17th International Conference on Environmental Science and Technology

Athens, Greece, 1 to 4 September 2021

Exploitation of Crowdsourcing Tools and Earth Observation data: A Systematic Literature Review

TSIAKOU D.[1,*], TSIMIKLIS G.[1], TSIAKOS V.[1], KARAGIANNOPOULOU K.[1] and AMDITIS A.[1]

[1]Institute of Communication and Computer Systems, 9, Iroon Politechniou str., 15773 Athens, Greece

*corresponding author:
e-mail: dimitra.tsiakou@iccs.gr

Abstract Crowdsourcing is a method gaining ever wider use in practice and leverages human intelligence to solve problems in a considerable number of study fields. Howe (Howe 2006) coined the concept defining: "Crowdsourcing represents the act of a company or institution taking a function once performed by employees and outsourcing it to an undefined network of people in the form of an open call." This systematic review aims to understand the Crowdsourcing tools and Earth Observation (Satellite, aerial & in-situ) data and their contribution to environmental conservation and sustainability. The review involved 29 papers with particular focus on Technology Readiness Level (TRL), data fusion methods applied and topics such as types of users and their incentives and tools used to engage users and to collect the data. This article provides a glimpse of the Citizen Science (CS) data collection combined with Earth Observation data and explores the development of this swiftly emerging and evolving subject. We discuss two central implications in terms of research-implementation spaces for Crowdsourcing tools and EO data fusion: (1) the need to centralize the role of the people in order to have more accurate outcomes and (2) the most commonly used tools for crowdsourcing. This systematic literature review provides collective insights about the most commonly used tools for crowdsourcing that characterize well it's one of the monitored domains, concluding that in complex environments remote sensing still exceeds citizen information. Our review identified limitations and recommendations derived from the monitored papers, that will improve the efficiency and provide an opportunity to look at multifaceted problems from numerous standpoints.

Keywords: Remote Sensing, Earth Observation, Citizen Science, Crowdsourcing tools

1. Introduction

In this systematic literature review we collect and synthesize previous research about Earth Observatories and Crowdsourcing tools. We apply a double and a two level of screening method in order to prioritize and utilize the papers gathered for this systematic literature review. We begin by contextualizing our review through the Crowdsourcing tools and Earth Observation data fusion and focus on how these methods contribute to the integrated environmental monitoring and by extension environmental sustenance. In Methodology section we describe our approach to this systematic literature review and we funnel on the TRL, the incentives for users, the engagement tools and platforms used in each research, the data fusion method and finally the proposed future research. We also discuss about the implications in terms of research implementation spaces for citizen participation. Finally, we close with a call for more research on the benefits of CS data and the huge potential arising from the combination of Crowdsource tools and EO data and how we can improve the current situation in terms of data fusion methods for an optimal result.

2. Background

Howe (Howe 2006) presents "crowdsourcing" as "an idea of outsourcing a task that is traditionally performed by an employee to a large group of people in the form of an open call". The aforementioned author describes crowdsourcing as an act of taking a task traditionally performed by a designed agent and outsourcing it by making an open call to an undefined but large group of people. A report by Hosseini, Phalp, Taylor, and Ali (Hosseini 2014) focused on defining four essential pillars of crowdsourcing. The crowd that consists of workers who take part in a crowdsourcing project, the crowdsourcer which is the entity that plans, coordinates, and controls the crowdsourcing project, the crowdsourcing task consists of the activities to be solved by the workers and the fourth pillar refers to the crowdsourcing platform, which manages the crowd and the tasks. Another interesting subject is what drives potential users to take part in crowdsourcing. Moreover, a systematic review that examines the state of knowledge and trends in the peer-reviewed literature related to the use of smartphone technologies for community and citizen science environmental monitoring is a paper by Andrachuk (Andrachuk 2019), which leads to the conclusion that the future of environmental monitoring with smartphones is inherently unpredictable. Technological innovations will continue to drive what will be possible.

We decided to conduct a systematic literature review because this type of review has the essential tools for summarizing evidence accurately and reliably, collates all empirical evidence that fits pre-specified eligibility criteria to answer a specific research question and uses explicit, systematic methods that are selected with a view to minimizing bias (Liberati 2009). Besides, we wanted to identify the current research status of the combination of Crowdsourcing tools and Earth Observation data, prescribe directions for future investigation and fill the gap in some of the aforementioned gaps in literature.

