4-7-2022

Instagram of Rivers: Facilitating Distributed Collaboration in Hyperlocal Citizen Science

Srishti Gupta

Julia Jablonski

Chun-Hua Tsai

John Millar Carroll

Instagram of Rivers: Facilitating Distributed Collaboration in Hyperlocal Citizen Science

SRISHTI GUPTA, Pennsylvania State University, USA
JULIA JABLONSKI, Pennsylvania State University, USA
CHUN-HUA TSAI, University of Nebraska Omaha, USA
JOHN M. CARROLL, Pennsylvania State University, USA

Citizen science project leaders collecting field data in a hyperlocal community often face common socio- technical challenges, which can potentially be addressed by sharing innovations across different groups through peer-to-peer collaboration. However, most citizen science groups practice in isolation, and end up re-inventing the wheel when it comes to addressing these common challenges. This study seeks to investigate distributed collaboration between different water monitoring citizen science groups. We discovered a unique social network application called Water Reporter that mediated distributed collaboration by creating more visibility and transparency between groups using the app. We interviewed 8 citizen science project leaders who were users of this app, and 6 other citizen science project leaders to understand how distributed collaboration mediated by this app differed from collaborative practices of Non Water Reporter users. We found that distributed collaboration was an important goal for both user groups, however, the tasks that support these collaboration activities differed for the two user groups.

CCS Concepts: • Human-centered computing → Empirical studies in collaborative and social computing; Empirical studies in collaborative and social computing.

Additional Key Words and Phrases: distributed collaboration, citizen science, sustainability, collaboratory

ACM Reference Format:

Srishti Gupta, Julia Jablonski, Chun-Hua Tsai, and John M. Carroll. 2022. Instagram of Rivers: Facilitating Distributed Collaboration in Hyperlocal Citizen Science. *Proc. ACM Hum.-Comput. Interact.* 6, CSCW1, Article 97 (April 2022), 22 pages. https://doi.org/10.1145/3512944

INTRODUCTION

The world's water resources are in crisis due to human activities and global warming. Countdown to 'Day Zero' in Cape Town, South Africa [44], lead contamination in Flint, Michigan [9], and groundwater loss in northern India [3] are some vivid examples of recent water crisis around the world. In the last few decades, citizen-based approaches to ensure adequate local supplies of good quality water have become a popular way of monitoring and ensuring water quality. Greater citizen-based volunteer activity in local water systems is associated with better water quality [26]. Citizen-based monitoring activities are termed as *citizen science*, where citizens of a community participate in scientific endeavors in various capacities [41]. Citizens may participate as volunteers assisting scientists in activities such as data collection, transcription, categorization, and analysis, or may work as part of a community hobby group, enthusiastic about conserving their natural resources and environment [7]. The practice of citizen science has especially proliferated and gained traction in environmental science where researchers need to collect large volumes of data over a wide geographical area [51].

Citizen science project leaders and volunteers often face common challenges such as uncertainty about citizen science data quality [14], retention of participants in projects [60], and lack of appropriate technical infrastructure [46]. Innovations that address these socio-technical challenges can be shared across different citizen science groups through peer-to-peer collaboration. However, most citizen science groups practice in isolation, and end up re-inventing the wheel when it comes to addressing these common challenges [53]. Encouraging distributed or peer-to-peer collaboration between different citizen groups can potentially streamline citizen science activities by ensuring that groups don't spend unnecessary resources and time on addressing problems, that can be otherwise easily resolved by sharing and exchanging information and resources. Moreover, lack of appropriate technical infrastructure to support such distributed collaboration can potentially impede sharing and exchange of valuable information and practices thereby hampering unanticipated mutual benefits such as informal learning.

This study seeks to investigate distributed collaboration between different water monitoring citizen science groups. Citizen science is an inherently collaborative venture, where collaboration between volunteers and scientists, and amongst volunteers of a particular citizen science team has been studied extensively [42, 47]. Moreover, literature on distributed collaboration in citizen science largely refers to online citizen science programs, where distributed teams collaborate on an online platform [61], or volunteers collecting data are dispersed across a wide geographic area [62]. In this study, we define distributed collaboration in citizen science as peer-to-peer collaboration between dispersed citizen science groups. These citizen science groups can be non-profit organizations, part of academic institutions, or civic organizations. These groups can be either located in the same region (for example, in one state), or in widely different regions (for example, east coast and west coast).

We surveyed 50 water data and learning tools to identify what tools already support, or could support citizen science practices. Through this survey we discovered an application called *Water Reporter* [1] with a unique social network design to support citizen science watershed initiatives. This was unique because this was the only design in the survey that could possibly facilitate distributed collaboration between citizen science water monitoring groups. Even though distributed collaboration is theoretically beneficial for hyperlocal citizen science, lack of literature on such collaboration practices in CSCW and HCI makes it valuable to be able to study how technology mediated distributed collaboration can be beneficial. Therefore, we decided to interview citizen science project leaders who were Water Reporter users to understand their collaborative practices, and how this application supported, or could potentially support these practices. We also interviewed citizen science project leaders who were Non-Water Reporter users to understand their collaborative practices as well, and then compared the two sets of interviews to understand (1) goals and benefits of distributed collaboration, and (2) how appropriate technological infrastructure can support and improve distributed collaboration.

Our interview data analysis found that distributed collaboration was an important goal for both Water Reporter and non Water Reporter users, however, the tasks that support these collaboration activities differed for the two user groups. Water Reporter users were aware of both the collaborative activities, and tasks in the platform that could support those activities. Hence, Water Reporter *mediated* distributed collaboration by creating *visibility* and *transparency*. The technology mediation allowed users to become aware of one another's activities [34], discover

[1] https://www.waterreporter.org/

other citizen science groups and make their work visible [52], and allow users to see activities as they occur, making it more transparent. Hence, users became more aware of one others practices. On the other hand, non Water Reporter users were aware of the collaborative activities but not the tasks because their platforms did not have features with tasks to support the activities. Hence, we use the Water Reporter application and its distinctive collaborative features, to compare how technology can mediate distributed collaborations more efficiently and help ameliorate citizen science practices. The analysis of our findings provides two meaningful contributions to CSCW and HCI: (1) we describe how visibility and transparency mediated by technology can create awareness and facilitate distributed collaboration in citizen science, and (2) we suggest a collaboratory design for citizen science to support different collaborative activities.

STUDY CONTEXT - OVERVIEW OF WATER REPORTER APP

Water Reporter is a watershed monitoring application with a social network design to connect and support watershed initiatives. The social network interface design is similar to Instagram, where users can post pictures of their watershed, and interact with 'like' and 'comment' features on posts (Fig. 1.). However, it has additional functionalities that support water monitoring activities, that distinguish it from traditional social media. For instance, citizen science groups can create 'campaigns' which is a monitoring activity that allows volunteers to contribute data to a particular initiative. The data in this case would be a geo-referenced picture, that is usually ascribed with a data form. For example, a volunteer can post a picture of a polluted stream to a 'pollution tracking' campaign and fill out the corresponding data form that contains parameters like temperature, flow, pH etc. Another prominent feature of Water Reporter is its data mapping function. It creates aesthetically pleasing and simple data maps and simple visualizations that makes it easy for volunteers to see how their data is being used. This allows groups to provide quick acknowledgement to their volunteers about their contribution.

The social media design can potentially allow the users to explore a watershed and observe action being taken in their community. Users can search by individual, organization, hashtags, or watershed name (Fig. 2.). It can connect individuals and citizen science groups together. The simple social media design of the app also makes it an excellent tool for data collection, as it doesn't require extensive training for volunteers. Most WR participants described the app as “Instagram for Rivers”. However, our interviews with Water Reporter users also revealed some challenges and limitations of the application. For instance, even though the social media feature of Water Reporter was perceived unique and intriguing, most users reported limited use of this features. One reason for this limited use was problems with the design itself. Even though it resembles the user interface of Instagram, users don’t have control over what feeds they see. In addition, people have been socialized to use traditional social media like Facebook and Instagram a certain way, and that is different from how people perceive Water Reporter’s social media functions. Water Reporter users mentioned how they want to take measures to encourage their volunteers to get accustomed to using the social media feature. Regardless of limitations and design challenges, all Water Reporter users considered the app to be user-friendly and proficient in both data collection and facilitating social interactions. Even though the application was not specifically geared towards enabling distributed collaboration, we found that users appropriated the social media design to collaborate in different ways. This unintended appropriation of the design provided unanticipated benefits for the users.

LITERATURE REVIEW

Citizen science is a research technique where citizens of a community participate in scientific endeavors. However, public participation in science is not a recent phenomena. Citizens have been contributing to science for centuries, by collecting observational data to understand the natural world around them. Professionalization of science excluded the knowledge of these ‘amateur’ citizens. Collaboration between citizens and scientists in citizen science has allowed citizens to add their knowledge to institutional science [16]. More recently, citizen science has become a popular method of collecting large amounts of data in fields like environmental sciences, archaeology and astronomy [51].

In the case of water systems, water quality monitoring is a technical activity that can more broadly strengthen the local community. Watershed systems become measurably stronger with greater citizen-based water quality monitoring activity. In a study of 2,150 US watershed systems, Grant and Langpap [26] showed that the quality of local water is positively correlated with greater activity of citizen volunteers. Given the current global water crisis, it is timely to ask how we can make current citizen-based water monitoring activities measurably stronger.

