COMPUTATIONAL TECHNIQUES FOR DATA SCIENCE APPLIED TO BROADEN THE KNOWLEDGE BETWEEN CITIZEN SCIENCE AND EDUCATION

Miriam Calvera-Isabal, Nuria Varas and Patricia Santos
Research group on interactive and distributed Technologies for education, Pompeu Fabra University
Plaça de la Mercè, 10-12. 08002 Barcelona, Spain

ABSTRACT

This paper describes a preliminary study of how computational methods allow us to know more about citizen science and its connection with education. Citizen science is a practice involving a general public in scientific tasks and generating knowledge and scientific results. Previous studies have shown that the education sector can take benefit of the knowledge and activities organized or resources generated in CS projects. Previous studies have shown that the education sector can take advantage of the knowledge and activities organized in CS projects. In this papers, we analyze three citizen science platforms (Eu.Citizen science platform, Observatorio de la ciencia ciudadana and Oficina de la ciència ciutadana) with computational analytics techniques to provide initial insights of how educators can take benefit of the analysis of large amounts of data from CS. Finally, different visualizations and dashboards have been developed as illustrative examples of tools to support educators and learners. These tools provide information about citizen science projects, an overview of scientific vocabulary, access to validated resources and examples of technology used in scientific inquiry that can be used with educational purposes.

KEYWORDS

Citizen Science, Education, Data Analysis

1. INTRODUCTION

The objective of this paper is to show the potential that analytic computational techniques have to better understand the practice of citizen science (CS) and how the data extracted/analyzed can be used with educational purposes. There is a need to know more about the practice of CS, especially since there are more and more CS projects and interest in this field in Europe (Vohland et. al., 2021). The information of these CS projects is distributed online all over the world in many websites which causes that knowledge is not centralized and it is more difficult to have an overview.

Despite data analysis algorithms have been applied to many scientific fields (McNamara 2011, Koh and Tan 2011), there is a lack of knowledge of how to apply computational analytic techniques to better understand the gap of knowledge about CS projects practices and the connection with Education (Ginger et al. 2020, Lambers et al. 2019). For this research, data of 377 CS projects and resources available in three key CS platforms has been extracted. This preliminary study will also contribute to bringing citizen science closer to the education sector and equipping it with new tools and data to support the educational process. The study shows how this connection can be established and states the bases of many new analyses on citizen science, from another perspective than participating in scientific projects.

2. BACKGROUND

2.1 Citizen Science

The term Citizen Science is not clearly defined because there is no consensus on the various definitions proposed by the community (Auerbach et al. 2019, Haklay et al. 2021). Even so, we select the following

definition: “Citizen science broadly refers to the active engagement of the general public in scientific research tasks” (Vohland, V et. at. 2021).

CS can be applied to many different scientific disciplines such as natural sciences, social sciences, or humanities (Tauginienė et al. 2020) although it depends on the scientific objectives of the CS project. There is a long history of citizens contributing to science (i.e., Christmas Bird Count or The American Association of Variable Star Observers (AAVSO)) which causes an impact on policies, human health or in society (Hecker et al. 2018). CS promotes citizen participation and interest in science in addition to other qualities related to learning about science (Shah and Martinez 2016).

2.2 Citizen Science & Education

There is a lot of evidence showing how CS can contribute to Science, Technology, Engineering and Maths (STEM) career motivation (Hiller and Kitsantas 2014). CS promotes values such as ecology, knowledge about the environment (Kobori et al. 2016, Ballard, Dixon and Harris 2017, Kelemen-Finan, Scheuch and Winter 2018) or critical thinking (Masterson et al. 2019).

