Citizen science, data science and education: how to support teacher's inspiration during the learning activities design with technology enhance learning

Miriam Calvera-Isabal, Patricia Santos and Davinia Hernández-Leo

Pompeu Fabra University, Plaça de la Mercè, 10-12 08002, Barcelona, Spain

Abstract

This research will investigate the potential that Citizen Science (CS) projects data has in a formal educational context to learn. CS involves citizens in scientific research generating knowledge and scientific results. Citizens participate in CS projects conducting many activities developing skills, interest in science or scientific literacy. These projects share their information in online repositories also called CS platforms to inform citizens about the aim of the project, their advances on research or how to take part in it. Although the connection between CS and education has been explored, remains to be understood and must be made more explicit. Web scraping and data mining methods have the potential to obtain CS projects data available online and analyze it to extract conclusions. The main aim of this research is to understand how to analyze and visualize CS data in a technology enhanced learning (TEL) tool to support educators during the learning design of educational activities. It is also intended to improve scientific knowledge and to bring official science closer to educational environments.

Keywords [1]

Citizen science, technology enhanced learning, data science, web scraping

1. Introduction

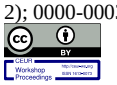
Scientific literacy and critical thinking are important skills for youth to raise awareness and to address today's societal challenges [1]. Citizen science (CS) by involving the general public (from youth to adults) in scientific work, might enhance public understanding of science, contribute to Science, Technology, Engineering and Mathematics (STEM) career motivation and promote values like ecology or respect for the natural environment [2, 3, 4, 5]. Scientific research in CS is organized into projects, in which participants carry out many types of activities, depending on the typology of participation defined [6]. On taking part of the project, citizens might develop: interest in science and the environment, self-efficacy for science and the environment, motivation for science and the environment, knowledge of the nature of science, skills of science inquiry and behavior and stewardship [7].

In order to promote and support CS projects, there are many online platforms that acts as repositories containing information about CS projects and other resources (i.e., EU-Citizen science platform [8]). The platforms share descriptive metadata about the CS project but only some of those follow metadata standards to show it in a structured way [9]. The Project Metadata Model (PMM) which is part of the Public Participation in Scientific Research (PPSR) metadata standard [10] describes project characteristics. Data shared in online platforms has the purpose of explaining to a particular audience what a citizen science

Proceedings of the Doctoral Consortium of Sixteenth European Conference on Technology Enhanced Learning, September 20–21, 2021, Bolzano, Italy (online).

EMAIL: miriam.calvera@upf.edu(A.1); patricia.santos@upf.edu (A. 2); davinia.hernandez-leo@mail.com (A. 3)

ORCID: 0000-0003-4117-6953 (A. 1); 0000-0002-7337-2388 (A. 2); 0000-0003-0548-7455 (A. 3)



project is about. These texts, in general, use special language related to science and contain information about the project aim or an explanation of the research. Web scraping techniques in combination with data mining methods have been used to extract and analyze data from online sites to obtain conclusions in many fields of science [11]. The use of data extracted from texts, that contain information about science, in combination with context of real problems and instructions has the potential of improving scientific literacy [12].

Designing, planning and developing activities requires pedagogical design capacity, design competencies and design expertise of teachers during the learning design process [13]. It is expected that teachers have subject matter knowledge (SMK) and pedagogical content knowledge (PCK) to develop learning designs [14] although it's possible that they need information, materials or training about specific topics. Citizen science projects support participants by developing educational materials (i.e., guides, posters, manuals, videos or podcasts) or sharing results or information about the project that teachers can use to inspire them during the learning design [15, 16]. Our hypothesis is that facilitating data access for educational purposes would have a positive impact on teachers' scientific knowledge and pedagogical skills that would influence on student's scientific literacy and its relation to science [17].

Technology tools that support teachers during the learning designs help them to ensure a better learning experience [18]. Although these tools try to cover teacher's needs like containing relevant resources (i.e Open educational resources (OER) [19] or supporting them during the design of activities, there are barriers or factors that affect the adoption or usage of the tool [20]. The benefits of involving teachers as designers during the design process of technology enhancing learning (TEL) are numerous: from improving student's learning, their own learning about technology or motivation and commitment of using technology and implementing it [21, 22]. The research conducted in this thesis will be considered to (co)design (with educators) and implement a digital tool to inspire and support teachers. Exploring a tool that uses and shows data about CS projects during the learning design process in a formal education context, will derive from improving SMK that will foster on developing PCK.