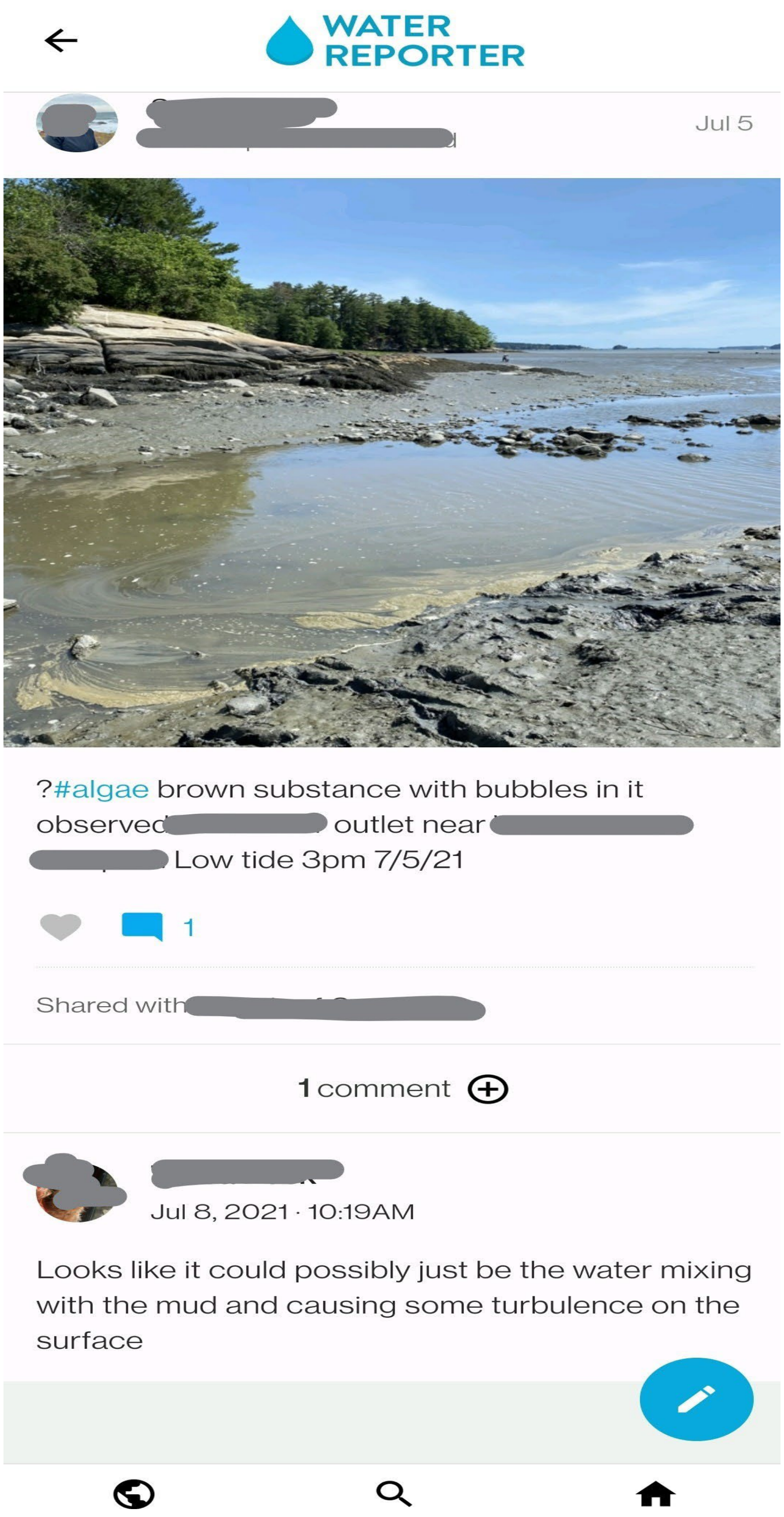

Fig. 1. Social Media Interaction on Water Reporter

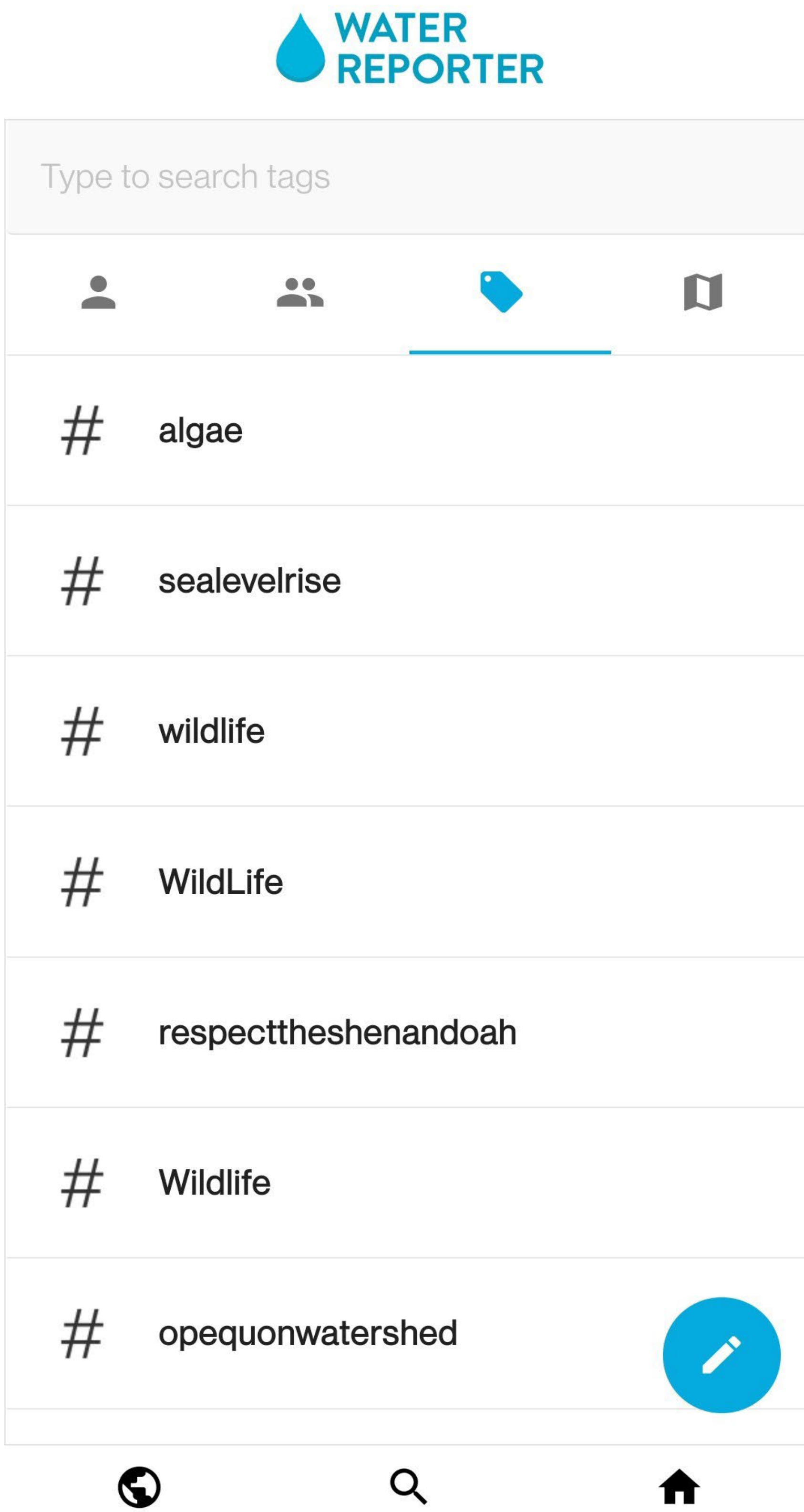

Fig. 2. Search Function: User can search by Name, Organization, Hashtag, or Watershed

Collaboration in Citizen Science

In citizen science, volunteer collaboration with professional scientist can be of three types: (1) *contributory* projects in which volunteers contribute data to scientists (2) *collaborative* projects, where volunteers assist scientists in project design, in addition to data collection (3) *co-created* projects where volunteers and scientists are equally involved in all parts of the project [6]. Most of the citizen science projects fall under the category of contributory projects. Instances of collaborative and co-created projects remains low. Volunteers usually collaborate with scientists by collecting data or doing partial analysis such as documentation and annotations [49].

Current research in citizen science has primarily focused on two types of collaboration: (1) collaboration between volunteers and professional scientists, and (2) collaboration among volunteers of a citizen science group.

Since volunteer-scientist relationship forms the crux of the citizen science project, research in exploring these relationships is important in sustaining citizen science projects. As mentioned above, most citizen science projects are contributory in nature. Contributory projects do not particularly seek to promote collaboration because of the rudimentary nature of volunteer participation. Hence, infrastructure focus remains on scientists to help them improve their activities. In order to promote more collaborative and co-created projects it is imperative to understand the motivations and needs of both volunteers and scientists. This will help design an environment that catalyzes collaboration in a constructive way [49]. Focusing on such collaboration is important because it determines the nature of public participation in science and success of the project. Shirk et al. [50] created a model of public participation in science that demonstrated that regardless of the type of citizen science project, its success depends on (1) degree of public participation, and (2) quality of public participation. Cox et al. [17]'s analysis of Zooniverse projects also corroborates this claim. Their findings indicated that successful citizen science projects depended on strength of scientific impact and public engagement.

