J Data Sci. Info. Citation Studies., 2023; 2(2):65-76.
https://www.jcitation.org

Research Article

Mapping Global Scientific Output on Citizen Science from 2012 to 2021: A Scientometric Analysis

Tapan Barui, Sourav Mazumder

Department of Library and Information Science, University of North Bengal, Darjeeling, West Bengal, INDIA.

ABSTRACT

Citizen Science (CS) methods are gaining importance in the scientific community. We conducted a comprehensive scientometrics analysis of global scientific output on CS. The bibliographic data of 1679 articles published between 2012-2021 were retrieved using the Scopus database. We used the R Bibliometrix, Google Sheets, VOSviewer, and Gensim Python library to analyse and visualise data. The results highlighted an increase in scientific output on CS, with a 40.68% annual average growth rate and a 28.96% exponential growth rate. It revealed that PLoS One was the most significant journal, Callaghan, C.T. was the most prominent author, the USA was the most dominant country, and National Science Foundation was the most acknowledged funder in terms of CS research productivity. The collaboration network analysis demonstrated strong connections between the USA, UK, Australia, and Germany. Therefore, the results showed the applicability of CS methods in the research of "birds distribution," "biodiversity," "species classification," "natural resource management," and "public engagement." This paper could be useful for policymakers, researchers, and funding agencies to identify the trend of research, potential collaborators and institutions, areas of research, and opportunities in the field of CS.

Keywords: Community science, Science mapping, Research productivity, Highly cited papers, LDA topic modelling.

Correspondence:
Mr. Sourav Mazumder
Department of Library and Information Science, University of North Bengal, Darjeeling, West Bengal, INDIA.
Email: smazumderlis91@gmail.com

Received: 01-12-2022;
Revised: 27-01-2023;
Accepted: 03-03-2023.

INTRODUCTION

Citizen Science (CS) is a participatory act of non-researchers– often called public in conducting scientific research. In CS, non-researchers participate and support scientists to expand the length of scientific knowledge. Public participation in the process of scientific research has significantly increased (Follett & Strezov, 2015). We may investigate the CS terminological evolution. Alan Irwin, a British sociologist (Irwin, 1995), Rick Bonney, and other scientists (Bonney *et al*., 2009a) termed and characterised CS. Oxford English Dictionary (OED) defines CS as "scientific work undertaken by members of the general public, often in collaboration with or under the direction of professional scientists and scientific institutions" (Daily Zooniverse, 2014). We also could find other definitions produced by different scientists and institutions (Wikipedia contributors, 2022).

Our society faces several challenges, such as COVID-19 and Climate change, which have shaken the global landscape. Science can safeguard the global environment, yet it needs the active involvement of common people who can collect, manage, and describe data for research. Traditional research activity is restricted to one researcher or a group of researchers, which may not be sufficient to achieve a specific goal. The advancement of Information Communication Technologies (ICTs) opens a wide-open door for researchers to communicate with the public and contribute scientific knowledge. A CS project can engage the public in collaborating toward a common goal (SciStarter, n.d.). For instance, the National Audubon Society's (USA) project Audubon Christmas Bird Count, is one of the longest-running community science bird projects (The Audubon Society, 2022). Volunteer birdwatchers participate in bird counts to provide bird census data in this project. Furthermore, citizen science projects involve the public monitoring on air quality, water quality, and fish, collecting rainforest, wildlife conservation, and marine science data (Conrad & Hilchey, 2011). Bonney *et al*. (2009a) divided CS projects into three main categories: (1) Contributory projects: public engagement in collecting data (e.g., Project Feeder Watch-winter bird populations); (2) Collaborative projects: public support in project design and analysing research findings (e.g., water quality monitoring) ; and (3) Co-created projects: public involvement in all stages of the research process (e.g., public health). West and Pateman (2016) recommended stages of the participation journey of a person who gets involved in any project: "Awareness of opportunity," where a person decides to take participation; "Initial participation," where a person matches

DOI: 10.5530/jcitation.2.2.10

the role in the project; "Sustained participation," where a person is communicated with the organiser regularly; and "Finish participation," where a person gives feedback on the outcomes.

Data derived from the website of the Association of Citizen Science (https://citizenscience.org/, accessed on 11/07/2022) reveals that there are more than 2000 projects, 10 lakhs volunteers, and 5000 members involved in CS projects. We can explore relevant, sustainable development projects such as "zero hunger," "no poverty," "quality education," and "climate action" (SciStarter, n.d.). Some remarkable CS research on "ecology and evolution," "public scientific literacy," "CS projects' impact," "science 2.0", and "monitoring socio-ecological systems" (Bonney *et al*., 2009b; Brossard *et al*., 2005; Cohn, 2008; Newman *et al*., 2012; Silvertown, 2009) were carried out in the past decades. CS is an evolving discipline and is required to support science and education (Newman *et al*., 2012). The rising growth of CS projects and activities has increased scholarly literature. Scientometrics analysis helps researchers to understand the growth of literature, intellectual structure and research patterns (Chaubey & Singh, 2021; Follett & Strezov, 2015; Hajibayova *et al*., 2021). It is an effective method of determining "citation count," "Journal Impact Factor (JIF)," "downstream citations," and "author index"(Cooper, 2015).

Main aim of this study is to map global scientific output on CS from 2012 to 2021. Some researchers conducted scientometrics studies on CS (Bautista-Puig *et al*., 2019; Chaubey & Singh, 2021; Follett & Strezov, 2015; Hajibayova *et al*., 2021; Kullenberg & Kasperowski, 2016; Zhang Xuanhui & Zhang Xuanhui, 2017). Still, there could be a comprehensive study, which might provide unique insights. To fill this gap, we performed a scientometrics analysis emphasising CS literature. This study depicts research output over the last 10 years' research production, research collaboration, and research topics in the context of CS. This study will provide valuable insights into the current state of CS research and practices and help researchers and practitioners with future research approaches.

Objectives of the study

This study attempts to achieve the following objectives: to assess the research productivity and growth of literature on Citizen Science (CS) during the period from 2012 to 2021, to analyse the knowledge structure of CS research, and to discover the latent topics from the published literature.

Literature Review

In the past years, various scientometrics studies have been carried out to examine the scientific literature on CS. Here we have presented the earlier work relevant to our study. Follett and Strezov (2015) analysed peer-reviewed published articles on CS. They collected the bibliographic data from the Web of Science (WoS) and Scopus databases. The results showed that there had been substantial methodology and validation techniques prior to the sharp increase in published research based on citizen science methods. Further, it indicated the growing interest among the researchers focused on CS project-based research and re-using the data of past projects. Kullenberg and Kasperowski (2016) conducted a scientometrics-meta analysis to define the conceptual characteristics and evaluate the scientific output of CS using data retrieved from the WoS database. The findings showed that "biology," "conservation," and "ecology" were the primary research areas. Using "geographic information research," the authors also discovered that citizens collect geographic data. Bautista-Puig *et al*. (2019) analysed 5100 articles from WoS. The results indicated that most CS publications were published from 2006 to 2017. The authors applied the co-occurrence technique to map the subject clusters, where "health," "bio," "geo,", and "public" were identified. Also, they observed the social shares of publications. Pelacho *et al*. (2021) approached quantitative and qualitative methods to pivot citizen science publications indexed to WoS to draw the evolution and collaboration networks. The results revealed: the exponential growth of highly-quality research articles, highly interconnected researchers producing scientific output, and many professional scientists considering citizen science a viable methodology in their research processes. Chaubey and Singh (2021) examined 2872 CS publications published between 1993 and 2020 using data they collected from WoS. They identified the "environmental sciences," "ecology," and "biodiversity" as the most prevalent research areas. Additionally, the results indicated that the USA was the most productive country and that 82% of papers had received citations. Hajibayova *et al*. (2021) discovered seven broad areas of scholarly interest in CS by analysing 92 peer-reviewed publications using WoS. Also, the result showed that "survey" and "mixed methods" were substantially more preferred in applied citizen science research.

