Proceedings of the 2014 IEEE 18th International Conference on Computer Supported Cooperative Work in Design

A Meta-Model for Crowdsourcing Platforms in Data Collection and Participatory Sensing

Ana Luiza Dallora Moraes[1], Felipe Fonseca[1], Maria Gilda P. Esteves[1], Daniel Schneider[1,2], Jano M. de Souza[1]
[1]PESC/COPPE/UFRJ, Graduate School of Engineering, Federal University of Rio de Janeiro, Brazil
[2]DEMAT/ICE/UFRRJ, Computer Science School, Federal Rural University of Rio de Janeiro, Brazil
Rio de Janeiro, RJ, Brazil
{analuiza, fmadureira, gilda, schneider, jano}@cos.ufrj.br

Abstract **- Crowdsourcing is both a new paradigm of collaboration and a new frontier for CSCWD. Since its introduction in 2006, the notion of Crowdsourcing has been the object of a number of studies. Crowdsourcing has a multidisciplinary nature with a huge diversity of applications that encompasses many practices. The concepts that shape crowdsourcing are still under construction and it is difficult to reach a consensus. Based on this scenario, this paper proposes a meta-model which aims to fit generic solutions related to a specific subset of crowdsourcing designed to accomplish a specific task: Data Collection and Participatory Sensing to leverage the engagement of volunteers with Science and Environmental issues.**

Keywords-crowdsourcing; data collection; participatory sensing; citizen science; environment-centric sensing

I. INTRODUCTION

"*Remember outsourcing? Sending jobs to India and China is so 2003. The new pool of cheap labour: everyday people using their spare cycles to create content, solve problems, even do corporate R & D*" [1]. These are the words of Jeff Howe, who introduced the concept of Crowdsourcing to the world in a publication in *Wired* Magazine in 2006. Since then, more and more studies on this subject have been produced and published, showing Crowdsourcing is a growing area that is attracting attention from the academic community and the industry[2].

A. Crowdsourcing

There are several ways to conceptualize crowdsourcing. Howe, in his article, defines it as the act of outsourcing a function, previously performed by employees of an organization, to an indefinite network of people in the form of an open call [1]. Defining crowdsourcing in this way may lead it to be mistaken for a synonym for outsourcing. [2] makes a disambiguation of these concepts in stating that in *outsourcing* there is a direct contact with the service provider, i.e., there is a selection process of a provider by knowing exactly who will conduct the service and by executing a contract. In crowdsourcing, on the other hand, the call for the execution of a task is open to an indefinite network of individuals who choose to do it or not, with no formal contract with the requesting organization. In any case, this definition of crowdsourcing draws attention to an important factor of crowdsourcing: the open call.

It is also possible to define the concept of crowdsourcing as a set of approaches that take advantage of the potential of crowds to perform tasks [3]. According to the Merriam-Webster dictionary, crowdsourcing is the practice of obtaining needed services, ideas, or content by soliciting contributions from a large group of people and especially from the online community rather than from traditional employees or suppliers.

This concept also features crowdsourcing as an emerging model of distributed production and problem-solving way where networked people work collaboratively to do a given job. This model has been used by organizations as a way to employ an online scalable workforce [4]. Other important factors can be derived from these other definitions, and these factors are the foundations of crowdsourcing: the crowd and collaboration in order to accomplish a task.

Crowdsourcing initiatives can take the shape of peer-production (when a job is done collaboratively), but is also often undertaken by sole individuals. The crucial prerequisite is the use of an open call format, and the wide network of potential workers [1].

The diversity of crowdsourcing applications allows it to be an effective and powerful practice that may be identified virtually with any type of Internet-based collaborative activity, such as co-creation or user innovation. However, the fact that crowdsourcing encompasses many practices makes it difficult to define and categorize [5].