In general, the scientific objectives of CS projects are not educational, but a common practice is to create educational materials and adapt practices to support learning (National Academies of Sciences, Engineering, and Medicine 2018, Schuttler et al. 2019). Posters, videos or guides are the most common educational resources created by projects to support participants during their participation in scientific research (Brossard, Lewenstein and Bonney 2005) These materials are helpful to support literacy, understand how investigation will be developed, know the timings, promote open discussions and advance in scientific knowledge and improve scientific skills (Bonney et al. 2009). Citizen science involves a participatory process which implies, for instance, events to train volunteers or workshops, carried out during the development of the project (Cohn, 2008).

2.3 Citizen Science Platforms

Since CS is attracting more and more attention every year (Bautista et al. 2019), several associations are being created to help in the development and management of CS. Divided by geographical areas, we could say that there are three key CS associations: the Citizen Science Association (CSA, North America), the European Citizen Science Association (ECSA, Europe) and the Australian Citizen Science Association (ACSA, Australia) (Storksdieck et al. 2016). Covering regional areas, we can find national or regional associations like Observatorio de la ciencia ciudadana in Spain or Flemish Knowledge Center for Citizen Science in Belgium (SCivil). These associations have a website or online digital platform containing information about CS projects, events, related resources and sometimes have a space for communications (i.e., news, forum, etc) (Sanz, Gold and Mazzonetto 2019). In ‘The science of citizen science’ it has been defined five different types of online CS platforms; in this preliminary study are analysed two National CS platforms (Observatorio de la ciencia ciudadana and Oficina de ciència ciutadana) and one European CS platform (EU.Citizen science) (Vohland et. al., 2021). These online portals, act as a repository having metadata information about regional CS projects in a structured or unstructured way, but only few of them use the metadata standards (i.e., PPSR metadata standard (https://core.citizenscience.org/)). The fact that the data is distributed in many platforms, in different languages and data structures, makes the process of analyzing and exploring the data from CS more complicated.

2.4 Data Extraction, Analysis and Visualization

In order to address the problem defined in the previous section, we propose to apply web scraping techniques. These are typically used for extracting data automatically from websites and storing it structured in a database or file. Web scraping software or personalized ones (also called robots or crawlers) is used in many fields like arts and humanities or biology (Diouf et al. 2019). In the context of CS, these techniques have been used to analyze citizen participation in online project forums (Ponti et al. 2018). However, there is no evidence of having used it to obtain metadata about CS projects although it has been proved a powerful tool to extract and classify data (Karthikeyan et al. 2019).

Data mining techniques such as Natural Language Processing (NLP), Sentiment Analysis or Machine Learning is applied in combination with web scraping to better understand data retrieved. These analytical techniques are typically used in CS projects to analyze the data obtained from volunteers' participation (Caruana et al. 2006, Fink and Hochachka 2012). Nevertheless, although many qualitative studies are done to analyze CS projects metadata (Bonney et al. 2009), there is a lack of knowledge related to the application of other automatic methods (i.e., text mining) to analyze CS online data.

In addition to computation analytic methods, data visualization is useful to validate hypotheses, get insight from the data (i.e., identifying patterns) and as a communication tool for end-users. Depending on the type of data (i.e., one-dimensional or two-dimensional), visualization technique (i.e., Standard 2D/3D displays) and the interaction (i.e., filtering or zooming), many techniques can be applied to the linked data such as the one extracted for the study presented in this paper (Keim 2002). Due to the number of different visualizations that can be created using this data, dashboards are a good solution that facilitate the monitoring of the data by integrating them into a single tool. For example, it has been shown the positive impact of dashboards to help educators to improve as professional facilitators, analysts, and designers. Moreover, these tools also support the learning process of students (Michaeli, Kroparo and Hershkovitz 2020).

3. METHODOLOGY

Three CS platforms have been selected: The Eu.Citizen science platform, The Observatorio de la ciencia ciudadana and Oficina de la ciència ciutadana. The purpose of this analysis is to know the connections between citizen science and education and identify information that can support the education process. Data from these three platforms has been analyzed to better understand what information is shared, vocabulary used and resources available that could be used in an educational context. To automate the manual work of extracting data from websites, web scraping tools are used to extract this data from multiple sites and store it, structured, in a database or file (Zhao 2017).