Some studies focused on specific aspects in the context of CS. For instance, Odenwald (2018) conducted a citation analysis of 143 articles based on 23 CS projects about astronomy and space science. The study reported the statistics of highly-cited papers cited more than 200 times. In another study, Odenwald (2020) obtained 783 peer-reviewed papers using WoS and "FedCats" to assess the scientific output of the earth science projects in CS. The author presented different scientometrics indicators such as impact factors and *h*-index. The results revealed that "Christmas Bird Count," "eBird," and "The GLOBE Project" were the top-three projects comprising 72% of the publications and the highest number of citations. De Filippo *et al*. (2020a) Used the Scopus and CORDIS data to investigate scientific output to determine the adoption of CS in energy efficiency research projects. The findings revealed that only 336 articles closely fit the study's main aim. However, it was stated that the number of articles addressing the adoption of CS in energy efficiency research has increased. In 2020, De Filippo *et al*. (2020b) carried out a study to investigate the scientific output regarding the application of CS methods in the field of "water resource management." The results determined that research processes on the topic were increasing day by day, and Australia was one of those countries that produced the most research. Bedessem *et al*. (2021) conducted a comparative analysis of the scientific impact of CS on biodiversity. The analysis was based on 123 research papers produced by Vigie-Nature from 2007 to 2019. The findings showed that the yearly citation growth was higher than other publications in the same field.

In all the studies reviewed here, scientometrics is recognised as one of the effective methods to understand CS research. The first part of this section presents the main results that are closely related to our studies. Most of the studies focused on analysing the CS research areas, the growth of literature, and research productivity. The second part of the section added some specific types of studies, focusing on the application of CS methods in different domains. Here the authors preferred citation studies to achieve their research goals. However, there are few studies on CS using scientometrics. It creates room for expanding more on the current state of knowledge regarding CS research. This study aims at closing the gaps by achieving the objectives of the study.

METHODS

A comprehensive scientometric analysis of published scientific literature on CS between 2012 and 2021 is the methodology used to carry out this study.

Search strategy and data collection

The content and references of the CS Wikipedia article (2022) were studied to comprehend the context of CS. We identified a total of five CS-related terms. However, we selected only four terms- "citizen science," "community science," "crowd science," and "civic science," which were further used to retrieve bibliographic data for this study. We used the Scopus database (Elsevier, n.d.) to search published literature on CS. Scopus is a product of Elsevier, and it covers more than 84 million records and 1.8 billion cited references. A total of 1679 articles were retrieved and exported as a BibTex file from the database using the search string "(TITLE("citizen science") OR TITLE("community science") OR TITLE("crowd science") OR TITLE("civic science")) AND PUBYEAR > 2011 AND PUBYEAR < 2022 AND (LIMIT-TO(DOCTYPE, "ar")) AND (LIMIT-TO (LANGUAGE, "English")). The scope of the present study is restricted to some parameters. First, the search strategy was built around search terms that appeared in the titles of articles in the Scopus database (see Figure 1). Second, we selected the only journal articles (in English) published in the field of CS during the period 2012-2021. Third, we used four popular broader search terms during bibliographic data retrieval. We may find more terms in the study of Pelacho *et al*. (2021).

Data analysis and visualisation

Several desktop or web-based scientometrics and visualisation software packages are available for analysing bibliographic data, such as BibExcel, Bibliometrix (R-package), Biblioshiny, CiteSpace, SciMat, ScientoPy, VOSviewer, and Gephi (Moral-Muñoz *et al*., 2020). In this study, we used the Bibliometrix package to quantify different scientometrics measures like the number of articles, total citations, and author index. Further, VOSviewer (version 1.6.18) was used to visualise the collaboration network of authors and countries. We created subsets of the main dataset based on specific sponsors from the Scopus database in order to analyse the funding sponsors tabulated in Table 6. Also, we have presented the most cited articles and references (see Table 7). We came across many software-generated duplicates regarding cited references before acquiring the unique list for analysing. We found that different variations of journal names appeared in many works. "Trends Ecol. Evol." was frequently used to refer to the publication "Trends in Ecology and Evolution." In Google Sheets, we used Regular Expressions (Regex) (*REGEXEXTRACT*, n.d.) to eliminate duplicates and create a unique list.

Topic modelling

We then applied Latent Dirichlet Allocation (LDA) (Blei *et al*., 2003) to discover the top-five topics discussed in the 1679 articles. In 2003, Blei *et al*. (2003) introduced LDA, "a generative probabilistic model" which allows for uncovering hidden topics from a corpus, a collection of textual data. Earlier studies (Lamba & Madhusudhan, 2019; Mazumder & Barui, 2021) show that LDA is a robust automatic topic modelling approach. In this study, we used Gensim, an open-source Python library (Rehurek & Sojka, 2010) for topic modelling to generate the latent topics from the titles of the 1679 articles. Since our dataset (article titles) contained four searched keywords, we excluded them (along with 18 non-representative words) to improve the quality of our analysis. Further, we labelled the keywords as topics (see Figure 6).

The term "scientific output" has also been written as research or published- "articles," "literature," "papers," and "publications" throughout the paper. Figure 1 illustrates the flowchart that incorporates the steps involved in this study.

RESULTS

Scientific output based on searched keywords

The first set of analyses provides the scientific output statistics based on the searched keywords on Scopus. Table 1 provides the results from the initial study of the number of articles associated with the four keywords. This analysis was based on the keywords that explicitly appeared in the titles of articles. The term "citizen science" (*n*=1609, 95.83%) is most frequently mentioned in the article titles by the authors. The use of the other three keywords was relatively low.

Table 1: Number of articles based on searched keywords on Scopus

R	Searched keywords	TA	Percentage
1	Citizen science	1609	95.83%
2	Community science	53	3.16%
3	Civic science	10	0.60%
4	Crowd science	7	0.42%

R= Rank; TA= Total Articles

Table 2: List of top-ten journals.

R	Journal title	TA	Percentage	TC	*h*-index	SJR score	Quartile
1	PLOS One	72	4.29%	2318	25	0.85	Q1
2	Biological Conservation	57	3.39%	3346	26	2.14	Q1
3	Journal of Science Communication	42	2.50%	579	15	0.42	Q2
4	Sustainability	41	2.44%	182	7	0.66	Q1
5	Citizen Science: Theory and Practice	37	2.20%	125	7	0.61	Q1
6	Scientific Reports	29	1.73%	407	11	1.01	Q1
7	Marine Pollution Bulletin	21	1.25%	434	14	1.51	Q1
8	Science of the Total Environment	21	1.25%	592	15	1.81	Q1
9	Ecology and Evolution	18	1.07%	115	5	0.72	Q2
10	International Journal of Environmental Research and Public Health	16	0.95%	166	7	0.81	Q1

TC= Total Citations

Scientific output during 2012-2021

Figure 2 depicts the annual and exponential growth of scientific output indexed in Scopus. It demonstrates CS-related research interests. After 2014, we saw a sharp increase (104.26%) in published articles. The average annual growth is approximately 40.68%. We also computed the exponential growth of the papers at an annual growth rate of 28.96%. It ascertains that the application of CS methods in research processes has been rapidly adopted. Nearly 61% of articles were published between 2019 and 2021. Also, we can determine the most articles published in the year 2021 (*n*=464), followed by 2020 (*n*=300) and 2019 (*n*=267).