Research in HCI has also identified the significance of volunteer-scientist relationship, and has proposed design directions and systems that can empower volunteers, and improve volunteer-scientist collaboration. Oliveira et al. [47] conducted an empirical study to understand the importance of volunteer-scientist collaboration, and design directions to improve this collaboration. Their research revealed the importance of designing systems that can provide more agency and active involvement in design and execution of projects, and facilitating interaction between volunteers and scientists to maximize both scientific and social benefits of citizen science. Hsu et al. [31] developed an air quality monitoring system as a collaborative project with the local community. The collaborative design process led to the development of a system that provided strong scientific evidence for the community to successfully fight with the local government to shut down the local coke plant. The system design also facilitated informal learning that led to more rational discourse amongst all the stakeholders involved.

Another type of collaboration discussed in the citizen science literature is among members of a citizen science team. The advantages of citizen science go beyond data collection. Citizen science is also a growing paradigm of community engagement, that can create a sense of community, and enhance informal learning and support for science[21]. It plays an important role in public engagement with science, which in turn builds social capital, cultivates scientific literacy, and overall helps build a more resilient community. However, contributory model of citizen science often fails to engender these social benefits. Public engagement in science requires more than just data collection. Volunteer involvement in scientific inquiry is important to meet these social goals. Research in the HCI community has focused on designing platforms that can support collaboration among volunteers to enable engagement with scientific inquiry. Luther et al. [42] designed Pathfinder, a wiki-based collaborative platform for citizen scientists to analyze and discuss the data collected. Willett et al. [63] designed a similar collaborative platform called "Common Sense Community Site" to enable collaborative analysis of air quality data. This platform can accommodate larger datasets as compared to Pathfinder. Additionally, Willett et al. [64] designed a collaborative visual analysis tool called CommentSpace, that uses links and tags to identify and organize information.

Social Media and Citizen Science

Social media platforms are increasingly becoming an important part of scientific enquiry. These platforms provide an efficient way for scientists to reach out to communities for participation in research studies, use the interactions to study human behavior, use data for different kinds of analysis, or use it to facilitate scientific communication [29]. Citizen science has also started leveraging social media platforms for various purposes such as volunteer recruiting, community engagement, education, and social interactions [1]. The familiarity of using social media has also facilitated citizen science project leaders to design social media style platforms for collecting observational data. Using such platforms for data collection requires minimal training, and can be used to engage a wider audience in data collection process. For instance, social media apps have been developed to track urban flooding [58]. Such apps allow volunteers to post pictures and related meta-data, which can provide huge datasets to scientists for analysis and decision-making at minimal cost. Other emergent use of social media in field-based citizen science, has been to facilitate social interactions amongst volunteers to increase and sustain volunteer engagement. A New Zealand Garden Bird Survey citizen science project developed a Facebook group for its volunteers to interact with each other [40]. The researchers studied the interactions on Facebook through the lens of Community of Practice [59], and demonstrated how social media fostered social interactions that led to sharing of knowledge among volunteers, and helped build a sense of community. Volunteer retention in citizen science projects is a common concern across different projects and domains [60]. Facilitating such interactions can be helpful in motivating volunteers to participate, and promote & sustain community engagement.

Advancement in collaborative technologies has also facilitated large scale citizen science projects on the internet. We now have several online citizen science platforms, where program leaders host projects and tools to collect data. Anyone with an internet connection can join these projects and become a citizen scientist [43]. Out of these platforms, Zooniverse is one of the oldest and biggest citizen science platform with more than 1.6 million registered users and around 40 projects. Although their primary goal was to collect data, they eventually created a discussion space called ‘Talk’ for volunteers and scientists. This discussion space led to an unprecedented rise of new scientific discoveries, as citizen scientists went beyond their usual task of collecting data, and started asking questions giving rise to rational scientific discourse and deliberations [55]. iNaturalist is another popular citizen science platform that aims at collecting data on biodiversity across the globe. The platform is designed to facilitate social interactions like a traditional social media. Hence, its design equally focuses on data collection for scientists, and sharing knowledge about biodiversity in a social network of volunteers [27].

Social media has thus become an online space that can potentially challenge the traditional top-down approach to citizen science, where institutional standards often dictate judgements on quality and accuracy of citizen science data and techniques. This top-down approach causes conflicts between citizen science and expert knowledge and makes power structures more apparent [48]. Traditional citizen science infrastructure is representative of this top-down approach, where various data structures of the system govern quality control and assurance, validity, reliability, and knowledge production. Social media can afford a bottom-up approach to citizen science, by providing more flexibility and fluidity of discussion around data, practices and knowledge production.[28].

Collaboration in Science

Collaboration among distributed and diverse scientific communities has become critical to address global challenges such as climate change and healthcare crisis. In addition, emerging technical infrastructure has produced enormous amounts of data about earth, human bodies, communication, and behavior. This big data has provided an opportunity to conduct revolutionary collaborative scientific research which now requires complex collaborative systems to support this cross-disciplinary and distributed research. Use of distributed technologies or grid to facilitate these scientific collaboration has come to be known as e-science. E-science seeks to promote online collaboration of distributed scientists through an underlying computational infrastructure that enables data sharing, analysis, experiments, instruments etc. [20]. In the United States, NSF funded initiative of e-science came to be known as *collaboratories*. A collaboratory is a laboratory or a centre without walls, that promotes collaboration between distributed scientists across a widespread region, to access, view, manipulate and discuss artefacts [24, 66]. Although these two programs were launched separately, the overarching goal remains the same - to develop and provide necessary infrastructure to support cross-disciplinary and distributed research.

CSCW has a key role in driving these initiatives because of its extensive work and experience with distributed technologies and collaboration. CSCW research has been helpful in understanding some core problems faced by e-science teams such as data-sharing, methodological challenges, and conceptions of collaboration [32]. Edwards et al. [23] analyzed different case studies of interdisciplinary scientific collaboration in the context of metadata. They discovered that metadata was an essential element binding the scientific team and work together. They analyzed its role to better conceptualize and articulate interdisciplinary work practices. For example, their analysis revealed that even though metadata provided a common ground for different experts in the scientific group, it was also a source of friction because of the cost, time and effort that went into managing it. Metadata was a product and a process of scientific communication. This analysis is similar to other scientific collaboration studies that view technological infrastructure as a boundary object that is helpful in facilitating interdisciplinary exchange in e-science teams [45]. In addition to boundary objects, distributed and interdisciplinary collaboration in science also requires efforts to align disconnected elements such as work practices and technologies of experts from diverse disciplines. Anticipation work helps in conceptually framing this effort and has been helpful in understanding and designing organization practices and infrastructure [54].

Distributed collaboration in scientific research is a well-established area of study in CSCW and HCI literature. Even though citizen science work practices and infrastructure are different from professional scientific work, lessons and conceptual frameworks from this literature can be helpful in analyzing and articulating distributed collaboration in field-based citizen science.

Distributed Collaboration in Citizen Science

Distributed collaboration in citizen science has largely been studied in the context of online citizen science programs. Online citizen science platforms like Zooniverse and iNaturalist can sustain large scale citizen science programs with volunteers distributed across different geographical regions. In addition, these platforms also facilitate peer-to-peer collaboration between different citizen science groups [62]. In online citizen science, distributed collaboration also includes collaboration between distributed scientists and volunteers. HCI and CSCW research has focused on studying collaborations facilitated by the design of these online platforms [49].

However, citizen science programs that rely on field-based data collection in a hyperlocal community, often practice in isolation. Their interactions are limited to their existing network of hyperlocal groups. There has been little research on studying peer-to-peer collaboration in field-based citizen science projects. Koehler and Koontz [35] observed that local citizen-based watershed monitoring groups collaborate to monitor and restore local water quality [35]. Carroll et al. [11] also found the prevalence and benefits of peer-to-peer collaboration between co-located citizen science groups in the Spring Creek Watershed.

Hence, research on distributed collaborations in field-based hyperlocal citizen science remains scarce. Benefits of distributed collaboration in scientific communities are well established [5]. Combined with the available evidence of importance and success of distributed collaboration in online citizen science, it is imperative to leverage these benefits for field-based hyperlocal citizen science groups as well. Hence, we conducted an empirical investigation of water quality monitoring groups across United States, to understand importance and goals of collaborations of these groups, and how technology can facilitate distributed collaboration to meet these goals. The comparative investigation between users and non-users of Water Reporter helps in understanding how visibility and transparency afforded by Water Reporter distinguished collaborative activities for the two user groups.

METHOD

We surveyed 50 water data and learning tools. Our goal was to identify designs that could provide socio-technical affordances such as collaboration and informal learning, important for citizen science water quality monitoring practices[11]. The survey and analysis of these tools was systematically done over a period of 2 months where researchers iteratively searched the web, academic resources, and other known water monitoring and citizen science resources such as USGS and citsci.org. The researchers also gathered information on different types of water data and learning tools from experts in the field of hydrology and agricultural sciences. The exhaustive search concluded when we could not identify any new web or mobile applications for water quality monitoring. Each tool was analyzed based on its design rationale, design features, socio-technical affordances, tool type and intended users. This survey analysis is part of another research project which is currently in progress.