2. Thesis statement

The aim of this research is to better understand the connections between education and citizen science and to identify the potential that data from CS has to support the design of learning activities. The main research question is: *"How can data science methods be effectively used to gain understanding of the potential that web data about citizen science projects have to inspire teachers in designing for science learning outcomes?".* The related sub research questions are the followings:

- How web scraping and data mining methods can be used to collect/analyze data online about citizen science projects?
- How data from CS projects can be presented/analyzed in relation to their potential to support learning outcomes in formal settings?
- What features and content should be integrated into a digital tool to inspire teachers in the design process of scientific learning activities based on citizen science?

3. Methodology

The methodology selected for this study is Design-Based Research (DBR), it is expected to meet the objective of contributing to understanding the connections between citizen science and educational research. First DBR is applied to analyze the literature exploring the connection between citizen science and education, and identifying gaps. This is an iterative process where the educator's needs will be taken into account involving them in the design process and evaluation of the prototype. The tool will be used in a real educational context to assess whether it meets the teachers' needs and performs the function of supporting teachers during the learning designs. Finally, from the lessons learned, actions and recommendations will be proposed to connect citizen science with education in formal settings.

Mixed methods will be used in an iterative process of testing and refinement cycles of the design process [23, 24]. Both qualitative and quantitative methods will be applied to obtain

results for the research questions defined above:

- *Literature review and exploration* to explore previous work done and how data is structured in online platforms
- *Web scraping methods* will be used to extract CS projects metadata, from online platforms using web scraping tools [25]
- *Workshops with key stakeholders* will be conducted to identify citizen science community interest on data and teachers needs during the design of learning activities
- *Data mining Methods for data analysis* will be applied to the data extracted and to explore problems related to the ways of reporting data
- *TEL tool* development

The PPM model from the PPSR metadata standard will be used during this research to structure data extracted in a database. Data mining methods will be applied to analyze the data stored in the database and obtain conclusions about CS and its connection to education.

Moreover, collaborative partnerships with experts will be built, to apply research findings aligned to the teachers' needs, to have a positive effect on teaching and learning [26]. To fulfil the main question and research objectives, many activities have been designed, revised and evaluated during the research. Interviews and workshops with teachers and key users to define and validate data needs, co-design the tool and evaluate and test the final design of the tool will be the main actions developed during the evaluation process.

4. Research plan, possible limitations and risks and progress done so far

Part of the activities defined to achieve research objectives are framed within the CS Track project (European Union's Horizon 2020 research and innovation program under grant agreement No 872522) [27]. The work to be done during the research is divided into four different phases. There are planned activities to address each research objective by applying research methods defined above:

1. Initial phase

- *Literature review* about citizen science, its connection to education, TEL, data science methods, web scraping methods and teacher's learning designs process.
- *Research design and definition*

2. Data extraction

- *Explore and study how citizen science projects information is shared online*. At this stage, there is to explore how citizen science is conducted online and which information about citizen science projects is available
- *Crawling citizen science projects data from online platforms*
- *Defining descriptors/categories to classify the data and analyzing existing data* to detect issues with the data. Analyze metadata standards to be used to store classified data extracted and define characteristics of the central database
- *Applying computational techniques to address the data problem*

3. Data analysis

- *Defining and applying mixed methods to analyze the data*
- *Creating post-processed data datasets of interest to the community*

4. Tool development and validation *(iterative process)*

- *Identifying how teachers design learning activities and their needs* during the process
- *Designing tool functionalities with teachers*
- *Tool development*
- *Evaluating the tool with teachers to validate the design proposed and the usability*. It will also be evaluated if the tool fulfils the objective of inspiring teachers and students

Data extraction and crawler development was done during this first year (Initial phase). An initial version of the database has been created and data has been initially analyzed (data extraction). Some quantitative methods and data mining methods have been applied so far with initial outcomes (data analysis). Hereinafter, data analysis methods will be applied to the data stored (data analysis) and, in parallel, the gathering of key user tool requirements will begin before the end of the

second year. The outcomes of data analysis will be available at the middle of third year at the same time the tool will be being developed. Finally for the first quarter of last year the tool will be tested, improved and the potential will be validated by key users (Tool development and validation).