Significant journals

It is essential to highlight the journals where the researchers opted to publish their work. Our dataset encompassed 674 journals in total. We identified the top 10 journals that published 354 (21.08%) articles in various citizen science research contexts (Table 2). We additionally provide the Scimago Journal and Country Rank (https://www.scimagojr.com/) metrics. We believe it may offer valuable information on the importance of a distinct journal. The journals that cover essentially the same disciplines have been consolidated together. First, PLOS One (top-ranked, *n*=72, 4.29%) and Scientific Reports (*n*=29, 1.73%) cover multi-disciplines. Second, Biological Conservation (*n*=57, 3.39%), Marine Pollution Bulletin (*n*=21, 1.25%), Science of the Total Environment (*n*=21, 1.25%), Ecology and Evolution (*n*=18, 1.07%), and International Journal of Environmental Research and Public Health (*n*=16, 0.95%) focus on agriculture, biological, and environmental sciences. In addition, Biological Conservation received most citations (TC=3346, *h*-index=26). Third, Sustainability (*n*=41, 2.44%) publishes sustainable development research. Fourth, Citizen Science: Theory and Practice (*n*=37, 2.20%) emphasises CS theoretical frameworks and applications. Fifth, the Journal of Science Communication (*n*=42, 2.50%) deals with science communication research.

Country-wise scientific output

The top-ten countries contributing to CS research are distributed in Table 3. We found a total of 129 distinct countries that produced CS research articles. The data in Table 3 shows that the majority of the articles (*n*=639, 38.06%) were produced in the USA. UK (*n*=368, 21.92%) came second in the list. Australia produced 171 (10.18%) research papers, followed by Germany (*n*=125, 7.44%), Italy (*n*=120, 7.15%), Canada (*n*=105, 6.25%), and Spain (*n*=102, 6.08%). The remaining three countries contributed fewer than 100 articles. According to the statistics on citations, the USA got the most citations (*n* =13955).

Table 3: List of top-ten countries contributing to citizen science research.

R	Country	TA	Percentage	TC	ACPA
1	USA	639	38.06%	13955	21.84
2	UK	368	21.92%	8552	23.24
3	Australia	171	10.18%	3297	19.28
4	Germany	125	7.44%	2164	17.31
5	Italy	120	7.15%	2187	18.23
6	Canada	105	6.25%	2436	23.20
7	Spain	102	6.08%	1647	16.15
8	Netherlands	99	5.90%	2343	23.67
9	France	64	3.81%	1701	26.58
10	Austria	57	3.39%	1325	23.25

ACPA= Average Citation Per Article

Author productivity

We present Table 4, which shows the statistics of the top-ten productive authors and their impact on CS research. We found a total of 7463 authors in our dataset. As shown in Table 4, Callaghan, C.T. was the most productive (*n*=17) author in the domain of CS, with 236 citations and an *h*-index of 9. King, A.C. was the second most productive author (*n*=14, TC=210, *h*-index=9). The author with the third-highest output was Porfiri, M. (*n* = 13, TC = 277, *h*-index= 9). Interestingly, the top three authors who received more than 1000 citations for 32 research papers were Parrish, J.K., Wiggins, A., and Ballard, H.L. We closely inspected the articles (*n*=120) of the top-ten authors to identify the core research areas. The articles focused on eight research areas: birds, species, health, social-ecological system, application of machine learning, biodiversity, data quality, and natural resource management. Table 4 also reveals the author-level metrics. It indicates that the average *h*-index of each author is approximately 9. Parrish, J.K. and Ballard, H.L. had more than 100 ACPA.

Significant institutions

Table 5 provides the data of the top-ten leading institutions that produced the most articles. It measures the research activities of the colleges and universities in CS research. Table 5 shows that the University of California (*n*=53, 3.16%) and the University of Oxford (*n*=40, 2.38%) ranked first and second. Cornell University (*n*=37, 2.20%), Colorado State University (*n*=33, 1.97%), Imperial College London (*n*=32, 1.91%), and Wageningen University and Research (*n*=31, 1.85%) made impressive contributions. The remaining four universities that gained the top-ten spot contributed 110 (6.55%) articles. The majority of the institutions were from the USA.

Acknowledged funding sponsors

Universities and research funding sponsors strive to ensure that the research they fund has the biggest potential influence on the field of study (Achimas Cadariu, 2012). First, this section reveals the most frequently acknowledged funding sources. The data in Table 6 (columns 2 and 3) was prepared using the data of Scopus database. Our dataset confirmed that there was no funding acknowledgement information in 697 (41.51%) research papers. After analysing the dataset, we observed that the National Science Foundation (NSF) was the most prominent funder in CS research, with 157 (9.35%) articles. The European Commission (EC) and the Horizon Programme were the second and third most acknowledged sponsors, with 93 (5.54%) and 49 (2.92%) articles, respectively. However, the Horizon Programme was the programme of the EC (*Horizon 2020*, n.d.). Natural Environment Research Council was the fourth most frequent funder, and 46 (2.74%) articles were published under the sponsorship. UKRI appeared in the top five, with 36 (2.14%) articles. Funders from the USA, like NASA and NIH, also entered the top-ten list. The rest is on the list from Canada, Germany, and the UK. Second, we show focused research areas. Based on author keywords, we analysed and presented the top-five focused areas of research like species, sustainable development, climate change, ecology, and natural resource management. A total of 504 articles (Table 6) are closely associated with these research areas. For better understanding, the most frequent three keywords were assigned to each funding sponsor. Next, Figure 3 shows the evolution of the scientific output supported by the top-ten funders. It shows how consistently each sponsor stimulates researchers to conduct research on a variety of CS-related projects.

Citation analysis

This section provides the results of the citation analysis. First, we discovered the five most frequently cited documents (out of 1679 documents) that received 1746 citations (see Table 7A). One article received more than 400 citations, three articles received more than 300 citations, and one article received more than 200 citations. The top-ranked article was "Citizen science can improve conservation science, natural resource management, and

Table 4: List of top-ten productive authors.