In order to provide rules to this web scraping process, a Robots Exclusion Protocol (REP) was defined. This protocol defines a regulatory framework that allows site administrators to decide which crawlers are or are not allowed to access a specific part of the website (Sun, Zhuang and Giles 2007). These specifications have to be informed for each website in the robots.txt file with a certain structure for crawlers to read it before scraping the site and check if data can be extracted or not.

For this preliminary study, a crawler has been developed with python (https://www.python.org/). Selenium (https://selenium-python.readthedocs.io/) has been chosen as the main web scraping software. The Observatorio de la ciencia ciudadana and The Oficina de ciència ciutadana websites have been scraped with that crawler to obtain the data. The EU.citizen science website has its own Application Programming Interface (API) (https://eu-citizen.science/swagger/) so, a code has been developed with python to get the data through it. Once the data has been collected, it has been stored in different files (by platform) in a structured way by classifying it by type of data (Saurkar, Pathare and Gode 2018). To analyze the data again an algorithm developed with python and other data science libraries and tools (i.e., pandas, re or nltk) and Jupyter notebook (https://jupyter.org/) were used.

To pre-process the data before performing the analysis, different methods of text mining have been applied (i.e data cleaning, tokenization or lemmatization) (Vijayarani and Janani 2016). When the data has been prepared to be analyzed, a Term Frequency - Inverse Document Frequency technique (TF-IDF) has been applied to count the number of occurrences for a given word in a document (TF) it measures if a term is common or not in a collection of documents (IDF) (Dillon 1983). With this method, the most relevant words for a document can be computed. After that, Name entity recognition (NER) has been applied to identify people's names or institutions from text.

Finally, to continue with the data analysis data has been categorized. The way to classify projects or resources using text fields is generating a list of words that refer to a topic, for example, and looking if that project description contains any of those words. In order to extract correlations among variables such as projects topics and status, those variables need to be converted into numerical values. Therefore, label encoders must be used to determine which number corresponds to a specific type.

The same process has been applied to online resources metadata. In case of texts that contain a description of the resource, as the previous section explains, text mining has been used to obtain the abstract word cloud, length distribution and the most relevant terms for each resource (with TF- IDF). For resources has been developed an analysis of metadata information of the timeline of when the resource was published.

Dashboards with Tableau (https://www.tableau.com/es-es) have been generated to show conclusions about the data analyzed of the project's information and resources. This is a very useful data visualization tool to make data more understandable by creating interactive dashboards and visualizations.

4. STUDY OF THREE CITIZEN SCIENCE PLATFORMS: RESULTS AND ANALYSIS

Figure 1 shows a comparative analysis of the data from the three different platforms (Eu.Citizen science, Observatorio de la ciencia ciudadana and Oficina de Ciència ciutadana).