4.1 Identification and prevention of possible limitation and risks

This study will be focused on analyzing science teaching in formal education needs (primary and secondary school levels) and will not take into account other educational levels due to lack of time and resources to cover all. This research will not analyze either the impact on student's learning because of the same limitations. Nevertheless, teacher's will be asked about their experience and perceptions of in which way students have learned after developing activities designed by them. Finally, finding participants to join co-design sessions is challenging so we will do an open call to invite secondary teachers to join the case studies sessions. In addition to this, our research plan includes the organization of workshops with pre-service teachers that are studying a master degree in UPF for teaching science subjects in secondary schools.

4.2 Progress done so far

Being part of the CS Track project, gives us the opportunity to be in touch with the citizen science community and participate in conferences. It was on *"Knowledge for Change: A decade of Citizen Science (2020-2030) in support of the SDGs"* [28], which took place on 14th-15th October 2020 online and in Berlin where we presented our advances on data extraction and its potential on SDG [29]. Furthermore, during the *"CitSciVirtual"* [30] conference which took place online throughout May 2021, we presented a poster about database development [31] and a workshop where we presented metadata stored and got feedback from the community about their data needs.

A preliminary study of how data mining methods allow us to know more about citizen science and its connection with education has been accepted for the CELDA conference [32]. This proof of concept was developed to obtain initial results of the research, design research and select technology will be used.

Regarding workshops with science secondary teachers and tool development (RO3), we have received *"Grant for activities to increase the social impact of research"* from UPF [33] to conduct it during 2021. As planned, it will be done during the last quarter of second year. As part of the CS-Track project, workshops will also be held with key stakeholders (teachers and/or CS participants) Furthermore, as part of *"Makers a les aules (20-21)"* program [34], it has developed the first version of the tool to be tested with students (8 to 10 years old) and primary school teachers (Figure 1). The tool contained information about some citizen science projects related to the subject of the activity (i.e., sea pollution). It allowed teachers and students to read more about the project itself and the tools used by participants.

During the co-design activity with teachers and one of the activities with students, they explored the tool and the information shared about citizen science projects and it was evaluated at the end of the program the influence of this information on the activity designed and results. The influence of citizen science projects data visualization has been analyzed on the subject of *human-machine interaction* with UPF undergraduate students. The tool used by students contained information about citizen science projects and was assigned to a sustainable development goal (SDG). This relation was established because of the issue addressed by the activity. Apart from the description and tools used by participants, information from the web from where the data has been extracted was also added in case students wanted to explore the project in more detail (Figure 2).

Figure 1: First version of the tool used in *Makers a les aules* (2020-2021) with primary school students and teachers

Figure 2: First version of the tool used in *Human-computer interaction* subject with undergraduate students at UPF

5 Expected contributions

The proposed study will make contributions to the TEL and CS fields on the following aspects (Figure 3):

1. *Technical architecture of web scraping tools and data management methods* to analyze and describe how citizen science information is available online. As a result, datasets of citizen science projects will be developed.
2. The study will provide *evidence on how computational data analytics methods can be applied* in the context on citizen science to broaden the knowledge about this field and its relation to education
3. *Proposal and development of a technical environment* to extract and process data and tool to show it addressing the main RQ.

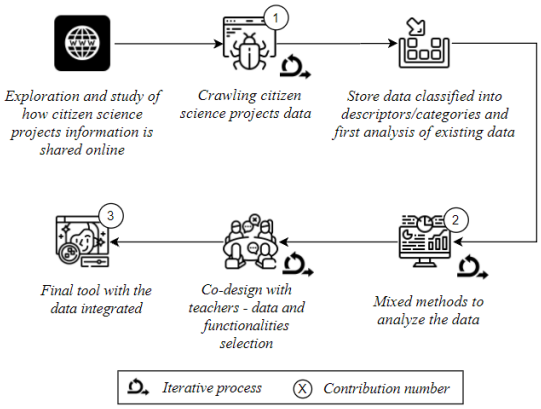


Figure 3: Data flow and expected contributions (identified with the numbers on the list).

This research aims to contribute to TEL and CS fields but mainly tries to impact on teaching/learning science in formal education settings. This study will be focused on analyzing science teaching in formal education needs during the conceptualization phase of the learning design process.