R	Author	TA	TC	ACPA	*h*-index
1	Callaghan, C.T.	17	236	13.88	9
2	King, A.C.	14	210	15.00	9
3	Porfiri, M.	13	277	21.31	9
4	Lintott, C.	12	499	41.58	10
5	Haklay, M.	12	297	24.75	8
6	Parrish, J.K.	11	1208	109.82	9
7	Wiggins, A.	11	1049	95.36	9
8	Ballard, H.L.	10	1201	120.10	8
9	See, L.	10	832	83.20	7
10	Roy, D.B.	10	582	58.20	8

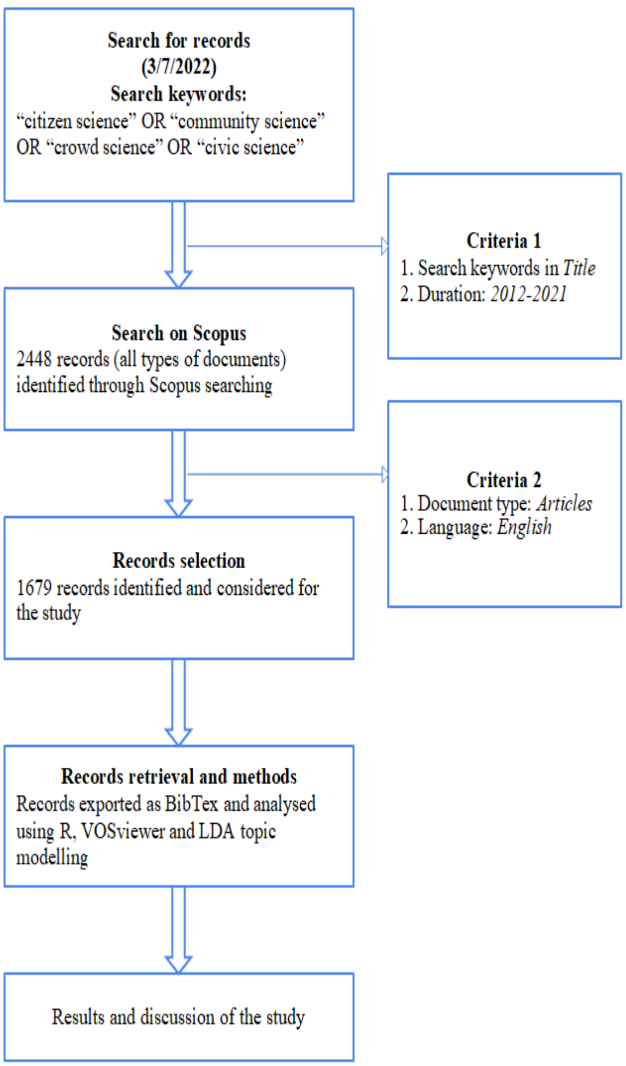

Figure 1: Flowchart of the study.

environmental protection", authored by McKinley *et al.* (2017). It was published in 2017. It has received 426 citations since it was published. The article's Average Citation Per Year (ACPY) is 71, whereas the ACPY of the other four articles range from 40 to 50. Biological Conservation published three out of the top-five articles. Second, we found 80790 references in our dataset and presented the top-five cited references (see Table 7B). The most frequently cited reference was an article by Bonney *et al.* (2009b) entitled "Citizen science: a developing tool for expanding science knowledge and scientific literacy," which received 349 citations. The ACPY of that article is 24.93, and the remaining four articles are between 14 and 25. One of the most intriguing revelations was that the contribution of Bonney, R. was found in both cited documents and cited references.

Network analysis and visualisation

The network analysis of two distinct maps–author and country collaboration–is shown in this section. Figure 4 illustrates the most significant authorship collaborations in CS research during 2012-2021. A document threshold of 5 documents published by authors was adopted to visualise the co-authorship map. Eight different clusters represent collaboration and communication across diverse CS scientific fields. The varying bubble sizes on the map show the strength of the scientific output and its impact on the processes of CS research. We identified Cluster 1, which consisted of 10 authors and was displayed in red bubbles using VOSviewer. The visualisation was based on the weights of the documents published by 55 authors. Heigl, F., Fink, D., Roy, D.B., See, L., Haklay, M., Parrish, J.K., Lintott, C.J., and Kuchner, M.J. were discovered as the most active authors from each cluster. Most of the authors have appeared in Table 4.

Figure 5 depicts the network map of country collaboration. We have already analysed the scientific productivity and impact of the top-ten contributing countries in Table 3. This map supports the earlier analysis. Hence, the top-fifty countries were selected to represent the map visually. The components of the map were the number of articles, citations, and link strength. Cluster 1 (red bubbles) incorporated 16 countries from Europe. The map clearly demonstrates the decent collaboration between the nations of North America, South America, Europe, and several Asian nations. Additionally, although the United States and the

Table 5: List of top-ten significant institutions.

R	Name of the institution	Country	TA	Percentage
1	University of California	USA	53	3.16%
2	University of Oxford	UK	40	2.38%
3	Cornell University	USA	37	2.20%
4	Colorado State University	USA	33	1.97%
5	Imperial College London	UK	32	1.91%
6	Wageningen University and Research	Netherlands	31	1.85%
7	North Carolina State University	USA	29	1.73%
8	University of Washington	USA	28	1.67%
9	School of Biological	Australia	27	1.61%
10	University of Florida	USA	26	1.55%

United Kingdom had the most articles, Cluster 5 (light purple) was determined to be relatively denser than other clusters.

Topic analysis

In this study, we employed the LDA to carry out topic modelling, which enables topic discovery from a set of textual data. As mentioned earlier, we selected five human-interpretable topics by inspecting ten keywords for each topic generated by the model. The left panel of Figure 6 (A) exhibits the multidimensional scaling of the five topics shown as bubbles, while the right panel (B) reveals the top keywords. The model was run multiple times to obtain the desired latent topics. Finally, the five topics were discovered. The bubbles in Figure 6 represent five different topics. Hence, Bubble 1 signifies Topic 1 and is bigger than the others. It means it has a more significant proportion of article titles in the corpus. For instance, if we interpret Topic 1—which includes the word "bird," "scale," "project," "population," and "survey"—we may state that extensive research has been done on either "Bird Distribution" or "Birds Survey." Topic 2 is associated with "Biodiversity," emphasising the sustainable development approach by carrying keywords like "biodiversity" and "conservation." The classification of species is covered in Topic 3. Next, "Natural Resource Management" is addressed in Topic 4, which includes keywords like "water," "quality," "ecology," and "area." Topic 5 is associated with "Public engagement" since it contains the words "project," "environmental," "program," and "participation." These topics confirm that most CS research focuses on the participation or engagement of people in CS. However, the keywords of each topic may refer to more than one topic. We consider the topics highlighted to be suitably pertinent.

DISCUSSION AND CONCLUSION

This section synthesises the findings of the study. The growth of research publications increased year by year. Between 2012 and 2014, fewer CS research articles (*n*=103, 6.13%) were published than the subsequent seven years. It is obvious that CS research interests were limited. However, from 2015 to 2018, there were more publications (*n*=545, 32.46%), which was a sign of the scientific community's growing interest in CS research. The most significant period, when the maximum number of articles (*n*=1031, 61.41%) were published, was from 2019 to 2021. We identified PLOS One as the most influential journal, publishing the highest number of articles. Applications of CS expanded into the humanities and social sciences (Tauginienė *et al*., 2020). It is possible to assert that CS has been incorporated into many academic fields. PLOS One is a multidisciplinary journal, so it tends to have more publications in CS. The result regarding country-wise scientific output reported that the USA was the most active country producing CS research. It reflects the findings of a couple of studies (Chaubey & Singh, 2021; Pelacho *et al*., 2021) which also found that the USA was the most productive country. However, European countries like the UK, Germany, Spain, and the Netherlands have made important contributions. Furthermore, it can be said that developed countries are more inclined towards CS research.

Our analysis showed some interesting patterns in terms of research productivity and impact among the top-ten authors. Callaghan, C.T. had the highest total number of articles, but the lowest ACPA (13.88). It suggests that Callaghan, C.T. is a highly productive author, but may not be as highly cited or impactful as from other authors. On the other hand, Parrish, J.K., Wiggins, A., and Ballard, H.L. had fewer articles than the top-five authors, but highest ACPA. This suggests that their articles are more impactful. Our further analysis revealed that the University of California and the University of Oxford were the most active institutions with the maximum publications. However, most institutions in the top ten belonged to the USA. It shows that the research culture and infrastructure of the institutions of the USA might be good.