The survey helped us identify the application "Water Reporter". This application was unique because it was designed as a social media platform that aimed to connect and support watershed initiatives. Although the design description did not explicitly articulate the app as a collaborative platform, its social media design could be appropriated to facilitate distributed collaboration. Most social media features or platforms are tailored to online citizen science projects since these projects tend to have large scale participation and many distributed teams. Collaborative technology to support field-based citizen science is limited. Hence, Water Reporter offered a unique design to facilitate distributed collaboration between water monitoring groups.

This study is a qualitative investigation of water monitoring citizen science groups' goals of collaboration and how technology can help in achieving these goals. We interviewed users of Water Reporter (WR) to understand how this app had helped facilitate collaboration with other citizen science groups. In addition, we also interviewed other Non Water Reporter (NWR) citizen science groups to understand their goals of collaboration, and how the platforms they used supported these goals. We used purposive sampling [56] to recruit 8 WR participants (P1 to P8) and snowball sampling [4] to recruit 6 NWR participants (P9 to P14). Purposive sampling was used to identify users that were exposed to a design that could facilitate distributed collaboration, and would help us understand what features of the platform afford collaboration. The participants recruited recommended and connected us with other NWR water monitoring groups, and hence through a snowball sampling we recruited NWR users. All 14 participants were involved in collaborative programs and activities on behalf of their group. Table 1 provides the demographic details like gender and title of each participant in their organization. The participants that were Water Reporter users, coordinated citizen science programs on the app on behalf of their group, and other Non Water Reporter participants were also leaders of citizen science programs in their group. The study was reviewed and approved by the researchers' university Institutional Review Board (IRB).

We conducted semi-structured interviews that lasted between 30 and 60 minutes. The overarching objective of the interview protocol for the two sets of participants was the same - to understand their goals of collaboration, and how technology can facilitate these goals. However, for WR participants, we also asked questions about specific design features of the app that could be appropriated for collaborative activities. All interviews were audio recorded with the consent of the participants, transcribed and analyzed using NVivo qualitative data analysis software.

We used thematic analysis to get meaningful insights from data. Thematic analysis is a qualitative data analysis method that provides a systematic approach to get meaningful insights and patterns across a dataset[8]. This method was helpful in identifying and constructing important themes around distributed collaboration in field-based citizen science. The goal of this research was to understand whether technology can mediate distributed collaboration between field-based citizen science groups. However without a tangible artefact that is specifically designed for this activity, it becomes difficult to deductively assess its usefulness. Water Reporter's design created visibility and transparency for its users to become aware of other citizen science activities in different regions. Hence, our assumption was that users of this app would have more exposure to technology mediated collaboration, than users who don't use this app. Therefore, the research aimed to discover if Water Reporter aided its users collaborative activities, and compare that with groups that did not use the platform.

For data analysis, two researchers independently coded small portion of the data, with the goal of evaluating and revising the codebook to more appropriately fit the research questions. Initial rounds of coding were used to analyze the data to find all relevant citizen science activities, tasks and interactions. We inductively analyzed and labelled similar codes in the data with low-level category labels. These low-level categories were further analyzed and scaffolded to more abstract categories. Finally, analysis of abstract categories helped discover major cohensive themes in the data. For instance, one category that emerged from the initial round of coding was *citizen participation*, from low-level categories *recruiting volunteers, types of participation, challenges with participation, and keeping volunteers engaged*. We provided brief descriptions of each low-level category, number of associated datasets, and numbers of references corresponding to each low-level category.

After the initial coding process, we iteratively examined the categories to construct themes in the context of collaboration. For instance, the example of the more abstract category and low- level categories above evolved into the theme of *community engagement*. As themes were further developed, we started comparing WR and NWR data on the basis of these categories. Researchers reviewed and compared each others codebook for a small number of interviews, and then repeated the procedure for the rest of the interviews until they reached a consensus. Although the first author lead the data analysis process, the entire research team closely worked together to frequently discuss the codes. The results of the analysis are presented in the next section.

Water Reporter Participant	Gender	Organization Title	Non-Water Reporter Participant	Gender	Organization Title
P1	F	Watershed Outreach Coordinator	P9	F	Program Coordinator
P2	F	Community Engagement Coordinator	P10	M	Scientist
P3	F	Pollution Prevention Specialist	P11	F	Water Quality Program Coordinator
P4	F	Lead Scientist	P12	F	Director
P5	F	Program Manager	P13	M	Science Coordinator
P6	M	Watershed Program Manager	P14	M	Executive Director
P7	M	Watershed Science and Restoration Program Director			
P8	F	Network Coordinator			

Table 1. Participant Demographics

RESULTS

We analyzed Water Reporter (WT) and Non Water Reporter (NWT) user interviews and compared the two groups in terms of distributed collaboration. Our analysis revealed four main facets of distributed collaboration: (1) kind of goals pursued, (2) how it strengthens community engagement, (3) social interactions, and (4) data sharing. In the following sections we discuss these facets in details and how they are manifested in WT and Non-WT user groups.

Goals of Distributed Collaboration

Both WT and Non-WT users emphasized the need for collaboration with wide range of partners for creating stronger and more successful citizen science programs. Although distributed collaboration was deemed important at both micro and macro level, collaborating regionally was often prioritized over national-level collaboration. Water quality issues are often regional and hence, more local buy-in was associated with more likelihood of success. *"for a lot of people water quality is a really hyper local issue. So this is a hot spot in our watershed, like, I want to know about that. Right? I don't care about how they're dealing with this issue in other states."* (P1)

P1 talks about importance of distributed collaboration in a local community. Citizen scientists are usually more concerned about their local environment, and the impact their work has on their local water system. Hence, collaboration at a micro level is often regarded more important. On the other hand, collaboration at a macro level requires identification of specific use-cases to drive data collection and broad engagement. For instance, data from a wide geographic area can be used to create models to see trends or make predictions about certain water quality issues. P13 gave an instance of such a use-case to explain how collaboration among a number of US states was helpful in creating a model to identify quality of trout habitat across the Appalachians and the Great Lakes. *"And basically it's a way of visually representing, where trout are, the quality of the habitat, and then various other metrics that allow us to visualize restoration opportunities on a watershed scale within a region. So we did one for like the Appalachian the brook trout, the Appalachian. We're working on one in the Great Lakes right now. It's for brook trout...the population model goes from Georgia to Maine. The Great Lakes model will be, you know, New York, Pennsylvania, Ohio, Michigan, Wisconsin, Minnesota"* (P13)

While both user groups acknowledged the importance of distributed collaboration, more established groups find it easier to find new groups across the nation to partner with. Whereas, smaller or newer groups often focused on collaborating more regionally. Out of these, smaller groups that used Water Reporter gave instances where the platform had helped them make connections that they wouldn't have considered or known otherwise.

"Yeah, that's pretty interesting to me, because I have just been working with local people here. And I'm just getting to talk to people in different I mean, most of them have been on East Coast, but it's still just like getting out of this space and talking to people who are doing monitoring, and a lot of citizen science programs in different areas, bigger cities and different demographics. That's just been really helpful for me to just understand how things are running in every in other places. And just to kind of bring that knowledge here.

And also, you know, just try to, you know, see opportunities where I can connect the folks over here to people in other places" (P5)

P5 is a member of a small citizen science group and a Water Reporter user in California. Since, most Water Reporter users are based on the East Coast of United States, P5 wouldn't have considered looking for collaborators in such a different geographic region. Though unanticipated, these collaborations facilitated by Water Reporter transpired to be beneficial. On the other hand, non- Water Reporter users usually use their existing regional networks to collaborate, and did not anticipate need for more distributed collaboration in other geographic areas.

Distributed collaboration was also recognized as important for sharing information, techniques and resources with partner organizations. Both user groups recognized the benefits of mutual learning and need to keep abreast of what other organizations were doing in different regions. Groups usually try to identify like-minded organizations who have similar interests in engaging in a project, and leverage these networks to learn and improve their practices, and create a broader impact.

"The person in Idaho that I mentioned, we've collaborated...they did a big, I guess, study and evaluation with his volunteers and our volunteers. Through his research, he found that a lot of volunteers want an educational experience. And so some of those people that we have do the training and monitor for a little bit and then stop, they kind of got what they wanted out of it. And where I used to take that was like, 'Oh, no! it didn't work, you know, like we're doing something wrong because they left!' It was kind of a new way of looking at it." (P11)

P11 explains how collaboration with group in Idaho helped them understand motivations behind their volunteer engagement which was eventually helpful in better recruitment and engagement strategies. In addition, some citizen science groups, that are well established often use their resources to collaborate and help smaller or newer regional groups. These established groups work as local champions, where their goal is to increase capacity in communities to collect data that they can understand, analyze, answer questions, and enhance their ability to protect local streams and rivers. *"In recent years, we've not only had other states approached us and asked us to share our information or model, but we've worked with three Canadian provinces now helping them set up a program similar to ours. And it's* *something we're very happy to do. We have had teams of researchers and also individuals for who are working to set up grassroots organizations, from those geographic areas come and visit us and shadow us and monitor what we do with our volunteers. So, we're very happy to be able to do that for other organizations from other areas."* (P14)

P14 explains how collaborating by sharing information and resources with other smaller organizations in different areas has been both helpful in augmenting citizen-based monitoring programs across different regions, and rewarding for their organization to be able to make a difference.

Citizen Science and Community Engagement

A longstanding goal of citizen science programs has been involving the community in scientific endeavors that may or may not involve data collection. This type of community engagement helps instill informal learning about their local environment and empower them by providing tangible ways to participate in conservation activities and make a change.