After the analysis of 40 original definitions for crowdsourcing, [5] showed that there are distinct definitions of crowdsourcing and clearly illustrated a lack of consensus and a certain semantic confusion. However, eight characteristics common to any given crowdsourcing initiative were found: the crowd; the task; the reward obtained; the *crowdsourcer* or requester of the crowdsourcing activity; what is obtained by their following the crowdsourcing process; the type of process; the call to participate; and the medium. The same authors proposed an integrated definition of crowdsourcing which is reproduced below:

"Crowdsourcing *is a type of participative online activity in which an individual, an institution, a non-profit organization, or company proposes to a group of individuals of varying knowledge, heterogeneity, and number, via a flexible open call, the voluntary undertaking of a task. The user will receive the satisfaction of a given type of need, be it economic, social recognition, self-esteem, or the development of individual*

978-1-4799-3776-9/14/$31.00 ©2014 IEEE

skills, while the crowdsourcer will obtain and utilize to their advantage what the user has brought to the venture, whose form will depend on the type of activity undertaken" [5].

B. Current Scenario

A secondary study under the current scenario of crowdsourcing was performed and highlights the immaturity of the field. It was shown that most of the published studies are based on a particular type of application, resulting in too many definitions and applications [2].

As well as new crowdsourcing applications are appearing every day, its definition is under constant evolution. Schneider and colleagues discuss a myriad of tools that allow the exchange of ideas, non-hierarchical decision-making, and the full use of the world's mind space – and the notion of Crowd Computing [6]. The authors demonstrate that crowdsourcing systems intersect with other classes of computational systems (human computation, social computing, and audience-computer interaction systems).

[7] argues that the term crowdsourcing has been used to describe several, related but different, phenomena and what might be successful with one form of crowdsourcing may not be with another. Through an open call, individuals from the crowd may be used to perform simple tasks such as data collection and translations of simple texts or to achieve more complex tasks (*e.g.* problem-solving). In between these extremes, there is an intermediate category related to creative tasks in fields, such as artistic design or software applications [8].

Crowdsourcing has the ability to harness volunteers who might not otherwise be able to contribute, so expanding the involvement of customers/users in the design and improvement of products, and in scientific and community projects.

The use of crowdsourcing platforms demands from CSCW designers attention both to the task requirements as the aspirations of potential and actual contributors so that the collaboration provides the desired results and benefits for all parties (requesters and participants) [9].

The use of crowdsourcing to improve the engagement of volunteers with Science is known as 'citizen science', 'crowdsourcing for science', 'networked science', or 'crowd science'[10].

In Citizen Science projects volunteers are invited to collaborate in scientific investigations applying some human cognitive ability [11][12][13][14][15]. Thanks to the Internet, Citizen Science is expanding its reach and establishing itself as a cheap and viable alternative for scientific work that was previously restricted to the limited capacity of professional teams [14][16]. The technological advances of the Web 2.0 and the increased use of mobile devices are breaking down the barriers that once separated amateurs from professionals [1] by harnessing volunteers who might not otherwise be able to contribute [7].

Today, there are more than 6.8 billion mobile phones around the world [17], each one of them embedded with cameras, microphones and GPS, to name but a few tools. The ubiquity of these new mobile advantages has driven the rise of a new paradigm for accomplishing large-scale sensing data collection, known in literature as participatory sensing [18][19].

Participatory sensing is a generic term that describes the use of mobile devices, such as mobile phones, as sensors where citizens voluntarily contribute sensor data for their own benefit and/or the benefit of the community [18].

The use of humans as a network of sensors to collect distributed geographic data is an approach that has already been discussed in the literature [19][20]. Following the citizen Science model of leveraging volunteers as data collectors, the use of volunteers and their mobile devices as data-collection tools provides unprecedented spatial-temporal coverage and makes it possible to observe unpredictable events, which may otherwise be excluded by static sensors [18].

C. Goal

Believing that conceptualizing what Crowdsourcing really is is a limiting factor in the advancement of this field, we need an understanding of the various solutions derived from it to be able to conceptualize it correctly [2]. Based on this fact, this article presents a meta-model that aims to generalize and study the solutions relating to data collection and participatory sensing, proposing a framework and conceptual modelling, to facilitate the understanding of current tools and to facilitate the building of new ones.