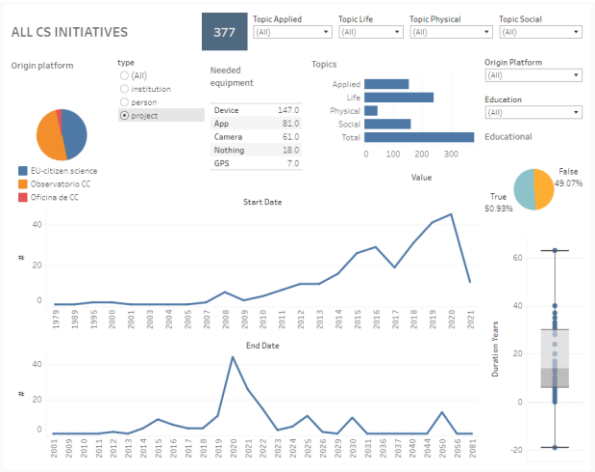


Figure 1. Information of CS projects of all platforms

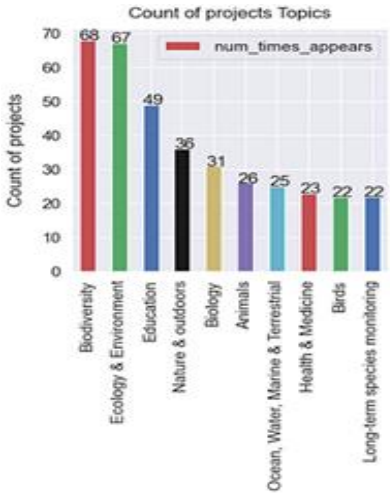


Figure 2. List of Topics obtained from CS projects information of the EU.Citizen science platform

As a result, data from 377 CS projects have been extracted from the 3 different CS platforms analyzed. The Spanish platform is the one with more content (188 CS projects), followed by the European one (176 CS projects) and the one from Barcelona (13 CS projects). Some data (16 CS projects) are duplicated within Spanish and European platforms.

First of all, data has been analyzed and related to education. CS projects that contain educational keywords (i.e., schools, education or classrooms) have been classified in the category “Education”. 241 over 468 projects (51,5%) meet the criteria. As a result, a filter has been designed for educators or learning designers to rapidly visualize specific data of CS projects related to educational initiatives.

In order to obtain a deeper understanding of the CS projects, the topics information have been distinguished in four different groups (depending on their science types) that correspond to the primary disciplines of citizen science projects: Applied sciences, social sciences, life sciences or physical sciences. To sum up, Life science is the discipline more common in the data being Biodiversity and Ecology & Environment the topics from this category most frequent in the Eu.Citizen science platform is (Figure 2).

As a result, data from 377 CS projects have been extracted from the 3 different CS platforms analyzed. The Spanish platform is the one with more content (188 CS projects), followed by the European one (176 CS projects) and the one from Barcelona (13 CS projects). Some data (16 CS projects) are duplicated within Spanish and European platforms.

First of all, data has been analyzed and related to education. CS projects that contain educational keywords (i.e., schools, education or classrooms) have been classified in the category “Education”. 241 over 468 projects (51,5%) meet the criteria. As a result, a filter has been designed for educators or learning designers to rapidly visualize specific data of CS projects related to educational initiatives.

Considering an educational perspective, we could say that topics of CS initiatives are similar to those addressed in the curriculum. This analysis has special interest to the educational community when exploring specific data of the project. We have analyzed the terminology related to citizen science projects for each platform and project. Figure 3 shows a word cloud graph with the most relevant words in the Eu.Citizen science platform.

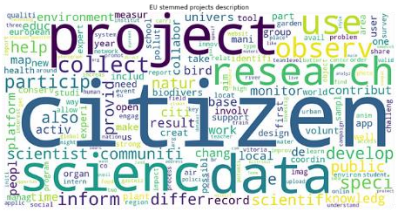
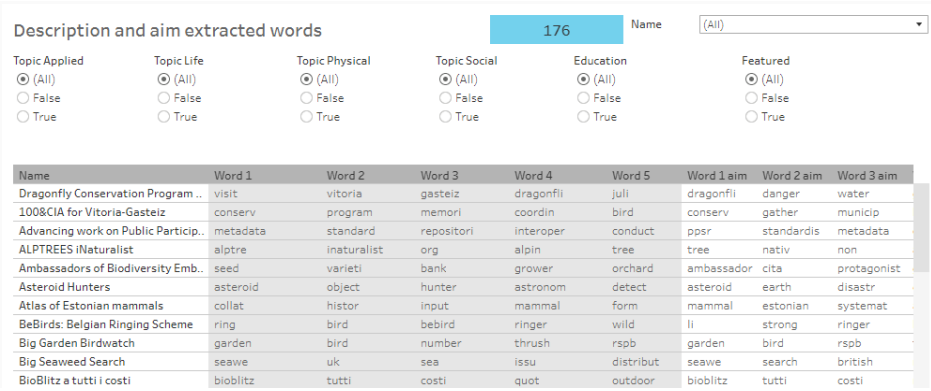