CS covers several projects on different topics (National Geographic Society, 2022). We can assume that funding agencies may support scientific outputs or projects. A total of 988 articles mentioning funding acknowledgement were found in our dataset. The National Science Foundation, Horizon Programme and European Commission were highly acknowledged funding sponsors and

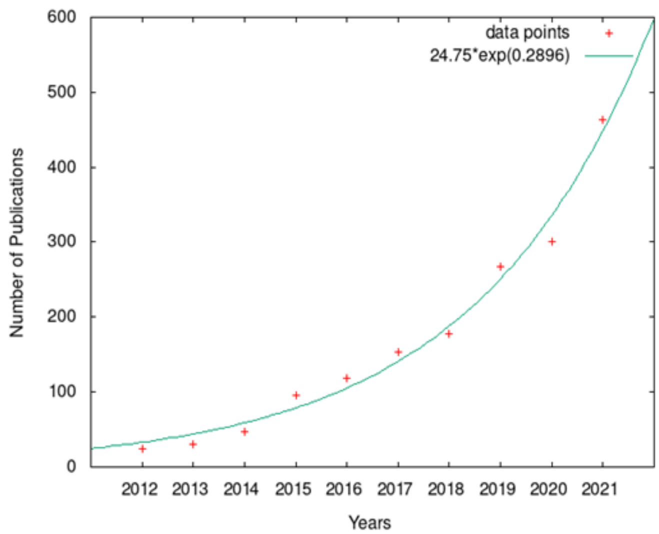

Figure 2: Scientific output in CS and exponential growth.

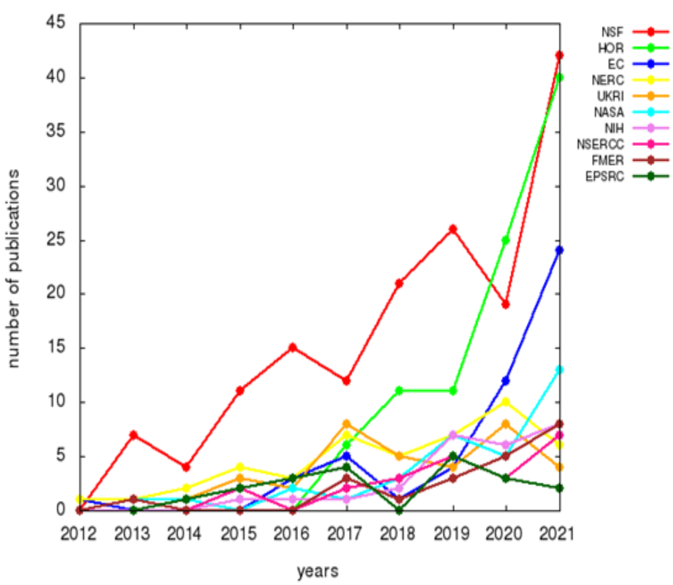

Figure 3: Evolution of the scientific output supported by funding sponsors.

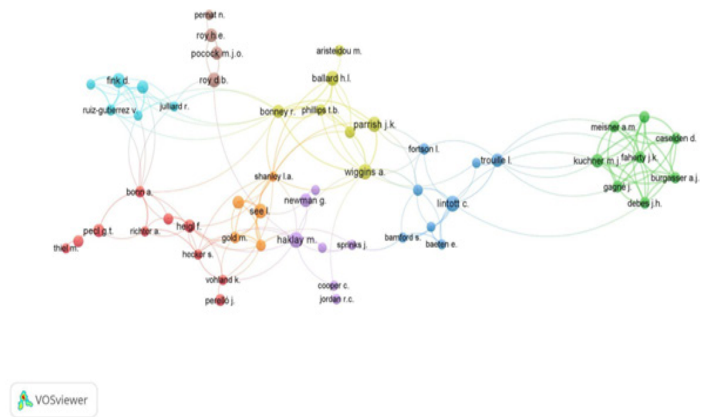

Figure 4: Author collaboration map in CS research (2012-2021). Map characteristics: 8 clusters, 55 items (authors), association method for normalisation, 164 links, total link strength– 446.

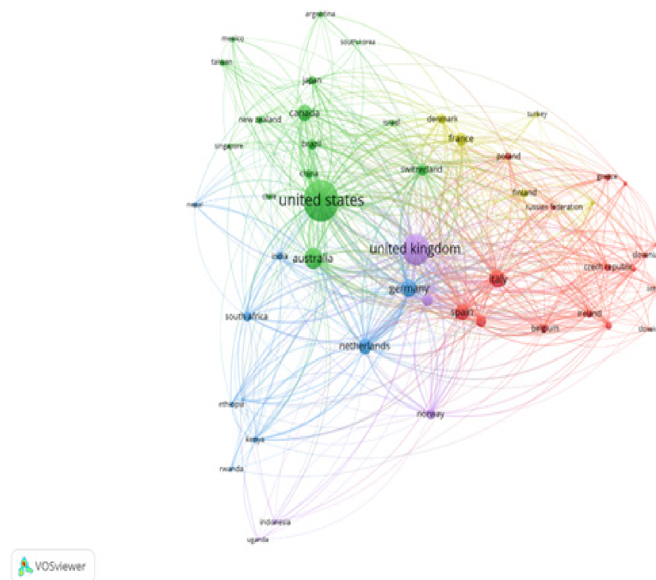

Figure 5: Country collaboration map in CS research (2012-2021). Map characteristics: 5 clusters, 50 items (countries), association method for normalisation, 590 links, and total link strength-2178.

programmes. Regarding top sponsors, we discovered that the Horizon Programme was ranked second highest and received the highest acknowledgement since 2016. The Horizon Programme, which funded research and innovation to address major global issues like climate change, started in 2014 and ended in 2020 (*Horizon 2020*, n.d.).

In this study, the most frequently cited documents and cited references were assessed. The most cited document (426 citations in Scopus) focused on the application of CS in "conservation science, natural resource management, and environmental protection." It was published in the journal Biological Conservation in 2017 and authored by McKinley *et al.* (2017). Following that, our further analysis determined the most cited references. Documents in our dataset (i.e., 1679 documents) frequently cited the article (349 citations in Scopus) authored by Bonney *et al.* (2009b) that was published in the journal BioScience in 2009. In the article, the authors introduced the "Citizen science program model," which can aid CS project developers in recruitment, education, conservation, information science, and statistics.

We mapped the author (Figure 4) and country (Figure 5) collaboration network using VOSviewer. The author collaboration map found the eight most influential authors with the highest link strengths. The most active collaboration was between Kuchner, M.J., Faherty, J.K., Meisner, A.M., Caselden, D., Debes, J.H., Gagne, J., and Burgasser, A.J. Other strong links were found between Fink, D., Ruiz-Gutierrez, V., and Julliard,

Table 6: List of top-ten acknowledged funding sponsors.