Lessons learned from the data scraping and data mining process connecting CS and Education, will be shared in the form of guidelines, datasets and other contributions. A tool will be co-designed with teachers as an inspirational resource to get inspiration in regards to certain scientific topics, and as a facilitator to help them to design activities about science. Furthermore, it is expected that the data exploration process will potentially improve teachers SMK and PCK.

6 Acknowledgements

This work has been partially funded by the CS Track project, European Union's Horizon 2020 research and innovation program under grant agreement No 872522., a Grant for activities to increase the social impact of research" from UPF, the EU Regional Development Fund and the National Research Agency of the Spanish Ministry of Science and Innovation under project grants TIN2017-85179-C3-3-R, PID2020-112584RB-C33. D. Hernández-Leo (Serra Húnter) acknowledges the support by ICREA under the ICREA Academia program.

7 References

[1] Siarova, H., Sternadel, D. & Szőnyi, E. 2019, Research for CULT Committee – Science and Scientific Literacy as an Educational Challenge, European Parliament, Policy Department for Structural and Cohesion Policies, Brussels

[2] Vohland, K., Land-Zandstra, A., Ceccaroni, L., Lemmens, R., Perelló, J., Ponti, M., ... & Wagenknecht, K. (2021). The Science of Citizen Science Evolves. Chapter 1. in Vohland, K. et al (Eds.).(2021) The Science of Citizen Science. Springer. pp 1-12. doi: 10.1007/978-3-030-58278-4.

[3] Strasser, B., Baudry, J., Mahr, D., Sanchez, G., & Tancoigne, E. (2019). " Citizen Science"? Rethinking Science and Public Participation. Science & Technology Studies, 32 (ARTICLE), 52-76.

[4] Hiller, S. E., & Kitsantas, A. (2014). The effect of a horseshoe crab citizen science program on middle school student science

performance and STEM career motivation. School Science and Mathematics, 114(6), 302-311.

[5] Kobori, H., Dickinson, J. L., Washitani, I., Sakurai, R., Amano, T., Komatsu, N., ... & Miller-Rushing, A. J. (2016). Citizen science: a new approach to advance ecology, education, and conservation. Ecological research, 31(1), 1-19.

[6] Bonney, R., Ballard, H., Jordan, R., McCallie, E., Phillips, T., Shirk, J., & Wilderman, C. C. (2009). Public Participation in Scientific Research: Defining the Field and Assessing Its Potential for Informal Science Education. A CAISE Inquiry Group Report. Online Submission.

[7] Phillips, T., Porticella, N., Constas, M. and Bonney, R., 2018. A Framework for Articulating and Measuring Individual Learning Outcomes from Participation in Citizen Science. Citizen Science: Theory and Practice, 3(2), p.3. URL: http://doi.org/10.5334/cstp.126

[8] EU-Citizen.Science platform. (2021, July 29). eu-citizen.science. URL: https://eu-citizen.science/ .

[9] Wagenknecht, K., Woods, T., Sanz, F. G., Gold, M., Bowser, A., Rüfenacht, S., ... & Piera, J. (2021). EU-Citizen. Science: A Platform for Mainstreaming Citizen Science and Open Science in Europe. Data Intelligence, 3(1), 136-149.

[10] Citizen Science Association Data & Meta Data Working Group. (2021, July 29). PPSR Core, A Data Standard for Public Participation in Scientific Research (Citizen Science). URL: https://core.citizenscience.org/docs/project .

[11] Thomas, D. M., & Mathur, S. (2019, June). Data analysis by web scraping using python. In 2019 3rd International conference on Electronics, Communication and Aerospace Technology (ICECA) (pp. 450-454). IEEE.

[12] Sørvik, G. O., & Mork, S. M. (2015). Scientific literacy as social practice: Implications for reading and writing in science classrooms. Nordic Studies in Science Education, 11(3), 268-281.

[13] Huizinga, T., Handelzalts, A., Nieveen, N., & Voogt, J. (2015). Fostering teachers' design expertise in teacher design teams: conducive design and support activities. The Curriculum Journal, 26(1), 137-163.

[14] Park, S., & Oliver, J. S. (2008). Revisiting the conceptualisation of pedagogical content knowledge (PCK): PCK as a conceptual tool to understand teachers as professionals. Research in science Education, 38(3), 261-284.