3. Methodology

3.1. Search Strategy

The main scope of this study was to examine and critically assess the recent literature, in terms of the utilization of the Crowdsourcing tools and Earth Observation (Satellite, aerial & in-situ) data and their combination in order to contribute to environmental conservation and sustainability. Additionally, this review systematically synthesizes the existing literature and map evidence underpinning the research question, and finally highlights any gaps in the literature and future research work that could be implemented. We adopted a systematic literature review procedure searching English peer-reviewed journals in the four electronic literature databases of Google Scholar, Scopus, ScienceDirect and Taylor and Francis, excluding the grey literature (conference papers, presentations, book chapters, commentary, extended abstracts, etc.). The searching period was oriented between the 1 January of 2015 until the 31 May of 2020 considering the research findings of (Fritz, Fonte & See 2017) and (Saralioglu & Gungor 2019), which clearly depict the rapid increase of the amount of the publications related to Earth Observation (EO) and Citizen Science (CS) fusion methods.

3.2 Inclusion/Exclusion criteria

A rapid research strategy was implemented adopting the methodological framework proposed by Tranfield (Tranfield, Denyer & Smart 2003). Only studies that meet all the inclusion criteria specified in the review protocol and which manifest none of the exclusion criteria need be incorporated into the review. The criteria used to select studies for inclusion in a review should be consistent with the focus. They should be explicit to base the review on the best-quality evidence. Similarly, the criteria that are used to assess the validity of the studies that are included should be explicit to minimize biased assessments and weighting of the included studies. After the implementation of in/out criteria, the papers that are mainly oriented in EO data and Citizen Science tools and the architecture development of the system is mostly described, a final 29 papers selected for analysis.

3.3 Data collection and handling

Across the literature there were some noteworthy tendencies correlated to what was being monitored, who was doing the monitoring, and how observations were recorded.

Technology readiness levels (TRLs) are a method for estimating the maturity of technologies during the acquisition phase of a program, developed at NASA during the 1970s. The use of TRLs enables consistent, uniform discussions of technical maturity across different types of technology (Héder 2017). Two papers considered a TRL 4 because they describe the basic technological components, which are integrated to establish that they will work together, and examples have been made which include integration of “ad hoc” hardware in the laboratory. The main difference between these two papers is that, even if they are TRL 4 only one of them has materialized an infrastructure and established benchmarks to regularly test the accessibility, performance and quality of existing activities. On the other side both of them need more research and engineer a full-scale. As for TRL 6, we included the papers that have been tested in a relevant environment. Examples include testing a prototype in a simulated operational environment. These technology readiness level papers have limitation as far as the definition of process variables and need more precise cost estimations. TRL 7 represents a major step up from TRL 6, requiring demonstration of an actual system prototype in an operational environment. The commonality, between these papers, is the lack of interface with platform validation. In the TRL 8 technology has been proven to work in its final form and under expected conditions. In almost all cases, this TRL represents the end of true system development. Examples include developmental test and evaluation of the system in its intended system to determine if it meets design specifications. Specifically, they analyze topics such as energy-efficiency data collection, phenological activity, in situ forest biomass measurements and a platform aiming to raise awareness. Even so, they are not tested in all environmental conditions and they weren’t performed in more than a few demonstrations. A TRL 9 application is actual application of the technology in its final form and under mission conditions, such as those encountered in operational test and evaluation, including examples of using the system under operational mission conditions. These papers display applications about ashfall reports, landslide data collection and video annotation tool for rapid mapping. The one thing they have in common is the lack of experience in terms of a full range of operating conditions.

Citizens can get involved in different stages of the scientific process: development of hypotheses, methodology design, data collection, data analysis and dissemination of data (Land-Zandstra 2016). In terms of user motivation, we particularly occupied with the citizen participation, if there was any kind of payment for the users or if the participation of the latter it was due to a natural interest. We also examined who were the intended users, whether they were experts or non-experts and in some papers both experts and citizen, included interest groups, employees, volunteers from a specific organization or local experts with education background connected with the research. Not all end users are the same and do not have the same skills and needs, as the types of users fall into many ranges, as mentioned above. Regardless of the category to which the users belong, everyone has the opportunity to participate in the research leading to actively include broad audiences to problem solving, establish communities and produce social innovation through crowdsourcing initiatives. To conclude, in order to successfully manage crowdsourcing communities, an understanding of different user roles and behaviors (e.g. contributions, knowledge sharing and social interactions) within the community needs to be achieved (Pratihast 2014). Additionally, we distinguished 4 different approaches concerning the future research, such as more human resources, scale up, change of strategy, trust in the source of information, gamification mechanism and payment.