R	Funding sponsors/ sources/ programmes	TA	Percentage	Focused research areas
1	National Science Foundation (NSF*), USA	157	9.35%	Data quality; Species; Science policy.
2	Horizon Programme (HOR**), UK	93	5.54%	Sustainable development; Technology application; Ecology.
3	European Commission (EC*), Belgium	49	2.92%	Sustainable development; Impact assessment; Ecology.
4	Natural Environment Research Council (NERC*), UK	46	2.74%	Pollution; Species; Climate change.
5	UK Research and Innovation (UKRI*), UK	36	2.14%	Species; Natural resource management; Conservation.
6	The National Aeronautics and Space Administration (NASA*), USA	33	1.97%	Species; Natural resource management; Climate change.
7	National Institute of Health (NIH*), USA	26	1.55%	Digital health; Environmental justice; Physical activity.
8	Natural Sciences and Engineering Research Council of Canada (NSERCC*)	23	1.37%	Species; Climate change; Water quality.
9	Federal Ministry of Education and Research (BMBF*, FMER**), Germany	21	1.25%	Disease; Science literacy; Sustainable development.
10	Engineering and Physical Sciences Research Council (EPSRC*), UK	20	1.19%	Biological recording; Urban agriculture; Sustainable development.

Note: *Some of the acronyms of funding sponsors are official and **a couple of acronyms were given by us.

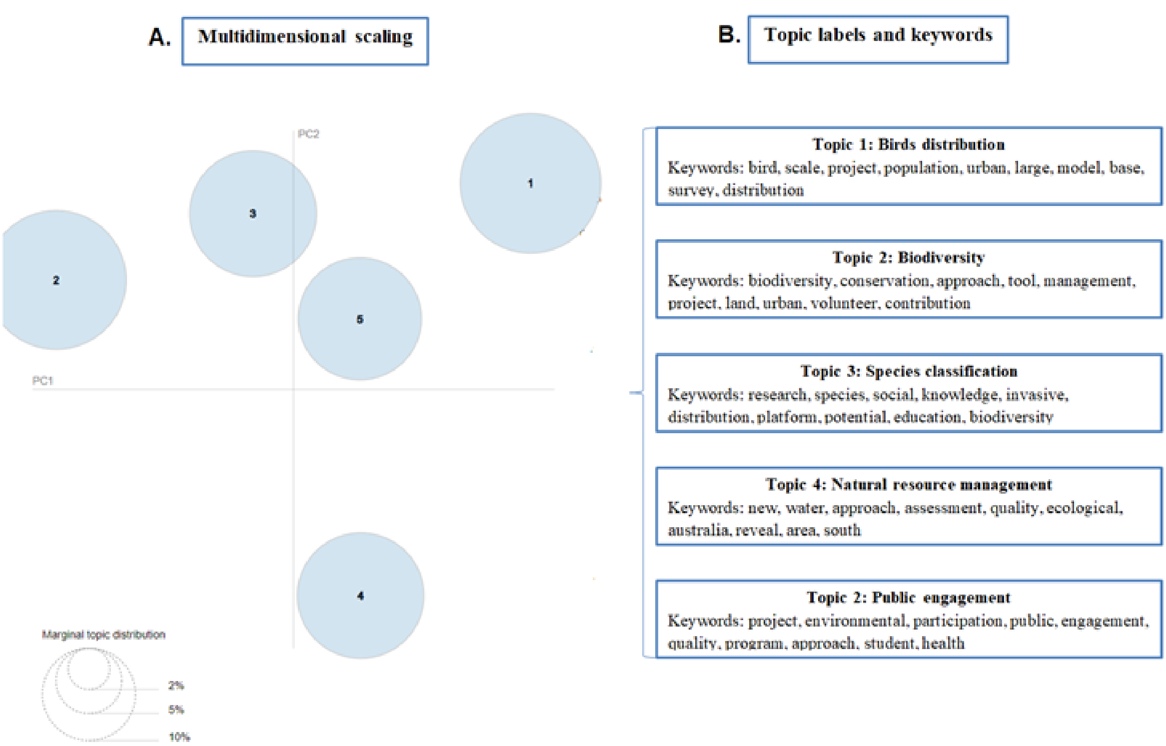

Figure 6: (A) Multidimensional scaling of five intertopics using pyLDAvis. The number inside the bubbles are topic numbers. (B) Topic labels and ten keywords associated with each Topic.

Table 7A: List of top-five cited documents.

R	Authors	Bibliographic information	TC	ACPY
1	McKinley *et al.* (2017)	McKinley, D. C., Miller-Rushing, A. J., Ballard, H. L., Bonney, R., Brown, H., Cook-Patton, S. C., Evans, D. M., French, R. A., Parrish, J. K., Phillips, T. B., Ryan, S. F., Shanley, L. A., Shirk, J. L., Stepenuck, K. F., Weltzin, J. F., Wiggins, A., Boyle, O. D., Briggs, R. D., Chapin, S. F., … Soukup, M. A. (2017). Citizen science can improve conservation science, natural resource management, and environmental protection. Biological Conservation, 208, 15–28.	426	71
2	Theobald *et al.* (2015)	Theobald, E. J., Ettinger, A. K., Burgess, H. K., DeBey, L. B., Schmidt, N. R., Froehlich, H. E., Wagner, C., HilleRisLambers, J., Tewksbury, J., Harsch, M. A., & Parrish, J. K. (2015). Global change and local solutions: Tapping the unrealized potential of citizen science for biodiversity research. Biological Conservation, 181, 236–244.	376	47
3	Bonney *et al.* (2016)	Bonney, R., Phillips, T. B., Ballard, H. L., & Enck, J. W. (2016). Can citizen science enhance public understanding of science? Public Understanding of Science, 25(1), 2–16.	331	47.29
4	Kosmala *et al.* (2016)	Kosmala, M., Wiggins, A., Swanson, A., & Simmons, B. (2016). Assessing data quality in citizen science. Frontiers in Ecology and the Environment, 14(10), 551–560.	318	45.43
5	Chandler *et al.* (2017)	Chandler, M., See, L., Copas, K., Bonde, A. M. Z., López, B. C., Danielsen, F., Legind, J. K., Masinde, S., Miller-Rushing, A. J., Newman, G., Rosemartin, A., & Turak, E. (2017). Contribution of citizen science towards international biodiversity monitoring. Biological Conservation, 213, 280–294	295	49.17

Table 7B: List of top-five cited references.

R	Authors	Bibliographic information	TC	ACPY
1	Bonney *et al.* (2009b)	Bonney, R., Cooper, C. B., Dickinson, J., Kelling, S., Phillips, T., Rosenberg, K. V., & Shirk, J. (2009b). Citizen Science: A Developing Tool for Expanding Science Knowledge and Scientific Literacy. BioScience, 59(11), 977–984.	349	24.93
2	Silvertown (2009)	Silvertown, J. (2009). A new dawn for citizen science. Trends in Ecology & Evolution, 24(9), 467–471.	333	23.79
3	Dickinson *et al.* (2010)	Dickinson, J. L., Zuckerberg, B., & Bonter, D. N. (2010). Citizen Science as an Ecological Research Tool: Challenges and Benefits. Annual Review of Ecology, Evolution, and Systematics, 41(1), 149–172.	313	24.08
4	Bonney *et al.* (2014)	Bonney, R., Shirk, J. L., Phillips, T. B., Wiggins, A., Ballard, H. L., Miller-Rushing, A. J., & Parrish, J. K. (2014). Next Steps for Citizen Science. Science, 343(6178), 1436–1437	200	22.22
5	Conrad and Hilchey (2011)	Conrad, C. C., & Hilchey, K. G. (2011). A review of citizen science and community-based environmental monitoring: Issues and opportunities. Environmental Monitoring and Assessment, 176(1), 273–291	171	14.25

R.; Parrish, J.K., Wiggins, A., Ballard, H.L., and Bonney, R.; and See, L., Shanley, L.A., and Gold, M., etc. Finally, inter-cluster collaboration determined influential partnerships, for instance, the link between Fink, D., Bonney, R., Ruiz-Gutierrez, V., Julliard, R., and Bonn, A. Figure 5 already reveals that there is a strong collaboration between the countries from North America, South America, Europe and Asia. The most active links identified between the United States and the United Kingdom. Some other active collaborations were between the United States and Canada; the United Kingdom, Germany, and the Netherlands; Italy, Spain, and Austria; the United States, Australia, and United Kingdom; and Switzerland, France, and Denmark. This particular outcome matches previous studies (Chaubey & Singh, 2021; Pelacho *et al*., 2021) in terms of the most active countries and partnerships in some instances.