This paper is structured as follows: Section II introduces the concepts of Data Collection and Participatory Sensing; Section III presents the existing challenges in these fields; Section IV shows the proposed meta-model which is the main purpose of this paper, and finally, Section V provides the conclusions.

II. DATA COLLECTION AND PARTICIPATORY SENSING

The Data Collection is an essential part of the scientific process. From a data set we can carry out observations and hypotheses, make scientific theories, which will be corroborated by the new data or recycled, if applicable. The use of crowdsourcing platforms is important to facilitate the work of collecting, whether to provide large spatial coverage or to increase the amount of data collected per unit time.

In the context of Crowdsourcing discussed in the previous section, a citizen Science data collection project is based on conducting an open call for individuals to collect and submit data. Thus, the data would not be collected only by a specific group of experts but by anyone interested and able to participate in the project. Several projects follow this paradigm, such as The *Great SunFlower* [21] project, whose participants must submit data on the number of bees visiting the sunflower to analyze data on pollination, the *eBird project* [22], where the goal is to collect information on the location and species of sighted birds, among others.

The main difference between traditional crowdsourcing data collection and participatory sensing data collection is the use of embedded existing sensing (e.g., mobile phones) and communication (cellular or WiFi) infrastructure to collect data.

According to [23] Participatory Sensing is the process through which individuals and communities are increasingly using mobile devices and cloud services to systematically collect and analyze data for several purposes.

Through the use of sensors (e.g., cameras, motion sensors, and GPS) built into mobile phones and Web services, a new collective capacity is emerging where people participate in activity of data collection and in the analysis of aspects of their lives that were previously invisible [24].

The broad range of new sensing applications can be categorized as either people-centric or environment-centric sensing [18]. People-centric applications mainly focus on documenting activities (e.g., sport experiences) and understanding the behaviour of individuals (e.g., eating disorders). In contrast, environment-centric sensing apps collect environmental parameters (e.g., air quality or noise pollution).

The crowdsourcing sensing tasks in mobile phones can be triggered manually, automatically, or based on the current context [18]. By simply enabling the sensor to the application, volunteers can collect data automatically, without any individual work. Cuff and colleagues [25] define mobile devices as 'passive sensors that can silently collect, exchange and process information all day'. And it is precisely with these features that many projects accomplish their collections. The *NoiseTube* [26] project, for instance, uses the smartphone sound recorder to map the noise level of the various areas of the globe.

In the context of a data collection project, Tweddle and colleagues define four entities, as follows [27]:

- Scientists - mainly interested in the results. It is the role that is behind the creation of a data collection project.
- Communities - a group of people who have a common interest, relevant to the project at hand.
- Participants - unpaid person who performs the various tasks of a project.
- Projects - the activity itself.

As regards participatory sensing, it is possible to identify the same entities shown above, and the difference from one to the other is after all related to the medium used for data acquisition. In participatory sensing, the volunteers have to necessarily use a mobile sensing component to complete the crowdsourcing task [18]. The sensing component is located in the participants' mobile phones and captures different kinds of sensor data, such as time, location, pictures, motion, pollution and sound data, to name a few.

III. CHALLENGES

The two Crowdsourcing modes shown above (data collection and participatory sensing) are part of a wider group called Citizen Science. It refers to the context in which individuals, mostly not part of the scientific or academic community, engage in scientific and environmental projects, expecting a reward or not for their efforts [27]. On one hand this research mode enables the collection of data on a scale larger than the traditional method, but on the other it brings challenges to motivate individuals to engage in data collection. Furthermore, the introduction of amateurs can promote a decrease in the accuracy of the results and a lack of confidence amidst end users.

A. Motivation

Since its first appearance in 2006, Crowdsourcing has been worked extensively in the technological aspects, but there are few studies that address the motivational aspects [16].