"And so community engagement wise, we have Water Reporter, which we would say is our most important effort currently. And when I say Water Reporter, I'm specifically talking about observational data. So we're specifically directing our Water Reporters to share anything that they see that they feel like they're changing, but within, you know if they need more guidance within those six areas. So Water Reporter is our most important community engagement effort currently" (P2)

P2 uses Water Reporter's photo sharing functionality for community engagement. Similar to Instagram, one can post pictures of water systems and add captions and hashtags to these posts. This activity requires very minimal volunteer training, and can be very enjoyable for community members. It gives community members a tangible way to participate in scientific endeavors, creating visible accountability, and instilling a sense of identity in their community. This inducement of intrinsic emotions and motivations can foster community engagement.

In addition, Water Reporter's accessible mapping and visualization capabilities can be easily used to convey volunteers' contributions. This will instill a sense of acknowledgement for their work, which is important for citizen science engagement. However, most Water Reporter users haven't fully leveraged this functionality, but have appropriated the platform in different ways to facilitate community engagement. For instance, some groups would publicize

Water Reporter with their collaborators to get more people in the community to use it.

"But this has been a really good partnership and it's created a really good relationship between our organization and other regional organizations. Before I launched the Water Reporter program, I connected with all those agencies and said, 'hey, look, I'm gonna run this program and a lot of pollution recording is going to be on your property. So I will be calling you about it, you should download the app and monitor it as well'. So that's been great, we have some great partnerships with an agency people." (P3)

P3 attempted to get her government agency collaborators to use Water Reporter to get wider community participation. This wider participation is helpful in both generating more data, and fostering community engagement.

Besides Water Reporter, other community engagement efforts included different ways groups tried to recruit volunteers for data collection, or engage citizens in various conservation activities. They often used their own network to reach out to community members.

"Using things with XY University, using things with the division of water quality, teachers that we have, just other partners that are doing it, nonprofit organizations, we, the Nature Center last year did a volunteer programs, all their teenagers in that program, got experience with water quality monitoring. " (P11)

P11 explains how her organization uses local networks to advertise and recruit volunteers for citizen science programs.

"The watershed project as a whole does a lot of like, there's the whole education part and capacity building, and like, as a whole, our organization works with community members and other groups to increase capacity. So, it's definitely our mission." (P5)

In addition, other community outreach initiatives like education, are also often done in partner- ship with other local citizen science groups. As P5 explains in the quote above, promoting water quality education by working with local partners helps in increasing the capacity of the community. Besides recruiting and outreach activities, one very common issue citizen science programs face is retaining volunteers. Citizen scientists often like to receive acknowledgement of their contributions, usually by seeing how their data is being used and making an impact. Hence, citizen science programs try to create efforts to create accessible reports or involve them in publications. *"That's why the platform, we developed with them,* *the water data hub, it's a map with keywords, you can see what sort of data they are posting, very transparently. I think that sort of is useful for user engagement because you can see what folks are doing all over."* (P1)

P1 is a member of a big citizen science organization that collaborates with various local watershed groups to collect water quality data. P1's organization developed an integrated data platform with one of their collaborators, so that all the data can be easily accessed by all collaborators in a single portal. P1 explains how the integrated data platform was designed to be transparent, so that volunteers could see the impact their data was making in the watershed. This transparency was helpful in user engagement, as acknowledgement of their contribution motivated volunteers to consistently engage in monitoring activities.

Social Interactions

Social interactions among citizen scientists has been observed as an important motivational factor in online citizen science projects and within field-based citizen science project group. Our interviews elucidated role of social interactions between distributed citizen scientists working in different groups. Since Water Reporter has been designed as a social media platform, it's users could easily identify benefits and challenges of such social interactions. However, interestingly non-Water Reporter users also alluded to the need to facilitate such distributed social interactions.

"One of the things that I think is missing from the databases, we just didn't have the funds to actually invest in like, the volunteer experience in making sure that there are things like Achievement Awards like "oh you just collected your 25th sample like wooho!" you gained this award or, you know, being able to integrate in some ability for volunteers to talk to each other. I remember when we were actually like sketching out ideal designs like wire-frames. We had this whole, like, interactive component where the volunteers would get to chat with each other. And so people in one area could like, you know, and then people would get alerts within that to be like, 'hey! so and so just got their hundred sample'" (P12)

P12 here talks about an integrated database that her group was building with their partners and wanted to add a social interaction feature for volunteers in different groups to communicate to each other. It is noteworthy that P12's example was not elicited by direct interview questions or prompts.

Most participants felt that social interactions between volunteers would be more useful for engagement and monitoring activities at a local or hyperlocal level. This is also connected to the goals of distributed collaboration, where most citizen science groups find collaboration at a micro level more valuable.

"But I think our public especially would be more apt to engage in something that's here locally, because they might be like, 'Oh! that looks similar to what I saw over there!' And then I could see maybe over time, more interest in like, 'okay, we're working on algae here. What are other people posting about algae?' and like clicking on the hashtag and seeing all the other algae posts." (P2)

P2's viewpoint on social interaction feature of Water Reporter was that it would be more useful to engage volunteers in their local community. Since water quality can vary regionally, people are more interested in learning and being cognizant about their local water system, and the impact their contribution has locally.

Facilitating social interactions among volunteers is especially important in a region with diverse demographics. In urban areas, where there is more diversity, volunteers usually come from different backgrounds. Having a space where these volunteers can communicate and build common ground, is important for sustaining engagement.

"We try and have our citizen scientists talk to each other and meet each other. They are all on different levels because there is that knowledge barrier and language barrier between the people. Trying to put everybody on the same playing field. Communicating saying, 'hey! I found that this works', or something like that." (P6)

P6 works for a citizen science organization in a big urban city with a diverse population. P6 explains how it's crucial to facilitate communication between these diverse groups of volunteers to ensure equal opportunity for everyone. P6 finds Water Reporter's social interaction feature valuable, but hasn't fully exploited this function to achieve the goal of bridging the gap between diverse volunteers.

Water Reporter's unique aspect is its social media feature. As described by most Water Reporter users, it has an interface similar to Instagram. The social media feature can not only facilitate social interactions between different citizen science groups, it also provides a very accessible way of collecting data. However, this feature remains under-leveraged partly due to certain design flaws and lack of accustomization to the use of social media.

"I think bringing people in the greater water cycle or whatever you want to call it, in the social media aspect a little more, like picking up trash is relevant to the local watershed group, it is relevant to the Surging River Group, that it is relevant to an Atlantic marine group. All because it's a feeder, to a larger feeder. You're not just cleaning up your local waterway, you are helping out the greater good in, you are helping the Surging River. I think people will do more." (P6, WR)

Even though Water Reporter's social media feature is not heavily used by P6's group, he sees potential in using it to facilitate social interactions between distributed citizen science groups. In the quote above he emphasizes the importance of such interactions in motivating people and increasing citizen science engagement. When different citizen science groups communicate with each other, exchange of information and social communication is helpful in cultivating knowledge that would otherwise remain unrealized.

"The smaller groups that have picked up this campaign, so they'll like, comment on other people in their watershed on their post. If you're on Water Reporter, and you work in like the Delaware River watershed, you know, and you can see who else in your watershed has posted. So you can see if there's a bunch of flood watchers in different regions" (P1)

P1 manages the 'flood watch' program on Water Reporter, where different smaller collaborating citizen science groups post observations. She explains how Water Reporter's social media feature can allow 'flood watch' volunteers to see posts from different parts of the watershed.

On the other hand, even though non-Water Reporter users alluded to the benefits of having a social interaction feature, only one participant P12 had done some tangible design thinking in this direction. P12 worked with her collaborators to prototype a social interaction feature to embed in an integrated database platform, but could not move forward with the idea due to lack of funds.

Integrated Technology for Data Sharing

Benefits of an integrated technological infrastructure has been recognized as an important way to address various citizen science data problems. An integrated technological infrastructure is beneficial for both distributed collaboration, and activities within an organization. Particularly, integrated database has been postulated as the future of citizen science [46]. Integrated database helps streamline citizen science data and address various data problems commonly identified in the citizen science literature. Integrated database in distributed collaboration helps all partner organizations to efficiently access and analyze data.

"We started a genetics project where we take a sample from every nest that gives us information from the mother that laid the eggs. And because females lay several months per season, we can track their output individuals across time, space and time. It would be really hard to do without a centralized database like this, so it has greatly facilitated that. That's across the three states. By using the database and tweaking it, I've been, able to capture more information about that threat and it's much easier having it standardized online that all the groups see." (P10)

P10 collaborates with organizations in three states. Having an integrated database helps standardize protocols across the collaborating groups, which streamlines data access and analysis processes. In addition, these integrated databases can be designed to have an in-built quality control and assurance function, which can substantially change the quality of data for all collaborators.

"And in a similar way, it's been much easier to get them to report data that way, I don't think I would have been able to get as much data from people had I said, Hey, here's a spreadsheet, or here's a paper form, can you fill us out and give it back to me? It's much better that they are using the database, which is all laid out in a standardized manner and has QA/QC built in it, it's really changed drastically the quality of our data." (P10)

P10 further explains how direct reporting of data in the database, instead of using paper data forms, and the in-built quality control and assurance feature has transformed their data quality.