Figure 3. A word cloud with the most common stemmed terms of the Eu.Citizen science platform

Aristeidou and Herodotou (2020) states that by participating in citizen science projects communication skills and vocabulary are improved. In Figure 4 the most important terms used by the citizen science projects of the European platform can be explored. They can also be filtered by project category (previously analyzed), which provides added value since it allows the words to be classified by scientific typologies. By exploring specific terminology, learners can improve functional science literacy (Utami, Saputro and Masykuri 2016).



Name	Word 1	Word 2	Word 3	Word 4	Word 5	Word 1 aim	Word 2 aim	Word 3 aim
Dragonfly Conservation Program ..	visit	vitoria	gasteiz	dragonfli	juli	dragonfli	danger	water
100&CIA for Vitoria-Gasteiz	conserv	program	memori	coordin	bird	conserv	gather	municip
Advancing work on Public Particip..	metadata	standard	repositori	interoper	conduct	ppsr	standardis	metadata
ALPTREES iNaturalist	alptre	inaturalist	org	alpin	tree	tree	nativ	non
Ambassadors of Biodiversity Emb..	seed	varieti	bank	grower	orchard	ambassador	cita	protagonist
Asteroid Hunters	asteroid	object	hunter	astronom	detect	asteroid	earth	disastr
Atlas of Estonian mammals	collat	histor	input	mammal	form	mammal	estonian	systemat
BeBirds: Belgian Ringing Scheme	ring	bird	bebird	ringer	wild	li	strong	ringer
Big Garden Birdwatch	garden	bird	number	thrush	rspb	garden	bird	rspb
Big Seaweed Search	seawe	uk	sea	issu	distribut	seawe	search	british
BioBlitz a tutti i costi	bioblitz	tutti	costi	quot	outdoor	bioblitz	tutti	costi

Figure 4. Dashboard with the keywords extracted from CS projects description of the Eu.Citizen science platform

Another aspect analysed is the use of information and communication technology (ICT, also called materials) by participants. Depending on the objective of the project and how the participation in the scientific process was designed, there are multiple tools used to conduct research (i.e., MammalNet: Watch Wildlife for Science (https://www.mammalnet.com/)). In general, the most common tools used in citizen science are ICT devices (i.e., smartphones) and cameras (Figure 5). ICT tools applied in an educational environment have an impact on student's scientific literacy (Luu and Freeman 2011). Knowing which technologies are useful according to certain scientific tasks helps teachers select which ones to use with their students.

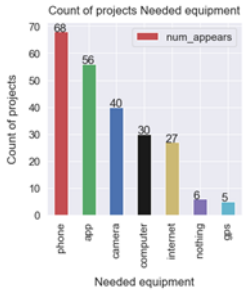


Figure 5. Eu.Citizen science platform most required ICT tools

Finally, the two temporal graphs (Figure 1) show when projects start and end. The first illustrates the growing interest that citizen science has because of the number of projects that started in the last ten years.

These dates can be used to understand how actualized the data and the material used are because the more recent the start date, the more innovative the project will be and the more interest it will have in current problems.

Figure 6 shows how many of the 212 resources extracted belong to each platform (62% are from the European platform and 38% are from the Spanish platform). The information has been classified into categories, depending on the typology of resources (most of them are guidelines, reports, websites or educational resources). Additionally, it has been informed if they are linked to a citizen science project or not (40% of the total are, while the remaining 60% are not).