Another objective of this paper is to better understand the tools that have been used in each paper through a systematic review. We focused on the different types of technology used for the purpose of crowdsourcing. More specifically we dealt with the tools used for engagement and data collection and the open platforms that facilitated their research. In the plethora of articles, the most common tools for citizen engagement were either Smartphones or Cameras followed by Social Media, Graphical User Interface (GUI) and WebGIS and sometimes they used crowdsourcing frameworks such as PyBossa and MicroMappers. One of the significant characteristics of citizen science is the wide coverage of participation of people from the public. It is based on the involvement of a large number of volunteers in the research process, mainly during the data-collection stage (Hochachka 2012). This phenomenon is referred as 'crowdsourcing', represented by the success of Volunteered Geographic Information (VGI) (Goodchild 2007). The other characteristic is that, with the help of modern sensors, e.g., sensors embedded in a smartphone which mainly reflect human activities, citizen science is able to expand the coverage of data content that are useful to advance the understanding of environmental science or human-environmental interactions from a human-centric perspective (Srivastava, Abdelzaher & Szymanski 2012; Pei 2013).

The other aspect of the tools used in the research papers was for the purpose of data collection. The tools that stood out among the papers were Unmanned Aerial Vehicles (UAVs), sensors, satellite imagery and OpenStreetMap or Google Maps. Other data collection tools, not so common in these research papers, were PostgreSQL or MySQL, GeoEye1 or GeoDIVA and MongoDB. The main purpose of using all these data collection tools is to lead to a more effective monitoring which will require an integrated approach, where detailed community-based observations are combined with remote sensing (Pratihast 2014). Thus, a Big Data approach is needed for efficient storage, management, analysis and visualisation of the data. In addition, the usefulness of the platforms, like Amazon Mechanical Turk (AMT), Geocrowd, MS4ME, PAYSAGES, DIYlandcover Open AQ, was significant. These platforms help to overcome the constraints of cost and production time, while retaining the advantages of human interpretation skill and connect them in an online marketplace. Also aggregate and share (via an open Application Programming Interface (API)) high-quality data from multiple official sources around the world. Importantly, these platforms check each data source for updated information, which guarantee that the data will be available from the platform almost immediately after they are published by the original data providers.

Data fusion techniques have been extensively employed on multisensory environments with the aim of fusing and aggregating data from different sensors. The fusion technique used in this review was the classification based on the Abstraction Levels (Castanedo 2013).

4. Conclusions

This systematic literature review explores the crowdsourcing and EO literature in order to fill the existing gap in enhanced tools, guiding principles and robustness of mechanisms for using the data intelligently. We have provided collective insights about the most commonly used tools for crowdsourcing that characterize well it's one of the monitored domains, concluding that in complex environments remote sensing still is preferred over citizen information. Our review identified limitations and recommendations derived from the monitored papers, that will improve the efficiency and provide an opportunity to look at multifaceted problems from numerous standpoints. By conducting this systematic literature review we hope that will help underpin and inform future research concerning EO data and crowdsourcing tools and how to combine them in the best way, in order to have a more holistic approach. A set of main results is listed below.

The goal of using data fusion in multisensor environments is to obtain a lower detection error probability and a higher reliability by using data from multiple distributed sources. As far as the type of users, it is essential to mention, that the papers who used only expertise citizens needed some fundamental understanding of the monitoring themedue to their complexity, that the non-experts aren't equipped with. These systems enable the experts to participate more keenly and so provide more accurate information. In addition, papers that engage citizen with no expertise led to the conclusion that their training would ultimately provide more accurate results. Even if most of the papers offer the opportunity of involvement in the process, there is still a substantial amount of papers that didn't rely on crowdsourced data. One thing that was noticed, in these papers, was that they engaged with highly dynamic time-variant environments and did not rely on accurate and mathematically solvable system models, therefore they reflected that remote sensing is still a more reliable source of information in the subject of accuracy compared with information derived from citizens.