Partners in distributed collaboration may have slightly different data protocols, based on the interpretations of protocols for local needs. Integrating such data can be challenging and requires work to make the protocols compatible.

"If there's anything that's a pain about her job, it's that you know, if that happens, and you don't make sure it fits the certain criteria, you have to change it, manipulate it,

to go to the government system. So we're trying to make it so that the database that we use can give her the format that is required, or at least the pieces that are required. So just it makes her life a little bit easier in many ways, because it's a lot of data. And it can be quite a headache to get it where it needs to be." (P9)

P9 works for a citizen science group that is managed and coordinated by a local environmental government agency. However, their data protocols are incompatible, which makes sharing data with the government agency arduous. This can make the process of data sharing prone to errors. Hence, It is essential to address the issue of incompatibility to maintain data quality and expedite the entire process.

"So, we do have our own database for inner tidal data. But what the problem is that we have our internal database which basically is an Excel sheet. And then you know, somebody south bound in the, say, Kallala area also collects the same kind of data, but they just keep theirs on their computer. And there might be somebody else in the Pinto area who collects the same kind of data that we don't even know about. So, we're hoping that we can have this one spot that's a repository for all of that data throughout the wider region." (P4)

P4 raises two important requirements for an integrated platform. First, the need to ensure data compatibility of all collaborators, and second, the ability to discover groups in the region collecting similar data. Therefore, a collaborative platform that not only strengthens collaborations but also helps make new collaborations.

Although Water Reporter doesn't have a database feature, it has customizable data forms, and capacity to hold, map, and visualize data. Such a design can be potentially worked on to also incorporate an integrated database to streamline distribution collaboration processes.

"The developers who made Water Reporter also developed something we call the Water Data Hub. The goal is to have a place where other watershed groups or water quality monitoring programs can use the database. It doesn't exactly have key social media function but again, it has that transparency to see what other groups are monitoring, what process results are. So that is maybe integrated pretty closely feeds that are coming in and out of Water Reporter." (P1)

P1 describes a separate integrated database her group built to have a centralized repository for all water monitoring groups to store their data in. An implication of having this centralized repository was increased transparency between participating groups. In addition, this database was linked to Water Reporter, so that all its feeds could automatically be stored in this database. This group's design of database coupled with Water Reporter, elucidates the significance of having an integrated platform that consists of functions of Water Reporter and a centralized integrated database.

An important implication of such a collaborative platform is that it can engender more transparency and visibility. Both these attributes are important to sustain and foster distributed collabo- ration. Transparency will enable citizen scientists to see how other groups are working, which can facilitate trust strengthening collaboration and community. In addition, visibility is important for awareness of unanticipated opportunities and practices.

DISCUSSION

This research is an exploratory study to investigate how technology mediated distributed collaboration between field-based citizen science groups can ameliorate their practices. Distributed collaboration is a well-studied area in CSCW research. It has proven valuable in understanding and mitigating barriers in interdisciplinary and distributed scientific groups. Hence, it seems reasonable to investigate and understand distributed collaboration in field-based citizen science. Citizen science is an inherently collaborative endeavor. Both virtual and field-based citizens science studies have studied the collaboration between scientists and citizens. However, characteristics of field-based citizen science makes distributed collaboration an important aspect, which remains under-explored in CSCW and HCI research. Hence, it is imperative to study how distributed collaboration manifests in citizen science teams, and how technology mediation can help scaffold citizen science practices. The primary contribution of this study was foregrounding the importance of distributed collaboration in citizen science and conducting a user study to investigate whether technology mediation can potentially be useful in improving citizen science practices. The user study with WR and NWR user groups highlights the importance of bonding and bridging. Communication and interactions within a group are important in strengthening a group, but communications and interactions beyond a single group helps brings resources and knowledge that weren't there before.

This research provides evidence that technology mediated distributed collaboration between field- based citizen science groups can ameliorate their practices. Through our empirical investigation of WR and NWR users we identified four facets of distributed collaboration that were important for both sets of participants, and can be more effectively mediated by technology. In this section we first reflect on our findings, and then discuss two contributions to CSCW and HCI: (1) we describe how technology can act as a mediator to provide visibility, transparency and awareness, and (2) suggest possible design directions to support distributed collaborations.

Reflections and Implications

In this section we reflect on our findings and examine their implications for CSCW and HCI.

All citizen science groups collaborate either regionally, nationally or both. Hence, all groups had clear goals of collaboration. Examining these goals helps elucidates what activities are important to incorporate in a collaborative technology. The kind of goals citizen science groups pursued slightly differed for more established and newer groups, however all participants emphasized the importance of collaborating regionally. Water quality is primarily a regional issue, and contributing to protection of local water system creates a sense of community identity. This is consistent with HCI research on creating collaborative systems that are *beyond being there* [30], that is, designing systems that facilitate telecommunication even when users are in a physical proximity. Hence, to better engage volunteers it is imperative to facilitate more awareness about citizen-based water monitoring activities in the hyperlocal community.

Community engagement has been an important aspect of citizen science [15]. Public participation in scientific activities not only helps in collecting large amounts of data, but also strengthens social capital, as it brings the community together to work towards protecting their water system, which is critical shared community resource. It inculcates informal learning about their environment and facilitates a more rationale discourse within the community [12]. All water monitoring groups we interviewed actively worked with their local partners to organize community engagement events. Collaborating with local partners for community engagement, helps groups reach a broader community, garner wider participation, and create broader impact. Such local collaborations also help groups in recruiting volunteers.

Both WR and NWR participants recognized the importance of social interactions. However, WR users had more experience with using technology to facilitate social interactions. The social media feature of Water Reporter gave the community members a tangible way to participate and contribute towards protection of their local water system. The Instagram like design provides a fun medium for community members to contribute data, interact with each other, and also become aware of similar activities in the region. In scientific communities, collaborative knowledge production is often considered a social process that takes place through discussions on a shared platform [2]. Citizen science collaborations are different from scientific collaborations, because the knowledge, skills and goals differ for the two groups. However, considering the evidence of benefits of social interactions for both scientific communities and online citizen science programs [55], it is important to leverage it for field based citizen science programs as well.

Data management is an important aspect of all citizen science programs. In water quality monitoring, data collection often requires sensitive and careful handling of equipment and samples. This makes data more prone to errors, and ensuring data quality and assurance becomes paramount in such field-based citizen science activities. In addition, groups often face challenges with efficiently sharing data with their collaborators. Effectively integrating citizen science data is one of the mainstream technical challenge in this field. Having a centralized repository where all groups can access and share data, was a core requirement for many groups interviewed. Current tools and platforms available to them do not support those interactions. Rather than facilitating citizen water quality data practices, current tools impede and limit these practices [10].

Technology Mediation: Visibility, Transparency, and Awareness

We describe the role of Water Reporter as a mediator for distributed collaboration through the lens of Vygotsky's [57] Activity Theory. It is a mediator because it assists citizen scientists in various activities such as data collection, interaction, and data mapping. These activities can be achieved without Water Reporter too, however having a mediator changes the process and outcomes of these activities. For instance, the observational data collected through the Instagram style interface, changes the way data is collected, and leads to additional social-technical outcomes such as greater transparency and interactions. On the other hand, the social media design provided citizen scientists with resources to carry out activities like data collection with minimal to no training. The familiarity and ease of use of the social media interface allowed program managers to circumvent the arduous and resource intensive task of volunteer training, while still collecting valuable data. This resonates with Activity Based Computing which states that systems should support user activities with minimal interference and resource overhead [34].

Water Reporter's mediator design also created visibility, allowing users to discover other groups and their work in the region. This affordance led to unintentional benefits, which the groups weren't necessarily looking for, but the presence and use of the particular design features augmented their activities. This can be characterized from the perspective of invisible work [52]. Collaborative systems often affect visibility of work and people, hence analyzing the context in which work or people become invisible has become an important concept for HCI and CSCW researchers. In the case of field based citizen science such as water monitoring, citizen-based initiatives can easily become invisible due to the regionality of water quality. When water monitoring groups have been working in a region for a reasonable amount of time, they organically become aware of other similar groups in their region. However, this method of building networks and collaborations holds the risk of missing out important and timely opportunities. Our findings showed how visibility engendered by Water Reporter led to unintentional connections and benefits.

A second implication of Water Reporter's mediator design was the transparency provided by the tool for its users. The social media design allowed users to see and comprehend each others activities as they occurred. This can help create awareness and provide mutual knowledge and common ground to all collaborators. In citizen science, such awareness can be helpful in providing a sense of acknowledgement, sense of community by building a common ground, and provide shared feedback [22]. In addition, activity awareness, which is awareness of citizen science actions and endeavors, can help create an environment of support for all distributed teams, by helping coordinate tasks such as data collection and discussions [13]. The importance of designing for transparency and awareness in collaborative systems is well established in software engineering. Collaborative software platforms such as GitHub provides transparency and awareness, making information more visible to the users. This visibility helps in coordinating projects and also advancing individual user's knowledge and skills [18]. Similarly, citizen science online platforms have started to incorporate design features to incorporate more transparency and awareness, however there is little research that investigates these design features and their benefits.

Design Direction for Distributed Collaboration

Some core collaborative activities identified such as data sharing, social interactions, and knowledge sharing, can all be effectively mediated by technology. However, having a different platform to support each activity can be onerous because it requires more resources, time and technical expertise to manage. While Water Reporter was not specifically built for distributed collaboration, it supported a number of collaborative activities, and WR participants suggested more design features they wished for in this system.