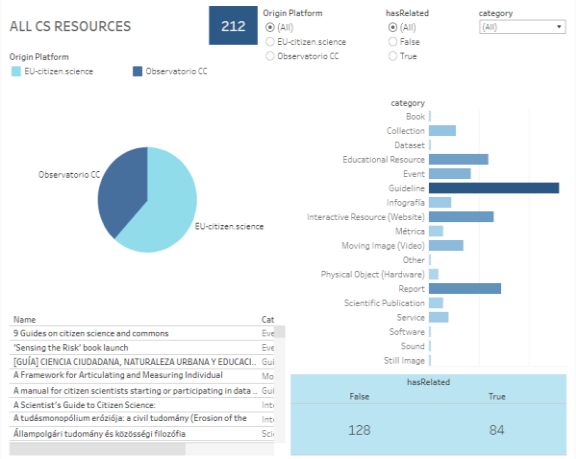


Figure 6. CS projects resources available in platforms

5. CONCLUSION AND FUTURE WORK

The goal of this preliminary study has been to use Data Science techniques to extract and analyze data from CS projects, create visualizations and analyze the connections and possibilities of the data with an educational perspective. The analysis done in this paper shows the potential to analyze large amounts of data to understand better the CS practice, and in particular how this data can be used with educational purposes.

The main limitation observed in this study is that there are inconsistencies on the data structure of CS platforms and the metadata standard use. This implies that specific algorithms have to be designed and implemented based on specific website structures. This limitation, along with the use of different national languages, entails a time consuming process and high cost of resources. In addition to this, another limitation of this preliminary study is that so far educators have not been involved in this research.

The lessons learned from the process followed in this study allowed us to identify four different potential benefits that can be further explored in the future: (1) Access to CS projects information ordered by different categories, (2) Overview of scientific vocabulary, (3) Access to scientific validated resources (4) Technologies that can be used in scientific inquiry.

Having access to CS projects information classified by different disciplines. This can potentially help teachers to design learning activities/content as it can be used as a source of scientific information. Besides that, the students that explore the information could find many examples about different scientific projects motivating or stimulating to learn more about scientific topics.

Having an overview of scientific vocabulary supports teachers during the teaching process and students to easily identify science concepts. Suggate et.al (2012) concluded that there is a positive significant correlation between lexicon and later literacy development. The combination of the use of scientific terms and specific teaching strategies during learning activities implies better vocabulary development by students and will affect literacy (Hong and Diamond 2012).

Enhance the accessibility to scientific validated resources that can be used by a non-academic audience to support the learning process in science. Teachers use resources available online to design their own learning design materials or activities or to use it as main materials**.**

Identifying technologies that can be used in scientific inquiry will help teachers to define which strategies follow on the use of ICT in the classroom. Many examples of how technology is being used in CS

projects for specific tasks can be found in the graphs developed. There is much evidence of how the use of ICT affects positively in students' learning (Fu 2013). Taking into account that one of the barriers of using technology in class is the self-confidence of teachers, it is important to bear in mind that, in many cases, there will be additional resources developed by projects to explain how to use it.

For future work, we plan to organize co-design workshops, as proposed by Scanlon, McAndrew and O'Shea (2015), to understand the needs of educators and students when interacting with Dashboards as the ones proposed in this paper. From the citizen science viewpoint, having CS projects metadata in a single repository will help the CS and Education communities in the analysis and expansion of knowledge of different aspects like CS projects communication or volunteer tasks and how participation is conducted. Additionally, we want to expand the amount of data by extracting information on more platforms to have more information about CS projects.

ACKNOWLEDGEMENT

This work has been partially funded by the CS Track project, European Union's Horizon 2020 research and innovation programme under grant agreement No 872522. Patricia Santos acknowledges the support by the Spanish Ministry of Science and Innovation under the Ramon y Cajal programme.

REFERENCES

Aristeidou, M., & Herodotou, C. (2020). Online citizen science: A systematic review of effects on learning and scientific literacy. Citizen Science: Theory and Practice, 5(1), 1-12.