In [16] the authors conduct an experiment aimed at studying possible motivational factors in crowdsourcing systems. They used volunteers that performed activities related to the resolution of small parts of a big task, and another image classification activity. Their results showed that motivational factors associated with the second activity were higher for all motivation indicators. Amongst the reasons that influenced the participants in the experiment, the major indicators were related to the collective, to the importance assigned to the goal of the task, and the intrinsic motives, which relate to altruism, fun, reciprocity, intellectual stimulation, and a feeling of obligation to contribute.

The motivation element has been little studied and should be taken seriously in the implementation of crowdsourcing systems, as with no participants the crowdsourcing platform is doomed to fail.

B. Quality

Data quality is a major concern in crowdsourcing activities. This concern arises directly from its definition, as when lay or amateur individuals are introduced directly in the context of a scientific project, the chance to introduce noise in the collected data may increase. So there is a need to promote means of increasing the confidence in and accuracy of its results [28].

In his research work, Tweddle and colleagues [27] emphasize the importance of training the participants of crowdsourcing tasks in order to collect quality data. In addition to training, it is also important to provide user support by creating forms, handouts, reference guides and consultation, or even a direct communication channel with the project support as a way of facilitating the work of the participant and minimizing the complexity of the task.

In [29], Antelio, M., and colleagues proposed a collaborative framework to improve data quality in Citizen Science projects. By associating data quality dimensions to scientists through a voting network, the authors aimed to create a continuous process for data quality validation.

IV. META-MODEL

Starting from the case study of existing systems, it was possible to observe common behaviours and concepts. From these observations, we mapped the activities of this crowdsourcing project type and built models using the *Unified Modelling Language* (UML). Initially, an activity diagram was created with the goal of mapping the project flow and steps. The diagram in question is shown in Figure 1.

this diagram we could see that there are three main actors. The *requester* is the person who wants a particular task to be performed, and the one that [27] refers to as a scientist and [18] as campaign administrator. The choice of the term 'requester' in this work was due to the goal of obtaining a higher degree of generalization. The *evaluator*, in its turn, is the actor who is responsible for evaluating the data and, therefore, the final quality of the data collected.