A virtual space that connects distributed citizen science teams and facilitates information, re- source and data sharing, is a notion similar to the concept of collaboratories [25] which views science as an "inherently collaborative enterprise". Collaboratories is a concept of creating an internet laboratory without walls where scientists from diverse geographic locations can collaborate with each other [66]. The idea is to leverage the well-established and accepted benefits of collaboration in improving scientific endeavors [39, 65], and successful design of technologies that afford collaboration [37, 38]. Hence, we suggest exploring how concept of collaboratories, which has been successfully employed in the scientific community, can be contrived to provide a similar collaborative environment to citizen scientists. A collaboratory for citizen science could be social media style design that allows easy data collection, a database to directly parse and store this data, social interactions and discussions, and methods for easy data analysis and interpretation. Social media design has been popular for data collection in a number of disciplines. In fact, algorithms can be developed to create a database than can directly parse and store social media data [19].

Even though skills and activities of professional scientists differs from citizen scientists, scientific collaboratories can be used as a model to guide the design of citizen science collaboratory. For instance, Kahin and Foray [33] provides a taxanomy of initiatives (community-centric, data-centric, computation-centric, interaction-centric) that a scientific collaboratory should support. We can map these initiatives to socio-technical affordances that these initiatives can serve for citizen scientists. Similarly, Kouzes et al. [36] developed categories of collaboration that can be useful in identifying different relationships and collaboration needs in citizen science.

Limitations and Future Work

Even though our empirical investigation provided evidence of distributed collaboration, and that technology can mediate and augment this activity, we also acknowledge some limitations of this research. Qualitative process of data collection from WR and NWR participants limited the comparative analysis on the two user groups. Even though our data provides some evidence of differences in collaborative activities of two user groups, the comparison is not balanced and can not be quantified. In addition, collaboratory nature of Water Reporter is limited. Even though it has features that can be appropriated to support collaboration, it lacks functions to support distributed collaboration. The findings from the user study provided us a starting point to initiate the discussion around distributed collaborative work in citizen science, and analyze it through the lens of invisible work, technology mediation and awareness. However, lack of quantitative data around specific features and their socio-technical affordances makes it difficult to further the analysis.

In this research, we try to capitalize on the existing ecology of collaborative tools to try and assess the value of collaborative features. For future work, it would be desirable to implement the best case of such features, that is, a collaboratory kind of design for citizen science. Also, it would be helpful to directly study the presence and absence of the feature using both qualitative and quantitative analysis to get a more robust analysis of benefits and limitations of such design.

CONCLUSION

This research attempted to investigate distributed collaboration between different water monitoring citizen science groups. We used a water monitoring platform called Water Reporter with a unique social network design to conduct a comparative investigation between users and non-users of this app. This helped us understand how visibility and transparency afforded by Water Reporter distinguished collaborative activities for the two user groups. We found that both sets of participants valued and benefited from collaboration, but Water Reporter users experienced some unanticipated benefits mediated by this application. The analysis of our findings provides two meaningful contributions to CSCW and HCI: (1) we describe how visibility and transparency mediated by technology can create awareness and facilitate distributed collaboration in citizen science, and (2) we suggest a collaboratory design for citizen science to support different collaborative activities, given its established significance and success in scientific communities.

As scientific endeavors have given rise to new research phenomena of public participation in science, it only seems reasonable to create collaborative environments for citizen scientists. Citizen scientists distributed across widespread area, and situated in different organizational structures will often have different goals. This is because volunteers usually participate in citizen science initiatives that seek to work on local issues. However, differences in participants' goals is true for every collaboration. Comprehensive requirement analysis of these volunteers will aid in the design of an appropriate collaboratory, that will have the potential to provide benefits to all participating actors.

Therefore, facilitating distributed collaboration between hyperlocal citizen science groups has the potential to address key citizen science issues such as informal learning, retention and engagement of volunteers. These challenges are common to various groups and are frequently studied and highlighted in citizen science literature. Distributed collaboration can help prevent re-inventing the wheel by helping exchange valuable information and ameliorate citizen practices.

REFERENCES

[1] Bianca Ambrose-Oji, AP van der Jagt, and Sue O'Neil. 2014. Citizen Science: Social Media as a supporting tool.

[2] Ernesto Arias, Hal Eden, Gerhard Fischer, Andrew Gorman, and Eric Scharff. 2000. Transcending the individual human mind—creating shared understanding through collaborative design. *ACM Transactions on Computer-Human Interaction (TOCHI)* 7, 1 (2000), 84–113.

[3] Akarsh Asoka, Tom Gleeson, Yoshihide Wada, and Vimal Mishra. 2017. Relative contribution of monsoon precipitation and pumping to changes in groundwater storage in India. *Nature Geoscience* 10, 2 (2017), 109–117.

[4] Patrick Biernacki and Dan Waldorf. 1981. Snowball sampling: Problems and techniques of chain referral sampling. *Sociological methods & research* 10, 2 (1981), 141–163.

[5] Jeremy Birnholtz. 2004. Factors affecting the utility of technology-mediated

collaboration in science and engineering. In *CHI'04 Extended Abstracts on Human Factors in Computing Systems*. 1045–1046.

[6] Rick Bonney, Heidi Ballard, Rebecca Jordan, Ellen McCallie, Tina Phillips, Jennifer Shirk, and Candie C Wilderman. 2009. Public Participation in Scientific Research: Defining the Field and Assessing Its Potential for Informal Science Education. A CAISE Inquiry Group Report. *Online Submission* (2009).

[7] Rick Bonney, Jennifer L Shirk, Tina B Phillips, Andrea Wiggins, Heidi L Ballard, Abraham J Miller-Rushing, and Julia K Parrish. 2014. Next steps for citizen science. *Science* 343, 6178 (2014), 1436–1437.

[8] Virginia Braun and Victoria Clarke. 2012. Thematic analysis. (2012).

[9] Lindsey J Butler, Madeleine K Scammell, and Eugene B Benson. 2016. The Flint, Michigan, water crisis: a case study in regulatory failure and environmental injustice. *Environmental Justice* 9, 4 (2016), 93–97.

[10] John M Carroll and Jordan Beck. 2019. Co-designing platform collectivism. *CoDesign* 15, 3 (2019), 272–287.

[11] John Millar Carroll, Jordan Beck, Elizabeth W Boyer, Shipi Dhanorkar, and Srishti Gupta. 2019. Empowering Community Water Data Stakeholders. *Interacting with Computers* 31, 3 (2019), 492–506.

[12] John M Carroll, Jordan Beck, Shipi Dhanorkar, Jomara Binda, Srishti Gupta, and Haining Zhu. 2018. Strengthening community data: towards pervasive participation. In *Proceedings of the 19th Annual International Conference on Digital Government Research: Governance in the Data Age*. 1–9.

[13] John M Carroll, Dennis C Neale, Philip L Isenhour, Mary Beth Rosson, and D Scott McCrickard. 2003. Notification and awareness: synchronizing task-oriented collaborative activity. *International Journal of Human-Computer Studies* 58, 5 (2003), 605–632.

[14] Jeffrey P Cohn. 2008. Citizen science: Can volunteers do real research? *BioScience* 58, 3 (2008), 192–197.

[15] Cathy C Conrad and Krista G Hilchey. 2011. A review of citizen science and community-based environmental monitoring: issues and opportunities. *Environmental monitoring and assessment* 176, 1-4 (2011), 273–291.

[16] Myriah L Cornwell and Lisa M Campbell. 2012. Co-producing conservation and knowledge: Citizen-based sea turtle monitoring in North Carolina, USA. *Social Studies of Science* 42, 1 (2012), 101–120.

[17] Joe Cox, Eun Young Oh, Brooke Simmons, Chris Lintott, Karen Masters, Anita Greenhill, Gary Graham, and Kate Holmes. 2015. Defining and measuring success in online citizen science: A case study of Zooniverse projects. *Computing in Science & Engineering* 17, 4 (2015), 28–41.

[18] Laura Dabbish, Colleen Stuart, Jason Tsay, and Jim Herbsleb. 2012. Social coding in GitHub: transparency and collaboration in an open software repository. In *Proceedings of the ACM 2012 conference on computer supported cooperative work*. 1277–1286.

[19] Jens A de Bruijn, Hans de Moel, Brenden Jongman, Marleen C de Ruiter, Jurjen Wagemaker, and Jeroen CJH Aerts. 2019. A global database of historic and real-time flood events based on social media. *Scientific Data* 6, 1 (2019), 1–12.

[20] David De Roure, Nicholas R Jennings, and Nigel Shadbolt. 2003. The semantic grid: A future e-science infrastructure. *Grid Computing-Making the Global Infrastructure a Reality* (2003), 437–470.

[21] Janis L Dickinson, Jennifer Shirk, David Bonter, Rick Bonney, Rhiannon L Crain, Jason Martin, Tina Phillips, and Karen Purcell. 2012. The current state of citizen science as a tool for ecological research and public engagement. *Frontiers in Ecology and the Environment* 10, 6 (2012), 291–297.

[22] Paul Dourish and Victoria Bellotti. 1992. Awareness and coordination in shared workspaces. In *Proceedings of the 1992 ACM conference on Computer-supported cooperative work*. 107–114.