The LDA topic modelling technique exposed the five topics from the titles of the articles. For example, one of the top research topics among the researchers was "Birds Distribution" or "Birds Survey," where citizens engaged as volunteers to survey birds and collect data for analysing the population of birds (McKinley *et al*., 2017). Surveying birds is one of the most popular CS projects. Breeding Bird Survey (BBS) is one of the significant examples that we observed (Bonney *et al*., 2016; McKinley *et al*., 2017). Another popular research topic was "Biodiversity" (Topic 2). Prior studies (Bedessem *et al*., 2021; Chaubey & Singh, 2021; Kullenberg & Kasperowski, 2016; Pelacho *et al*., 2021) used other methods and identified that there had been significant research on Topic 2. There are several scopes (e.g., "spatial extent," "taxonomic and system breadth," and "economic worth") of biodiversity CS and in the last 30 years, biodiversity-based CS projects have sharply increased, which resulted in scientific output (Theobald *et al*., 2015). Topic 3 was about the classification of species. Researchers of the 1679 documents showed extensive interest in studying species distribution or management. Community-Based Monitoring (CBM) programs are such examples that involve phenology-oriented research (Chandler *et al*., 2017). The focus of Topic 4 was "Natural Resource Management." As people could help scientists by producing scientific information for natural resource management and decision-makers, it kept substantial research interests among scientific communities involved in CS (McKinley *et al*., 2017). Finally, Topic 5 emphasised "Public engagement," which enables gathering massive amounts of scientific data from various regions to examine large-scale natural patterns (Bonney *et al*., 2009b). This novel result is emphatically unique and has not been presented in earlier studies.

The implications of our results could have a significant impact on CS researchers, policymakers, practitioners, educators, funding agencies, and science lovers. In CS, anyone can assist scientists. Therefore, this study also cited how the engagement of people could be made in arrays of scientific endeavours (Bonney *et al*., 2009b). The results highlighted that the interest in CS research has been increased; consequently, CS researchers can explore the research trends, funding agencies, and research areas in which they can plan for their further studies. Moreover, future researchers can identify the research collaboration patterns and ask for international collaboration.

Our study has some limitations, such as the size of the dataset, research period, limited search terms, and ambiguity regarding exported bibliographic data. Nonetheless, the analyses performed here in this study provide key outcomes. The potential for future research is broad in the environmental and biological sciences, as CS researchers are extensively researching these domains, according to our study. Future scientometrics studies may conduct co-citation analysis, co-occurrence analysis, citation burst analysis, cluster analysis, and impact factor analysis. Our analysis revealed that the countries with the highest levels of CS research are from Europe, North America (especially the United States), and South America. Therefore, one of the possible analyses could be a comparative study between these continents or specific countries, depending on the researchers' interests. We only used our search terms that occurred in the "Title" of the articles. Hence, the volume of the dataset for a similar study can be increased by picking some parameters, such as (a) search terms that occurred in keywords, titles, and abstracts in more than one bibliographic databases; (b) including both articles and reviews; and (c) including more search keywords (Pelacho *et al*., 2021). These approaches could significantly contribute to this particular field of study.

CONFLICT OF INTEREST

The authors declare that there is no conflict of interest.

REFERENCES

Achimas CA. (2012) Global Access to Knowledge. In R. Chadwick (Ed.). *Encyclopedia of Applied Ethics (Second Edition)* (pp. 497–503). Academic Press. https://doi.org/10.1016/B978-0-12-373932-2.00066-1

Bautista-Puig N, Filippo DD, Mauleón E, Sanz-Casado E. (2019) Scientific Landscape of Citizen Science Publications: Dynamics, Content and Presence in Social Media. *Publications. 7*(1), Article 1. https://doi.org/10.3390/publications7010012

Bedessem B, Julliard R, Montuschi E. (2021) Measuring epistemic success of a biodiversity citizen science program: A citation study. *PLOS ONE, 16*(10), e0258350. https://doi.org/10.1371/journal.pone.0258350

Blei DM, Ng AY, Jordan MI. (2003) Latent dirichlet allocation. *The Journal of Machine Learning Research, 3*(null), 993–1022.

Bonney R, Ballard HL, Jordan RC, McCallie E, Phillips TB, Shirk J, Wilderman CC. (2009a) *Public Participation in Scientific Research: Defining the Field and Assessing Its Potential for Informal Science Education. A CAISE Inquiry Group Report*.

Bonney R, Cooper CB, Dickinson J, Kelling S, Phillips T, Rosenberg KV, Shirk J. (2009b) Citizen Science: A Developing Tool for Expanding Science Knowledge and Scientific Literacy. *BioScience, 59*(11), 977–984. https://doi.org/10.1525/bio.2009.59.11.9

Bonney R, Phillips TB, Ballard HL, Enck JW. (2016) Can citizen science enhance public understanding of science? *Public Understanding of Science, 25*(1), 2–16. https://doi.org/10.1177/0963662515607406

Bonney R, Shirk JL, Phillips TB, Wiggins A, Ballard HL, Miller-Rushing AJ, Parrish JK. (2014) Next Steps for Citizen Science. *Science, 343*(6178), 1436–1437. https://doi.org/10.1126/science.1251554

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. https://doi.org/10.1080/09500690500069483

Chandler, M., See, L., Copas, K., Bonde, A. M. Z., López, B. C., Danielsen, F., Legind, J. K., Masinde, S., Miller-Rushing, A. J., Newman, G., Rosemartin, A. & Turak, E. (2017) Contribution of citizen science towards international biodiversity monitoring. *Biological Conservation, 213*, 280–294. https://doi.org/10.1016/j.biocon.2016.09.004

Chaubey, A. & Singh, A. (2021) Analysis of citizen science scientific publications: A scientometric study. *Library Philosophy and Practice (e-Journal)*. https://digitalcommons.unl.edu/libphilprac/5645

Cohn, J. P. (2008) Citizen Science: Can Volunteers Do Real Research? *BioScience, 58*(3), 192–197. https://doi.org/10.1641/B580303

Conrad, C. C. & Hilchey, K. G. (2011) A review of citizen science and community-based environmental monitoring: Issues and opportunities. *Environmental Monitoring and Assessment, 176*(1), 273–291. https://doi.org/10.1007/s10661-010-1582-5

Cooper, I. D. (2015) Bibliometrics basics. *Journal of the Medical Library Association : JMLA, 103*(4), 217–218. https://doi.org/10.3163/1536-5050.103.4.013

Daily Zooniverse. (2014, September 16) 'Citizen science' added to Oxford English Dictionary. *Daily Zooniverse*. https://daily.zooniverse.org/2014/09/16/citizen-science-in-dictionary/