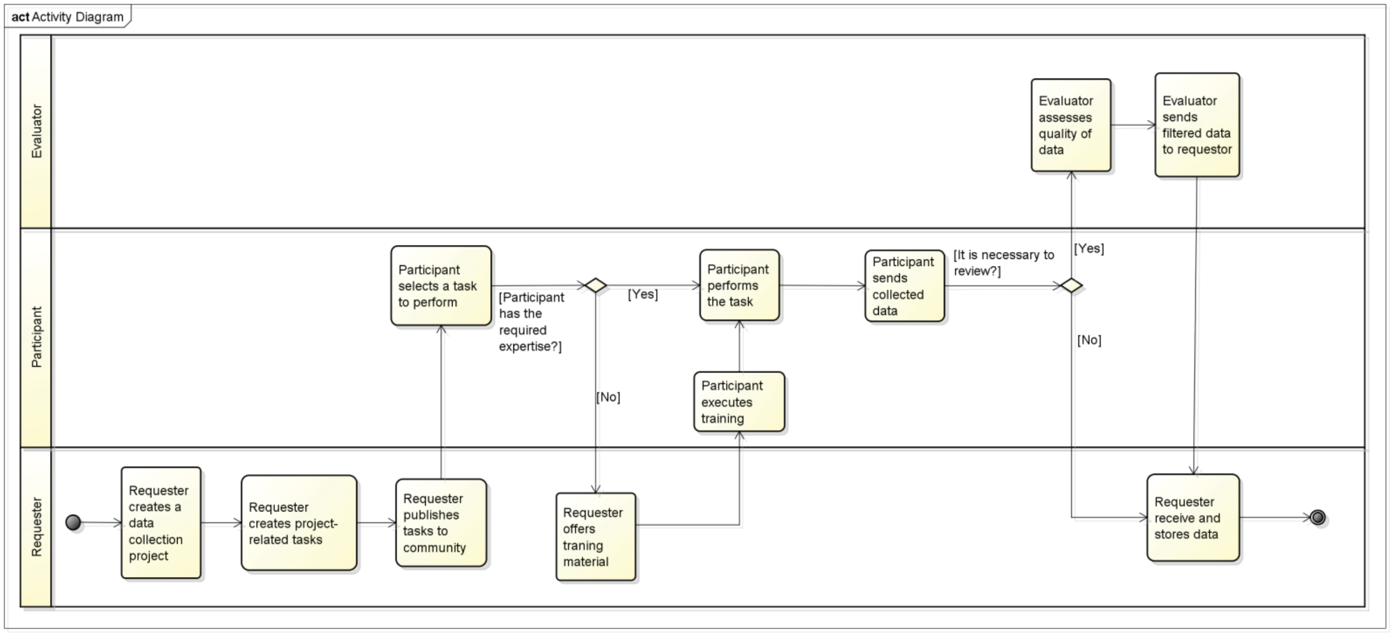

Figure 1. Activity diagram of data collection in crowdsourcing

The standard flow begins with the creation of a data collection project or participatory sensing data collection campaign by the requester. This project is broken into tasks, such as counting a particular event or gathering sensor readings with the mobile phones. Thereafter, the requester performs an open call to the crowd and participants can start contributing. One participant, since endowed with the ability or *expertise* required to perform the task, can run it. If one does do not have the adequate knowledge, one can be trained to perform that task. Once the task has been performed, the collected data should be sent to the requester and it may undergo a process of quality assessment before being effectively stored by the requester. The evaluation process should be done by the evaluator, who may be the actual requester, the community or a third party and, therefore, a role that can be played by different entities.

Once this investigation was done, we decided to create a use-case diagram with the three actors: requester, participant, and evaluator. Participants may be classified as automatic or manual, depending on the human interaction required to submit their contributions. They can perform tasks or undergo training to get the necessary expertise. The requester, in turn, can create designs for data collection and request tasks for the community. The evaluator, if any, will be responsible for evaluating the submitted measures. The use-case diagram is shown in Figure 2.

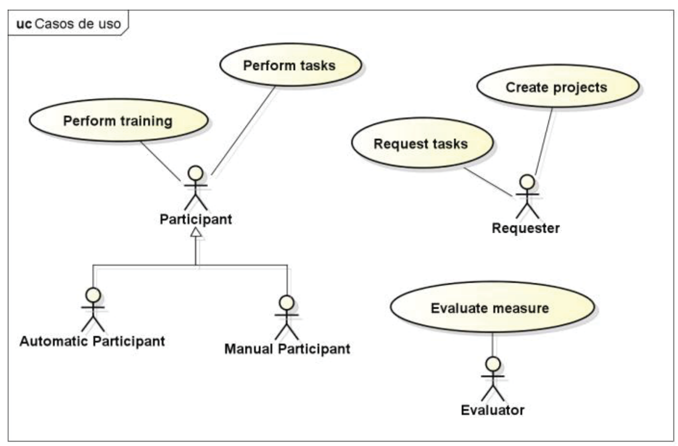

Figure 2. Use-case diagram

After this initial mapping it was possible to create a conceptual class diagram to represent the concepts and links between them. This model seeks to be generic enough to meet existing applications and serve as a basis for creating new initiatives. The model is presented below in Figure 3.