[23] Paul N Edwards, Matthew S Mayernik, Archer L Batcheller, Geoffrey C Bowker, and Christine L Borgman. 2011. Science friction: Data, metadata, and collaboration. *Social studies of science* 41, 5 (2011), 667–690.

[24] Umer Farooq, Craig H Ganoe, John M Carroll, and C Lee Giles. 2009. Designing for e-science: Requirements gathering for collaboration in CiteSeer. *International Journal of Human-Computer Studies* 67, 4 (2009), 297–312.

[25] Thomas A Finholt. 2002. Collaboratories. (2002).

[26] Laura Grant and Christian Langpap. 2019. Private provision of public goods by environmental groups. *Proceedings of the National Academy of Sciences* 116, 12 (2019), 5334–5340.

[27] Yurong He and Andrea Wiggins. 2017. Implementing an environmental citizen science project: Strategies and concerns from educators' perspectives. *International Journal of Environmental & Science Education* 12, 6 (2017), 1459.

[28] Christine Hine. 2020. Knowledge infrastructures for citizen science: The taming of knowledge. In *A History of Participation in Museums and Archives*. Routledge, 93–108.

[29] Christian Pieter Hoffmann, Christoph Lutz, and Miriam Meckel. 2014. Impact factor 2.0: Applying social network analysis to scientific impact assessment. In *2014 47th Hawaii International Conference on System Sciences*. IEEE, 1576–1585.

[30] Jim Hollan and Scott Stornetta. 1992. Beyond being there. In *Proceedings of the SIGCHI conference on Human factors in computing systems*. 119–125.

[31] Yen-Chia Hsu, Paul Dille, Jennifer Cross, Beatrice Dias, Randy Sargent, and Illah Nourbakhsh. 2017. Community- empowered air quality monitoring system. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*. 1607–1619.

[32] Marina Jirotka, Charlotte P Lee, and Gary M Olson. 2013. Supporting scientific collaboration: Methods, tools and concepts. *Computer Supported Cooperative Work (CSCW)* 22, 4-6 (2013), 667–715.

[33] Brian Kahin and Dominique Foray. 2006. Toward a Cyberinfrastructure for Enhanced Scientific Collaboration: Providing Its "Soft" Foundations May Be the Hardest Part. (2006).

[34] Victor Kaptelinin and Bonnie Nardi. 2012. Activity theory in HCI: Fundamentals and reflections. *Synthesis Lectures Human-Centered Informatics* 5, 1 (2012), 1–105.

[35] Brandi Koehler and Tomas M Koontz. 2008. Citizen participation in collaborative watershed partnerships. *Environmental management* 41, 2 (2008), 143.

[36] Richard T Kouzes, James D Myers, and William A Wulf. 1996. Collaboratories: Doing science on the Internet. *Computer* 29, 8 (1996), 40–46.

[37] Robert E Kraut and Paul Attewell. 1997. Media use in a global corporation: Electronic mail and organizational knowledge. *Culture of the Internet* (1997), 323–342.

[38] Robert E Kraut, Jolene Galegher, and Carmen Egido. 1987. Relationships and tasks in scientific research collaboration. *Human–Computer Interaction* 3, 1 (1987), 31–58.

[39] Betty Wolder Levin and Alan R Fleischman. 2002. Public health and bioethics: the benefits of collaboration.

[40] Andrea Liberatore, Erin Bowkett, Catriona J MacLeod, Eric Spurr, and Nancy Longnecker. 2018. Social media as a platform for a citizen science community of practice. *Citizen Science: Theory and Practice* 3, 1 (2018).

[41] Kathleen E Little, Masaki Hayashi, and Steve Liang. 2016. Community-based groundwater monitoring network using a citizen-science approach. *Groundwater* 54, 3 (2016), 317–324.

[42] Kurt Luther, Scott Counts, Kristin B Stecher, Aaron Hoff, and Paul Johns. 2009. Pathfinder: an online collaboration environment for citizen scientists. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 239–248.

[43] Karen Masters, Eun Young Oh, Joe Cox, Brooke Simmons, Chris Lintott, Gary Graham, Anita Greenhill, and Kate Holmes. 2016. Science learning via participation in online citizen science. *arXiv preprint arXiv:1601.05973* (2016).

[44] Amy Maxmen. 2018. As Cape Town water crisis deepens, scientists prepare for'Day Zero'. *Nature* 554, 7690 (2018).

[45] Matthew S Mayernik, Jillian C Wallis, and Christine L Borgman. 2013. Unearthing the infrastructure: Humans and sensors in field-based scientific research. *Computer Supported Cooperative Work (CSCW)* 22, 1 (2013), 65–101.

[46] Greg Newman, Andrea Wiggins, Alycia Crall, Eric Graham, Sarah Newman, and Kevin Crowston. 2012. The future of citizen science: emerging technologies and

shifting paradigms. *Frontiers in Ecology and the Environment* 10, 6 (2012), 298–304.

[47] Nigini Oliveira, Eunice Jun, and Katharina Reinecke. 2017. Citizen science opportunities in volunteer-based online experiments. In *Proceedings of the 2017 CHI conference on human factors in computing systems*. 6800–6812.

[48] Gwen Ottinger. 2010. Buckets of resistance: Standards and the effectiveness of citizen science. *Science, Technology, & Human Values* 35, 2 (2010), 244–270.

[49] Dana Rotman, Jenny Preece, Jen Hammock, Kezee Procita, Derek Hansen, Cynthia Parr, Darcy Lewis, and David Jacobs. 2012. Dynamic changes in motivation in collaborative citizen-science projects. In *Proceedings of the ACM 2012 conference on computer supported cooperative work*. 217–226.

[50] Jennifer L Shirk, Heidi L Ballard, Candie C Wilderman, Tina Phillips, Andrea Wiggins, Rebecca Jordan, Ellen McCallie, Matthew Minarchek, Bruce V Lewenstein, Marianne E Krasny, et al. 2012. Public participation in scientific research: a framework for deliberate design. *Ecology and society* 17, 2 (2012).

[51] Jonathan Silvertown. 2009. A new dawn for citizen science. *Trends in ecology & evolution* 24, 9 (2009), 467–471.

[52] Susan Leigh Star and Anselm Strauss. 1999. Layers of silence, arenas of voice: The ecology of visible and invisible work. *Computer supported cooperative work (CSCW)* 8, 1-2 (1999), 9–30.

[53] Martin Storksdieck, Jennifer Lynn Shirk, Jessica L Cappadonna, Meg Domroese, Claudia Göbel, Muki Haklay, Abraham J Miller-Rushing, Philip Roetman, Carla Sbrocchi, and Katrin Vohland. 2016. Associations for citizen science: regional knowledge, global collaboration. *Citizen Science: Theory and Practice* 1, 2 (2016).

[54] Andrea K Thomer, Michael Bernard Twidale, and Matthew J Yoder. 2018. Transforming Taxonomic Interfaces: " Arm? s Length" Cooperative Work and the Maintenance of a Long-lived Classification System. *Proceedings of the ACM on Human-Computer Interaction* 2, CSCW (2018), 1–23.

[55] Ramine Tinati, Max Van Kleek, Elena Simperl, Markus Luczak-Rösch, Robert Simpson, and Nigel Shadbolt. 2015. Designing for citizen data analysis: A cross-sectional case study of a multi-domain citizen science platform. In

Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems. 4069–4078.

[56] Ma Dolores C Tongco. 2007. Purposive sampling as a tool for informant selection. *Ethnobotany Research and applications* 5 (2007), 147–158.

[57] L Vygotsky. [n.d.]. Instrumental method in psychology [Instrumentalnyj metod v psikhologii]. *LS Vygotsky. Collected Works* 1 ([n. d.]).

[58] Ruo-Qian Wang. 2018. Big data of urban Flooding: dance with social media, citizen science, and artificial intelligence. *EGUGA* (2018), 404.

[59] Etienne Wenger. 2011. Communities of practice: A brief introduction. (2011).

[60] Sarah Elizabeth West and Rachel Mary Pateman. 2016. Recruiting and retaining participants in citizen science: what can be learned from the volunteering literature? *Citizen Science: Theory and Practice* (2016).

[61] Andrea Wiggins and Kevin Crowston. 2010. Developing a conceptual model of virtual organisations for citizen science. *International Journal of Organisational Design and Engineering* 1, 1-2 (2010), 148–162.

[62] Andrea Wiggins and Kevin Crowston. 2010. Distributed scientific collaboration: research opportunities in citizen science. In *CSCW 2010 Workshop*. 1–4.

[63] Wesley Willett, Paul Aoki, Neil Kumar, Sushmita Subramanian, and Allison Woodruff. 2010. Common sense community: scaffolding mobile sensing and analysis for novice users. In *International Conference on Pervasive Computing*. Springer, 301–318.

[64] Wesley Willett, Jeffrey Heer, Joseph Hellerstein, and Maneesh Agrawala. 2011. CommentSpace: structured support for collaborative visual analysis. In *Proceedings of the SIGCHI conference on Human Factors in Computing Systems*. 3131–3140.

[65] Judith D Wilson, Nathan Hoskin, and John T Nosek. 1993. The benefits of collaboration for student programmers. *ACM SIGCSE Bulletin* 25, 1 (1993), 160–164.

[66] William A Wulf. 1993. The collaboratory opportunity. *Science* 261, 5123 (1993), 854–856.