De Filippo, D., Lascurain, M. L., Pandiella-Dominique, A. & Sanz-Casado, E. (2020a) Scientometric Analysis of Research in Energy Efficiency and Citizen Science through

Projects and Publications. *Sustainability, 12*(12), Article 12. https://doi.org/10.3390/su12125175

De Filippo, D., Sanz Casado, E., Berteni, F., Barisani, F., Bautista Puig, N. & Grossi, G. (2020b) Assessing citizen science methods in IWRM for a new science shop: A bibliometric approach. *Hydrological Sciences Journal, 65*(S3), 179–192. https://doi.org/10.1080/02626667.2020.1851691

Dickinson, J. L., Zuckerberg, B. & Bonter, D. N. (2010) Citizen Science as an Ecological Research Tool: Challenges and Benefits. *Annual Review of Ecology, Evolution, and Systematics, 41*(1), 149–172. https://doi.org/10.1146/annurev-ecolsys-102209-144636

Elsevier. (n.d.) *About Scopus—Abstract and citation database | Elsevier*. Retrieved 14 July 2022, from https://www.elsevier.com/solutions/scopus?dgcid=RN_AGCM_Sourced_300005030

Follett, R. & Strezov, V. (2015) An Analysis of Citizen Science Based Research: Usage and Publication Patterns. *PLOS ONE, 10*(11), e0143687. https://doi.org/10.1371/journal.pone.0143687

Hajibayova, L., Coladangelo, L. P. & Soyka, H. A. (2021) Exploring the invisible college of citizen science: Questions, methods and contributions. *Scientometrics, 126*(8), 6989–7003. https://doi.org/10.1007/s11192-021-04050-6

Horizon 2020. (n.d.) European Commission - European Commission. Retrieved 18 July 2022, from https://ec.europa.eu/info/research-and-innovation/funding/funding-opportunities/funding-programmes-and-open-calls/horizon-2020_en

Irwin, A. (1995) *Citizen Science: A Study of People, Expertise and Sustainable Development*. Routledge.

Kosmala, M., Wiggins, A., Swanson, A. & Simmons, B. (2016) Assessing data quality in citizen science. *Frontiers in Ecology and the Environment, 14*(10), 551–560. https://doi.org/10.1002/fee.1436

Kullenberg, C. & Kasperowski, D. (2016) What Is Citizen Science? – A Scientometric Meta-Analysis. *PLOS ONE, 11*(1), e0147152. https://doi.org/10.1371/journal.pone.0147152

Lamba, M. & Madhusudhan, M. (2019) Mapping of topics in DESIDOC Journal of Library and Information Technology, India: A study. *Scientometrics, 120*(2), 477–505. https://doi.org/10.1007/s11192-019-03137-5

Mazumder, S. & Barui, T. (2021) Discovering Topics from the Titles of the Indian LIS Theses. *Library Philosophy and Practice (e-Journal)*. https://digitalcommons.unl.edu/libphilprac/5924

McKinley, D. C., Miller-Rushing, A. J., Ballard, H. L., Bonney, R., Brown, H., Cook-Patton, S. C., Evans, D. M., French, R. A., Parrish, J. K., Phillips, T. B., Ryan, S. F., Shanley, L. A., Shirk, J. L., Stepenuck, K. F., Weltzin, J. F., Wiggins, A., Boyle, O. D., Briggs, R. D., Chapin, S. F., … Soukup, M. A. (2017) Citizen science can improve conservation science, natural resource management, and environmental protection. *Biological Conservation, 208*, 15–28. https://doi.org/10.1016/j.biocon.2016.05.015

Moral-Muñoz, J. A., Herrera-Viedma, E., Santisteban-Espejo, A. & Cobo, M. J. (2020) Software tools for conducting bibliometric analysis in science: An up-to-date review. *Profesional de La Información, 29*(1), Article 1. https://doi.org/10.3145/epi.2020.ene.03

National Geographic Society. (2022, May 22) *Citizen Science*. Encyclopedia. https://education.nationalgeographic.org/resource/citizen-science

Newman, G., Wiggins, A., Crall, A., Graham, E., Newman, S. & Crowston, K. (2012) The future of citizen science: Emerging technologies and shifting paradigms. *Frontiers in Ecology and the Environment, 10*(6), 298–304. https://doi.org/10.1890/110294

Odenwald, S. (2018) A Citation Study of Citizen Science Projects in Space Science and Astronomy. *Citizen Science: Theory and Practice, 3*(2), Article 2. https://doi.org/10.5334/cstp.152

Odenwald, S. F. (2020) A citation study of earth science projects in citizen science. *PLOS ONE, 15*(7), e0235265. https://doi.org/10.1371/journal.pone.0235265

Pelacho, M., Ruiz, G., Sanz, F., Tarancón, A. & Clemente-Gallardo, J. (2021) Analysis of the evolution and collaboration networks of citizen science scientific publications. *Scientometrics, 126*(1), 225–257. https://doi.org/10.1007/s11192-020-03724-x

REGEXEXTRACT. (n.d.) Retrieved 14 July 2022, from https://support.google.com/docs/answer/3098244?hl=en

Rehurek, R. & Sojka, P. (2010) Software Framework for Topic Modelling with Large Corpora. *Proceedings of the LREC 2010 Workshop on New Challenges for NLP Frameworks*, 40–50.

SciStarter. (n.d.) *What is Citizen Science*. SciStarter. Retrieved 11 July 2022, from https://scistarter.org/citizen-science

Silvertown, J. (2009) A new dawn for citizen science. *Trends in Ecology & Evolution, 24*(9), 467–471. https://doi.org/10.1016/j.tree.2009.03.017

Tauginienė, L., Butkevičienė, E., Vohland, K., Heinisch, B., Daskolia, M., Suškevičs, M., Portela, M., Balázs, B. & Prūse, B. (2020) Citizen science in the social sciences and humanities: The power of interdisciplinarity. *Palgrave Communications, 6*(1), Article 1. https://doi.org/10.1057/s41599-020-0471-y

The Audubon Society. (2022) *Christmas Bird Count*. Audubon. https://www.audubon.org/conservation/science/christmas-bird-count

Theobald, E. J., Ettinger, A. K., Burgess, H. K., DeBey, L. B., Schmidt, N. R., Froehlich, H. E., Wagner, C., HilleRisLambers, J., Tewksbury, J., Harsch, M. A. & Parrish, J. K. (2015) Global change and local solutions: Tapping the unrealized potential of citizen science for biodiversity research. *Biological Conservation, 181*, 236–244. https://doi.org/10.1016/j.biocon.2014.10.021

West, S. & Pateman, R. (2016) Recruiting and Retaining Participants in Citizen Science: What Can Be Learned from the Volunteering Literature? *Citizen Science: Theory and Practice, 1*(2), Article 2. https://doi.org/10.5334/cstp.8

Wikipedia contributors. (2022) Citizen science. In *Wikipedia*. 2022. https://en.wikipedia.org/w/index.php?title=Citizen_science&oldid=1097085063

Zhang Xuanhui, Z. Y. & Zhang Xuanhui, Z. Y. (2017) Evolution Path and Hot Topics of Citizen Science Studies. *Data Analysis and Knowledge Discovery, 1*(7), 22–34. https://doi.org/10.11925/infotech.2096-3467.2017.07.04

Cite this article: Mazumder S, Barui T. Mapping Global Scientific Output on Citizen Science from 2012 to 2021: A Scientometric Analysis. Journal of Data Science, Informetrics, and Citation Studies. 2023;2(2):65-76.