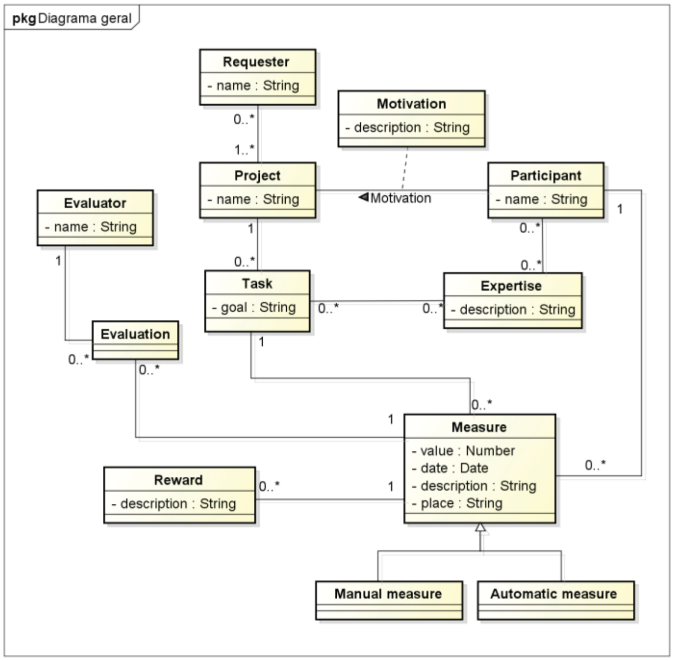

Figure 3. Conceptual class diagram

In this model we could implement several previously-presented concepts, such as the concept of participant, requester and project, for instance. Besides the already established concepts, others were created to meet the requirements of such a system. For a given task, the participant can take measurements and send them to the requester. The Measurement class represents this concept and can be either a manual measurement when the participant is directly involved in measuring the task (e.g. manual reports and taking photographs) or an automatic measurement, when done by some kind of automatic sensor (e.g. Monitoring noise). For every performed measurement, the participant may obtain one or more rewards, benefits and possible financial benefits to the group or individual advantages. The rewards, however, are not restricted to these groups. The concept of motivation was discussed earlier in this article and has been mapped as an association between participant and project to represent the bond the participant has with the project. The *expertise,* in its turn, represents skills, ability or tools the task requires for its completion. In addition, a participant may have a particular set of elements of *expertise.* The *expertise* required to complete the task as well as the *expertise* the participant has will determine whether one is feasible or not to contribute to the task.

To validate the model we used a third existing model in UML: the object diagram. By using the project *The Great SunFlower Project*, we simulated the operation and created instances of mock objects to verify that the proposed model conformed to an existing system. This result is displayed in the object diagram of Figure 4.