[15] Bonney, R., Cooper, C. B., Dickinson, J., Kelling, S., Phillips, T., Rosenberg, K. V., & Shirk, J. (2009). Citizen science: a developing tool for expanding science knowledge and scientific literacy. BioScience, 59(11), 977-984.

[16] Brossard, D., Lewenstein, B., & Bonney, R. (2005). Scientific knowledge and attitude change: The impact of a citizen science project. International Journal of Science Education, 27(9), 1099-1121.

[17] Chan, K. K. H., & Yung, B. H. W. (2018). Developing pedagogical content knowledge for teaching a new topic: More than teaching experience and subject matter knowledge. Research in Science Education, 48(2), 233-265.

[18] Bennett, S., Agostinho, S., & Lockyer, L. (2015). Technology tools to support learning design: Implications derived from an investigation of university teachers' design practices. Computers & Education, 81, 211-220.

[19] Hylén, J. (2021). Open educational resources: Opportunities and challenges.

[20] Dagnino, F. M., Dimitriadis, Y. A., Pozzi, F., Asensio-Pérez, J. I., & Rubia-Avi, B. (2018). Exploring teachers' needs and the existing barriers to the adoption of Learning Design methods and tools: A literature survey. British Journal of Educational Technology, 49(6), 998-1013.

[21] Kali, Y., McKenney, S., & Sagy, O. (2015). Teachers as designers of technology enhanced learning. Instructional science, 43(2), 173-179.

[22] Recker, M., Walker, A., Giersch, S., Mao, X., Halioris, S., Palmer, B., ... & Robertshaw, M. B. (2007). A study of teachers' use of online learning resources to design classroom activities. New Review of Hypermedia and Multimedia, 13(2), 117-134.

[23] Amiel, T., & Reeves, T. C. (2008). Design-based research and educational technology: Rethinking technology and the research agenda. Educational Technology & Society, 11(4), 29–40.

[24] Creswell, J. W. (2014). Research design: Qualitative, quantitative, and mixed methods approaches (4th ed.). Sage publications.

[25] Daniel Glez-Peña, Anália Lourenço, Hugo López-Fernández, Miguel Reboiro-Jato, Florentino Fdez-Riverola, Web scraping technologies in an API world, Briefings in Bioinformatics, Volume 15, Issue 5, September 2014, Pages 788–797, URL: https://doi.org/10.1093/bib/bbt026

[26] Getenet, S. (2019). Using design-based research to bring partnership between researchers and practitioners. Educational Research, 61(4), 482-494.

[27] CS Track project. (2021, July 29). CS Track, investigating citizen science. URL: https://cstrack.eu/

[28] Museum für Naturkunde – Leibniz Institute for Evolution and Biodiversity Science. (2021, July 29). Knowledge for Change: Citizen Science SDG Conference. URL: https://www.museumfuernaturkunde.berlin/en/museum/events/knowledge-change-citizen-science-sdg-conference

[29] Patricia Santos, Miriam Calvera, Laia Albó, Julia Lorke, Ulrich Hoppe, & Davinia Hernández-Leo. (2020). CS-Track database: a central database of CS projects in Europe that can be key to understand the connection of CS and SDGs. Presented at the Citizen Science SDG Conference (CS SDG), Zenodo. URL: http://doi.org/10.5281/zenodo.4095157

[30] Citizen Science Association (CSA). (2021, July 29). CitSciVirtual 2021. URL: https://citizenscience.org/home/events/conferences/citscivirtual/

[31] Miriam Calvera-Isabal, Patricia Santos, Ishari Amarasinghe, & Davinia Hernandez-Leo. (2021). A central database of CS projects that can open a new perspective on CS knowledge. Presented at the CitSci virtual (CitSci virtual), Zenodo. URL: http://doi.org/10.5281/zenodo.4773098

[32] 18th International Conference on Cognition and Exploratory Learning in the Digital Age (CELDA 2021). (2021, July 29). CELDA conference 2021. URL: https://www.celda-conf.org/ .

[33] Pompeu Fabra University. (2021, July 29). Grants for activities to increase the social impact of research. URL: https://www.upf.edu/recercaupf/convocatoria-ajuts-impacte-social-recerca

[34] Pompeu Fabra University. (2021, July 29). Makers a les aules: Edició 2020-2021 URL: https://www.upf.edu/es/web/tide/news/-/asset_publisher/JP7iPN458OrU/content/id/244364382/maximized#.YPaDbuj7RPY