Auerbach, J., Barthelmess, E. L., Cavalier, D., Cooper, C. B., Fenyk, H., Haklay, M., ... & Shanley, L. (2019). The problem with delineating narrow criteria for citizen science. Proceedings of the National Academy of Sciences, 116(31), 15336-15337.

Ballard, H. L., Dixon, C. G., & Harris, E. M. (2017). Youth-focused citizen science: Examining the role of environmental science learning and agency for conservation. Biological Conservation, 208, 65-75.

Bautista-Puig, N., De Filippo, D., Mauleón, E., & Sanz-Casado, E. (2019). Scientific landscape of citizen science publications: Dynamics, content and presence in social media. Publications, 7(1), 12.

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.

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.

Caruana, R., Elhawary, M., Munson, A., Riedewald, M., Sorokina, D., Fink, D., ... & Kelling, S. (2006, August). Mining citizen science data to predict orevalence of wild bird species. In Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining (pp. 909-915).

Cohn, J. P. (2008). Citizen science: Can volunteers do real research?. BioScience, 58(3), 192-197.

Dillon, M. (1983). Introduction to modern information retrieval. Information Processing & Management, 19(6). Vijayarani, S., & Janani, R. (2016). Text mining: open source tokenization tools-an analysis. Advanced Computational Intelligence: An International Journal (ACII), 3(1), 37-47.

Diouf, R., Sarr, E. N., Sall, O., Birregah, B., Bousso, M., & Mbaye, S. N. (2019, December). Web Scraping: State-of-the-Art and Areas of Application. In 2019 IEEE International Conference on Big Data (Big Data) (pp. 6040-6042). IEEE.

Fink, D., & Hochachka, W. M. (2012). Using data mining to discover biological patterns in citizen science observations. In Citizen science: public participation in environmental research (pp. 125-138). Ithaca, NY: Comstock Publishing Associates.

Fu, J. (2013). Complexity of ICT in education: A critical literature review and its implications. International Journal of education and Development using ICT, 9(1), 112-125.

Ginger Tsueng, Max Nanis, Jennifer T Fouquier, Michael Mayers, Benjamin M Good, Andrew I Su, Applying citizen science to gene, drug and disease relationship extraction from biomedical abstracts, Bioinformatics, Volume 36, Issue 4, 15 February 2020, pp 1226–1233.

Haklay, M. M., Dörler, D., Heigl, F., Manzoni, M., Hecker, S., & Vohland, K. (2021). What Is Citizen Science? The Challenges of Definition. The Science of Citizen Science, 13.

Hecker, S., Haklay, M., Bowser, A., Makuch, Z., Vogel, J., & Bonn, A. (2018). Innovation in open science, society and policy–setting the agenda for citizen science. Citizen science: innovation in open science, society and policy. UCL Press, London, UK, 1-23.

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.

Hong, S. Y., & Diamond, K. E. (2012). Two approaches to teaching young children science concepts, vocabulary, and scientific problem-solving skills. Early Childhood Research Quarterly, 27(2), 295-305.

Karthikeyan, T., Sekaran, K., Ranjith, D., & Balajee, J. M. (2019). Personalized content extraction and text classification using effective web scraping techniques. International Journal of Web Portals (IJWP), 11(2), 41-52.

Keim, D. A. (2002). Information visualization and visual data mining. IEEE transactions on Visualization and Computer Graphics, 8(1), 1-8.

Kelemen-Finan, J., Scheuch, M., & Winter, S. (2018). Contributions from citizen science to science education: An examination of a biodiversity citizen science project with schools in Central Europe. International Journal of Science Education, 40(17), 2078-2098.

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

Koh, H. C., & Tan, G. (2011). Data mining applications in healthcare. Journal of healthcare information management, 19(2), pp 65.

Lambers, K., Verschoof-van der Vaart, W. B., & Bourgeois, Q. P. (2019). Integrating remote sensing, machine learning, and citizen science in Dutch archaeological prospection. Remote Sensing, 11(7), pp 794.