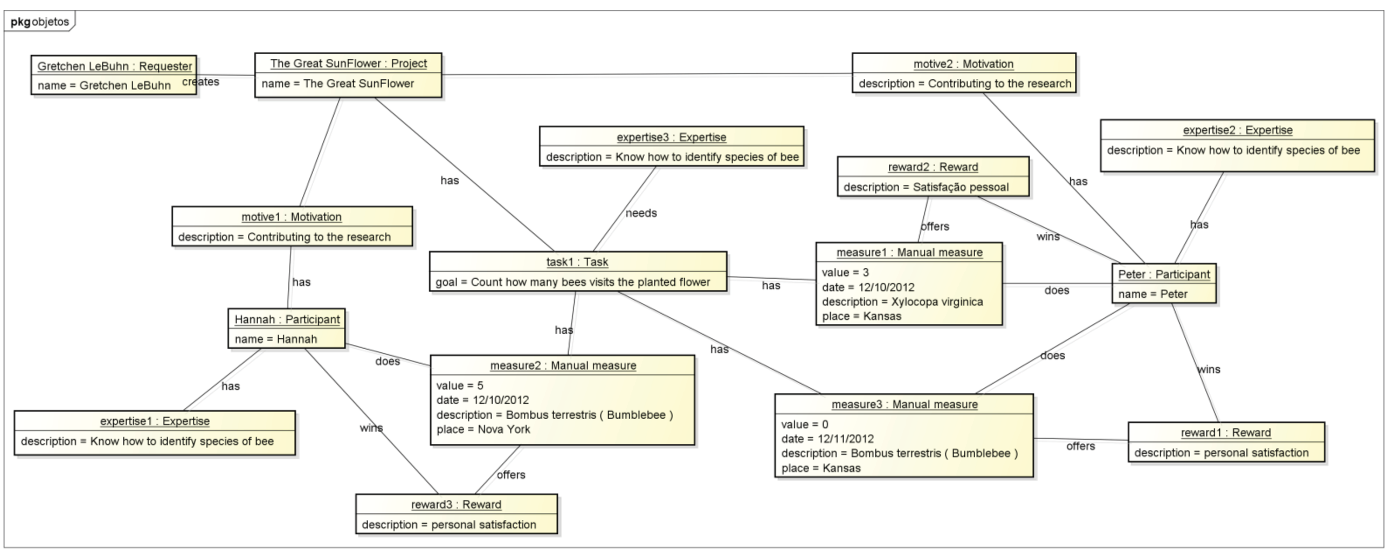

Figure 4. Object diagram applied to the Great SunFlower project.

In the diagram we can see that there is a participant named "Gretchen LeBuhn". This requester creates a project called "The Great SunFlower" whose main goal was broken into a single task: "Count how many bees visit the planted flower". To perform this task an expertise is requested: "Know how to identify species of bees." To facilitate the understanding, we created only two participants in the project. Participants have a particular motivation to contribute to the project. In this case, participants have a motivation described as "Contributing to the research".

To understand how the dynamics of contribution to the project operates, observe participant "Peter" who has the expertise to assist in the requested task. In the example, Peter has contributed three times to the project. In each contribution, he got a reward described as "Personal Satisfaction". The example project has no explicit data validation. Therefore, we cannot say that there is an evaluator and an assessment of the performed collection. For this reason no objects were presented with the role of *evaluator*.

V. Conclusions and Future Work

The main goal of this article was to present a meta-model that aims to generalize and study the solutions relating to data collection and participatory sensing, and to propose a framework to facilitate the understanding of current tools, and to facilitate the construction of new ones.

In the current context of Data Collection and Participatory Sensing Crowdsourcing Systems, many of the existing studies are based on specific cases and end up with special features that cannot be reused in other projects. With the creation of a meta-model, the steps and concepts that are common to projects of this nature can be reused and, should the need arise, be extended for a specific case.

Future work includes investigating whether other projects related to participatory sensing and data collection are consistent with the meta-model proposed here and, if so, find out where it can be improved.

References

[1] Howe J. The rise of crowdsourcing. Wired 2006; 14(6).

[2] Zhao, Y., Zhu Q. (2012). Evaluation on crowdsourcing research: Current status and future direction.

[3] Geiger, D. (2011). Crowdsourcing Information Systems – A Systems Theory Perspective.

[4] Vukovi, M. (2009). Crowdsourcing for Enterprises.

[5] Estelles-Arolas, E., and Gonzalez-Ladron-de-Guevara, F., “Towards an integrated crowdsourcing definition,” Journal of Information Science, vol. 38, no. 2, pp. 189–200, Mar. 2012.

[6] Schneider, D., Moraes, K., Souza, J.M., , Esteves, M. G. P. CSCWD: Five characters in search of crowds. Proceedings of the 16th International Conference on Computer Supported Cooperative in Design (CSCWD 2012), 2012, pp. 634-641.

[7] Rouse, Anne C., "A Preliminary Taxonomy of Crowdsourcing" ACIS 2010 Proceedings. Paper 76, 2010. http://aisel.aisnet.org/acis2010/76

[8] Schenk, E. and Guittard, C. “Towards a characterization of crowdsourcing practices,” Journal of Innovation Economics, no. 1, pp. 93–107, 2011.

[9] Uchoa, A. P., Esteves, M. G. P., and Souza, J. M. “Mix4Crowds - Towards a framework to design crowd collaboration with Science,” in 2013 IEEE 17th International Conference on Computer Supported Cooperative Work in Design (CSCWD), 2013, pp. 61–66.

