, Y. He, C. Boston, J. Hammock. "Gamifying Citizen Science: A Study of Two User Groups," in *Proceedings of the companion publication of the 17th ACM conference on Computer supported cooperative work & social computing*, 2014, pp. 137-140.

[3] C. Cordes and E. Miller. "Fool's gold: A critical look at computers in childhood. Alliance for Childhood.", Internet: http://waste.informatik.hu-berlin.de/diplom/DieGelbeKurbel/pdf/foolsgold.pdf, 2000 [Oct. 27, 2014].

[4] E. Damiani and G. Sissa. "An agent based model of environmental awareness and limited resource consumption," in *Proceedings of 5th Inernational. Conference Management of Emergent Digital EcoSystems - MEDES '13*, 2013, pp. 54-59.

[5] C. DiSalvo, P. Sengers, H. Brynjarsdóttir. "Mapping the Landscape of Sustainable HCI," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 2010, pp. 1975-1984.

[6] A. Dix and C. Cowen. "HCI 2.0? Usability Meets Web 2.0," in *Proceedings of the 21st British HCI Group Annual Conference on People and Computers: HCI…but not as we know it*, 2007, pp. 185-186.

[7] A. Eveleigh, C. Jennett, S. Lynn, A. Cox. "I want to be a Captain! I want to be a Captain!: Gamification in the Old Weather Citizen Science Project," in *Proceedings of the First International Conference on Gameful Design, Research, and Applications*, 2013, pp. 79-82.

[8] A. Eveleigh, et al. "Designing for Dabblers and Deterring Dropouts in Citizen Science". *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 2014, pp. 2985-2994.

[9] J. Fails, A. Druin, M. Guha. "Mobile collaboration: collaboratively reading and creating children's stories on mobile devices." *Proceedings of the 9th International Conference on Interaction Design and Children*, 2010, pp. 20-29.

[10] J. Fails, et.al. "GeoTagger: A Collaborative and Participatory Environmental Inquiry System," in *Proceedings of the companion publication of the 17th ACM conference on Computer supported cooperative work & social computing*, 2014, pp. 157 – 160.

[11] V. Rideout, U. Foehr, D. Roberts. "Generation M2: Media in the Lives of 8- to 18-Year-Olds." Internet: http://kff.org/other/event/generation-m2-media-in-the-lives-of, Jan. 20, 2010 [Nov. 15, 2014].

[12] D. Helbing, "New Ways to Promote Sustainability and Social Well--Being in a Complex, Strongly Interdependent World: The FuturICT Approach" Internet: http://arxiv.org/ftp/arxiv/papers/1310/1310.3498.pdf, Oct. 13, 2013 [Nov. 17, 2014].

[13] K. Herbert, et al. "Scientific Data Infrastructure for Sustainability Science Mobile Applications," in *Proceedings of Big Data (BigData Congress), 2014 IEEE International Congress*, 2014, pp. 804-805.

[14] T. Hewett, et al. “Human-Computer Interaction” in *Acm SIGCHI Curricula for Human-Computer Interaction*. B. Hefley, Ed. New York: ACM Inc., 1992, pp. 5-27.

[15] S. Kim, J. Mankoff, E. Paulos. “Sensr: Evaluating a Flexible Framework for Authoring Mobile Data-Collection Tools for Citizen Science,” in *Proceedings of the 2013 conference on Computer supported cooperative work*, 2013, pp. 1453-1462.

[16] M. Laituri and F. Sternlieb. Water Data Systems: Science, Practice, and Policy.” *Journal of Contemporary Water Research & Education*, vol. 153, pp. 1-3, July 2014.

[17] S. Levin and W. Clark. “Toward a Science of Sustainability.” Internet: https://www.nsf.gov/mps/dmsdocuments/SustainabilityWorkshop2009Report.pdf, Nov. 29, 2009 [Nov. 3, 2014].

[18] R. Louv. *Last Child in the Woods: Saving Our Children from Nature-Deficit Disorder*. New York: Algonquin Books of Chapel Hill, 2008.

[19] E. Orts and J. Spigonardo. “Sustainability in the Age of Big Data.” Internet: http://d1c25a6gwz7q5e.cloudfront.net/reports/2014-09-12-Sustainability-in-the-Age-of-Big-Data.pdf, Sept. 2014 [Nov. 15, 2014].

[20] Outdoor Foundation. “Outdoor Recreation Participation Report, 2008.” Internet: http://www.outdoorfoundation.org/pdf/ResearchParticipation2008.pdf, 2008 [Oct. 31, 2014].

[21] S. Park, et al. “Crowdsourcing Micro-Level Multimedia Annotations: The Challenges of Evaluation and Interface.” in *Proceedings of the ACM multimedia 2012 workshop on Crowdsourcing for multimedia,* 2012, pp. 29-34.

[22] R. Phillips, J. Blum, M. Brown, S. Baurley. “Testing a Grassroots Citizen Science Venture Using Open Design, The Bee Lab Project.” in *Proceedings of CHI '14 Extended Abstracts on Human Factors in Computing Systems*, 2014, pp. 1951-1956.

[23] J. Preece, A. Bowser. “What HCI can do for Citizen Science.” in *Proceedings of CHI’14 Extended Abstracts on Human Factors in Computing Systems*, 2014, pp. 1059-1060.

[24] C. Preist, E. Massung, D. Coyle. “Competing or Aiming to be Average? Normification as a Means of Engaging Digital Volunteers.” in *Proceedings of the 17th ACM conference on Computer supported cooperative work & social computing*, 2014, pp. 1222-1233.

[25] N. Prestopnik, K. Crowston. “Purposeful Gaming & Socio-Computational Systems: A Citizen Science Design Case.” in *Proceedings of the 17th ACM international conference on Supporting group work,* 2012, pp. 75-84.

[26] M. Rijmenam. “A Short History of Big Data.” Internet: https://datafloq.com/read/big-data-history/239, Jan. 7, 2014 [Sept. 8, 2014].

[27] A. Wiggins. “Technology and Work Practices in Citizen Science.” in *Proceedings of the 73rd ASIS&T Annual Meeting on Navigating Streams in an Information Ecosystem,* 2010, vol. 47, article no. 158.

[28] M. Sabou, K. Bontcheva, A. Scharl. “Crowdsourcing Research Opportunities: Lessons from Natural Language Processing.” in *Proceedings of the 12th International Conference on Knowledge Management and Knowledge Technologies.* 2012, Article No. 17.

[29] “New York Metropolitan Area.” Internet: http://en.wikipedia.org/wiki/New_York_metropolitan_area [Jan. 29, 2015].

[30] C. Dybas. “In wake of Hurricane Sandy, Oklahoma tornadoes, NSF awards $32 million in hazards sustainability grants.” Internet: http://www.nsf.gov/news/news_summ.jsp?cntn_id=129307, Oct. 21, 2013 [Jan. 29, 2015]