Luu, K., & Freeman, J. (2011). An analysis of the relationship between information and communication technology (ICT) and scientific literacy in Canada and Australia. Computers & Education, 56(4), 1072-1082.

Masterson, J., Meyer, M., Ghariabeh, N., Hendricks, M., Lee, R. J., Musharrat, S., Newman G., Sansoms G. & Van Zandt, S. (2019). Interdisciplinary citizen science and design projects for hazard and disaster education. International journal of mass emergencies and disasters, 37(1), 6.

McNamara, D. S. (2011). Computational methods to extract meaning from text and advance theories of human cognition. Topics in Cognitive Science, 3(1), pp 3-17.

Michaeli, S., Kroparo, D., & Hershkovitz, A. (2020). Teachers' Use of Education Dashboards and Professional Growth. The International Review of Research in Open and Distributed Learning, 21(4), 61-78.

National Academies of Sciences, Engineering, and Medicine. (2018). Learning through citizen science: enhancing opportunities by design. National Academies Press.

Ponti, M., Hillman, T., Kullenberg, C., & Kasperowski, D. (2018). Getting it right or being top rank: Games in citizen science. Citizen Science: Theory and Practice, 3(1).

Sanz, F., Gold, M., & Mazzonetto, M. (2019). D2.3: Platform Functionality Requirements & Specification Report (Version Version Final Submitted draft). Zenodo.

Saurkar, A. V., Pathare, K. G., & Gode, S. A. (2018). An overview on web scraping techniques and tools. International Journal on Future Revolution in Computer Science & Communication Engineering, 4(4), 363-367.

Scanlon, E., McAndrew, P., & O'Shea, T. (2015). Designing for educational technology to enhance the experience of learners in distance education: how open educational resources, learning design and MOOCs are influencing learning. Journal of interactive Media in education, 2015(1).

Schuttler, S. G., Sears, R. S., Orendain, I., Khot, R., Rubenstein, D., Rubenstein, N., Dunn R. R., Baird E., Kandros K., O'Brien T. & Kays, R. (2019). Citizen science in schools: Students collect valuable mammal data for science, conservation, and community engagement. Bioscience, 69(1), 69-79.

Shah, H. R., & Martinez, L. R. (2016). Current approaches in implementing citizen science in the classroom. Journal of microbiology & biology education, 17(1), 17.

Storksdieck, M., Shirk, J. L., Cappadonna, J. L., Domroese, M., Göbel, C., Haklay, M., ... & Vohland, K. (2016). Associations for citizen science: regional knowledge, global collaboration. Citizen Science: Theory and Practice, 1(2).

Suggate, S., Schaughency, E., McAnally, H., & Reese, E. (2018). From infancy to adolescence: The longitudinal links between vocabulary, early literacy skills, oral narrative, and reading comprehension. Cognitive Development, 47, 82-95.

Sun, Y., Zhuang, Z., & Giles, C. L. (2007, May). A large-scale study of robots. txt. In Proceedings of the 16th international conference on World Wide Web (pp. 1123-1124).

Tauginienė, L., Butkevičienė, E., Vohland, K., Heinisch, B., Daskolia, M., Suškevičs, M., ... & Prūse, B. (2020). Citizen science in the social sciences and humanities: the power of interdisciplinarity. Palgrave Communications, 6(1), 1-11.

Utami, B., Saputro, S., & Masykuri, M. (2016, January). Scientific literacy in science lesson. In Proceeding of International Conference on Teacher Training and Education (Vol. 1, No. 1).

Vohland, K., Land-Zandstra, A., Ceccaroni, L., Lemmens, R., Perelló, J., Ponti, M., ... & Wagenknecht, K. (2021). The Science of Citizen Science Evolves. The Science of Citizen Science, 1.

Zhao, B. (2017). Web scraping. Encyclopedia of big data, 1-3.