[10] Franzoni, C. and Sauermann, H. “Crowd Science: The Organization of Scientific Research in Open Collaborative Projects,” SSRN eLibrary, Nov. 2012.

[11] Wiggins, A. and Crowston, K. “Goals and Tasks: Two Typologies of Citizen Science Projects”, Proceedings of the 45th Hawaiian International Conference on Systems Sciences (HICSS-45), 2012.

[12] Alabri, A. e Hunter, J. “Enhancing the Quality and Trust of Citizen Science Data,” 2010, p. 81-88.

[13] Wiggins, A. and Crowston, K. “Developing a Conceptual Model of Virtual Organizations for Citizen Science,” International Journal of Organizational Design and Engineering, vol. 1, pp. 148–162, Sep. 2010.

[14] Nov, O., Arazy, O. and Anderson, D. "Crowdsourcing for science: understanding and enhancing SciSourcing contribution", Position paper: ACM CSCW 2010 Workshop on the Changing Dynamics of Scientific Collaborations, 2010.

[15] Rotman, D., Preece, J., Hammock, J., Procita, K., Hansen, D., Parr, C., Lewis, D. and Jacobs, D. “Dynamic changes in motivation in collaborative citizen-science projects”, 2012, p. 217.

[16] Nov, O., Arazy, O., Anderson, D. Technology-Mediated Citizen Science Participation: A Motivational Model. Proceedings of the AAAI International Conference on Weblogs and Social Media (ICWSM 2011), pp.249-256, Barcelona, Spain, July 2011.

[17] ITU World Telecommunication/ICT Indicators database. The key 2006-2013 ICT data for the world, by geographic regions and by level of development. http://www.itu.int/en/ITU-D/Statistics/Pages/stat/default.aspx

[18] Christin, D., Reinhardt, A., Kanhere, S. S., and Hollick, M. “A survey on privacy in mobile participatory sensing applications”, Journal of Systems and Software, vol. 84, no 11, p. 1928–1946, Nov. 2011.

[19] Burke, J. A., Estrin, D., Hansen, M., Parker, A., Ramanathan, N., Reddy, S. and Srivastava, M. B. “Participatory sensing”. In World Sensor Web Workshop - WSW’06 at SenSys ’06, October 31, 2006, Boulder, Colorado, USA, 2006.

[20] Goodchild, M. F. “Citizens as sensors: Web 2.0 and the volunteering of geographic information,” GeoFocus (Editorial), vol. 2, pp. 24–32, 2007.

[21] The Great SunFlower Project, http://www.greatsunflower.org.

[22] eBird, http://ebird.org.

[23] Estrin, D. “Participatory sensing: applications and architecture [Internet Predictions],” IEEE Internet Computing, vol. 14, no. 1, pp. 12–42, Jan. 2010.

[24] Goldman, J., Shilton, K., Burke, J., Estrin, D., Hansen, M., Ramanathan, N., Reddy, S., Samanta, V., Srivastava, M. and West, R. “Participatory Sensing: A citizen-powered approach to illuminating the patterns that shape our world,” Foresight & Governance Project, White Paper, 2009.

[25] Cuff, D., Hansen, M., King, J. Urban Sensing: Out of the Woods. Communications of the ACM, Volume 51, Number 3, pp. 24-33, 2008.

[26] NoiseTube, http://www.noisetube.net.

[27] Tweddle, J.C., L. Robinson, D., Pocock, M.J.O., Roy, H.E. (2012). Guide to citizen science: developing, implementing and evaluating citizen science to study biodiversity and the environment in the UK.

[28] Jordan, R. C., Brooks, W. R., Howe, D. V., Ehrenfeld, J. G.. Evaluating the Performance of Volunteers in Mapping Invasive Plants in Public Conservation Lands, 2010.

[29] Antelio, M., Esteves, M. G. P., Schneider, D., Souza, J.M. Qualitocracy: A data quality collaborative framework applied to citizen science. Proceedings of the International Conference on Systems, Man and Cybernetics (IEEE SMC 2012), 1012.