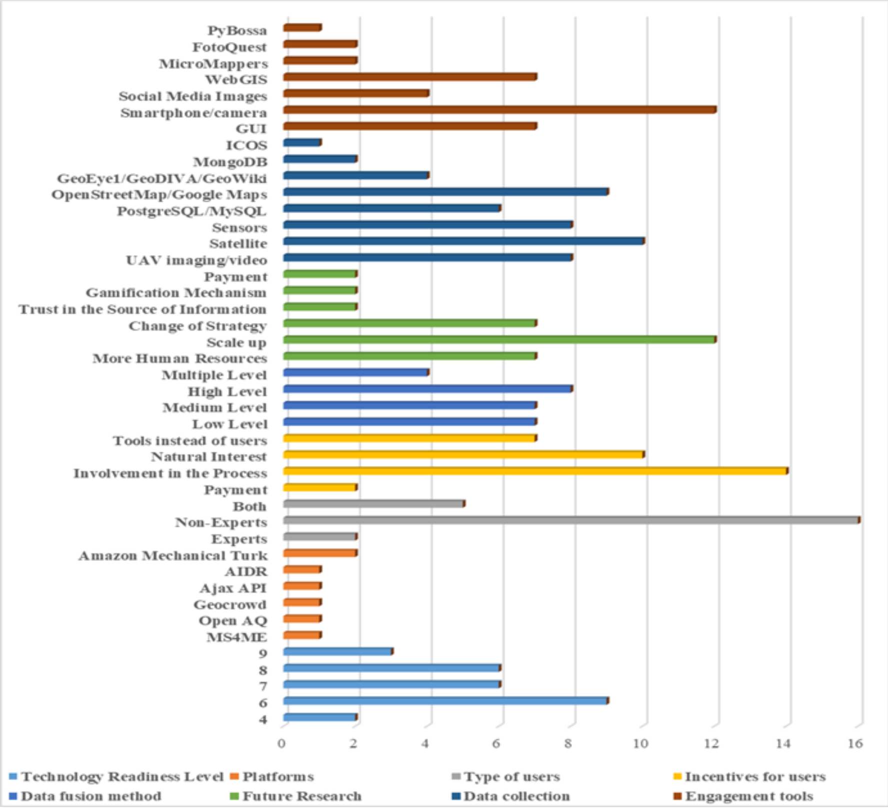

Figure 1. Summary of tendencies correlated to papers being monitored (n=29)

References

Andrachuk, M., Marschke, M., Hings, C. & Armitage, D., 2019, 'Smartphone technologies supporting community-based environmental monitoring and implementation: a systematic scoping review', *Biological Conservation*, 237(April), 430–442.

Castanedo, F., 2013, 'A review of data fusion techniques', *The Scientific World Journal*, 2013.

Fritz, S., Fonte, C.C. & See, L., 2017, 'The role of Citizen Science in Earth Observation', *Remote Sensing*, 9(4), 1–13.

Goodchild, M.F., 2007, 'Citizens as sensors: The world of volunteered geography', *GeoJournal*, 69(4), 211–221.

Héder, M., 2017, 'From NASA to EU: The evolution of the TRL scale in Public Sector Innovation', *Innovation Journal*, 22(2), 1–23.

Hochachka, W.M., Fink, D., Hutchinson, R.A., Sheldon, D., Wong, W.K. & Kelling, S., 2012, 'Data-intensive science applied to broad-scale citizen science', *Trends in Ecology and Evolution*, 27(2), 130–137.

Hosseini, M., Phalp, K., Taylor, J. & Ali, R., 2014, 'The four pillars of crowdsourcing: A reference model', *Proceedings - International Conference on Research Challenges in Information Science*.

Howe, J., 2006, 'The Rise of Crowsourcing', *Critical Inquiry*, 35(4), 996.

Land-Zandstra, A.M., Devilee, J.L.A., Snik, F., Buurmeijer, F. & Broek, J.M. Van Den, 2016, 'Citizen science on a smartphone: Participants' motivations and learning', *Public Understanding of Science*, 25(1), 45–60.

Liberati, A., Altman, D.G., Tetzlaff, J., Mulrow, C., Gøtzsche, P.C., Ioannidis, J.P.A., Clarke, M., Devereaux, P.J., Kleijnen, J. & Moher, D., 2009, 'The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: Explanation and elaboration', *PLoS Medicine*, 6(7).

Pei, L., Guinness, R., Chen, R., Liu, J., Kuusniemi, H., Chen, Y., Chen, L. & Kaistinen, J., 2013, 'Human behavior cognition using smartphone sensors', *Sensors (Switzerland)*, 13(2), 1402–1424.

Pratihast, A.K., DeVries, B., Avitabile, V., Bruin, S. de, Kooistra, L., Tekle, M. & Herold, M., 2014, 'Combining satellite data and community-based observations for forest monitoring', *Forests*, 5(10), 2464–2489.

Saralioglu, E. & Gungor, O., 2019, 'Use of crowdsourcing in evaluating post-classification accuracy', *European Journal of Remote Sensing*, 52(sup1), 137–147.

Srivastava, M., Abdelzaher, T. & Szymanski, B., 2012, 'Human-centric sensing', *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 370(1958), 176–197.

Tranfield, D., Denyer, D. & Smart, P., 2003, 'Towards a Methodology for Developing Evidence-Informed Management Knowledge by Means of Systematic Review* Introduction: the need for an evidence-informed approach', *British Journal of Management*, 14, 207–222.