CS TRACK
Investigating Citizen Science

Horizon 2020 / Science with and for Society Programme
Grant agreement number: 872522

D2.2: Final documentation of initiatives selected for analysis

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 872522

Title of project	CS Track
Full title of project	Expanding our knowledge on Citizen Science through analytics and analysis
Title of this document	Final documentation of initiatives selected for analysis
Number of this document	D2.2
Dissemination level	Public
Due date	30th November 2022
Actual delivery	30th November 2022
Versioning history	
Author	Coordination: Patricia Santos (UPF) Authors (*ordered alphabetically by institution*): FORTH: Marinos Anastasakis, Katherina Kikis-Papadakis JYU: Kai Nils Weeber, Aaron J. Peltoniemi, Ohto Sabel, Raija Hämäläinen MOFET: Yaela Golumbic, Reuma De Groot, Tslil Farchi, Anne Turbe RIAS: Ulrich Hoppe, Cleo Schulten UPF: Ishari Amarasinghe, Miriam Calvera Isabal, Patricia Santos, Nicolas Gutierrez Paez, Davinia Hernández-Leo URJC: David Roldán-Álvarez, Fernando Martínez-Martínez WiD: Marius Oesterheld, Vincent Schmid-Loertzer WLW: Christine Urban, Michael Strähle
Executive summary	One of the main aims of WP2 has been to compile a database of CS projects in the European Union and Associated Countries and document a collection of these projects to explore their availability of data for further analysis (through WP3 and WP4), which will be based on the criteria established by WP1. This deliverable documents the collection of resulted studies developed in the context of WP2. A total of 13 studies are presented and classified according to four different knowledge gaps identified by WP1. This documentation also contains an overview analysis of the final content collected in the CS Track database including general descriptors, and three detailed sections where the information from the database is classified according to research areas, Sustainable Development Goals and skills of

	science inquiry (aspects that have been identified as relevant by WP3 and WP4). Finally, a final section discussing what we have learned in the process of analysing CS project descriptions, and suggest guidelines for writing clear and informative project descriptions.

Table of Contents

1. Orientation and general approach

The following deliverable presents the work done by WP2. As presented in our previous deliverable "D2.1: Explorative study of CS projects in Europe, categorization and clustering to build a database of CS projects for analysis", in the context of WP2, we have compiled a database (named the CS Track database). The DB has been built by searching for existing CS platforms in Europe, associated countries and other relevant platforms in the world. As a result, a total of 4949 CS projects have been collected.

WP2 takes up the theoretical discussion presented in D1.1 (WP1) while exploring activities and initiatives as potential candidates for further analysis within CS Track. In the literature review presented in D1.1 a set of topics was identified according to existing knowledge gaps in the field of Citizen Science. These topics are: (1) People involved in Citizen Science; (2) Citizen Science and its relations with the science system; (3) Citizen Science and Education; (4) Visibility of Citizen Science; and (5) Economic considerations in Citizen Science.
Regarding the studies and analysis presented in this deliverable, it is important to take into consideration the type of information gathered in our database that mainly relies on CS project descriptions gathered online. Due to the previous analysis presented in D2.1, WP2 already had an idea of the type of topics and research questions we were able to explore further. For this reason, from these five topics we have focused our attention on four. We knew that it was not possible to understand in detail the economic considerations of CS with the information provided in project descriptions. The economic considerations were derived from, and analysed mainly on the basis of, the literature review conducted for D1.1. This, in combination with some additional evidence collected from the study presented in section 3.2 of this deliverable, contributed to the formulation of a policy recommendation addressing the availability of reliable information on CS projects, especially economic information (on costs, funding models, etc.). For these reasons, this topic is not further elaborated in this deliverable.
The results of the corresponding empirical studies are presented in sections 2-6. The studies described in these sections mainly rely on qualitative research, consequently only a selection of CS projects are considered from the total amount included in our DB. Most of the studies included in these sections are completed and have an accepted publication associated with them, or a confidential one (in case that the study has a pending submission to be accepted as a journal or conference publication). Only few studies describe ongoing research that the corresponding leading authors will continue after the submission of this deliverable.

Based on the lessons learned in the empirical studies presented in sections 2-6, we have been able to scale some research aspects of our analysis. These aspects have been analysed by considering a higher number of projects contained in our database (in some cases the total number of projects). In section 7, we present the final overview analysis of projects contained in the CS Track DB. The analysis presented in this section mainly relies on the application of computational analytics methods derived from WP3. The analysis is done according to relevant aspects of research that have been identified during the duration of the project as main aspects to be further understood. These are: (a) Research Areas represented in CS; (b) Sustainable Development Goals; and (c) Skills of science inquiry. In addition to these ones, we include a section discussing other aspects of interest that could be analysed using computational analytical methods as future research.

Finally, based on the knowledge acquired after analysing different CS Platforms and project descriptions, we have been able to identify good practices to be recommended for reporting data from CS. The final section of this deliverable presents a set of guidelines designed and distributed to guide the process of documenting CS projects.

2. Selected CS project studies

The following sections (3, 4, 5 and 6) have been divided into four main topics of interest (based on the ones identified in WP1 D2.1). The topics and the corresponding studies as listed as follows:

(1) **People involved in Citizen Science**:

This topic is analysed through three studies: (a) An explorative study on the effects of the recent pandemic on online Citizen Science: lessons learnt for improving project management and implementation; (b) Very short Questionnaire; and (c) Is it a match? Motivations for citizen science volunteers and recruitment arguments in project descriptions

(2) **Citizen Science and its relations with the science system**:

This topic is analysed through three studies: (a) Investigating the potential of citizen science to respond to emerging challenges - The case of COVID-19; (b) Mapping Sustainable Development Goals to Citizen Science projects; and (c) Tasks in Citizen Science: proposing a hierarchical framework for categorising citizen scientists' activities in CS projects.

(3) **Citizen Science and Education**:

This topic is analysed through three studies: (a) Identifying learning dimensions in CS project descriptions; (b) Learning in citizen science: a triangulation approach; and (c) Educational uses of CS data.

(4) **Visibility of Citizen Science**:

This topic is analysed through two studies: (a) How to automate the extraction and analysis of information for educational purposes; and (b) Citizen science project descriptions as science communication texts - the good, the bad, and the ugly.

Each empirical study tackles the associated main topic from a different perspective and with its corresponding research questions. This deliverable contains an executive summary of each study. Some of these studies are associated with an accepted publication or a full report (pending to be published) that can be consulted for further detail.

Each study is summarised under the structure of:

- Title of the study
- Authors and Research Affiliation
- Period addressed by the study
- Main aim of the study
- Research question/s
- Research Context
- Research Method(s) applied
- Procedure(s) applied
- Summary of results/findings
- Conclusion
- Link to complete report (if any): This can be a link to the accepted publication, a link to Zenodo or other repository.
- Link to dataset (if any): link to Zenodo or another repository.
- Link to source code (if any): link to GitHub.
- References

3. People involved in CS

3.1 An explorative study on the effects of the recent pandemic on online Citizen Science: lessons learnt for improving project management and implementation

Authors and Research Affiliation

Marinos Anastasakis[a] and Kathy Kikis-Papadakis[a]

[a] Foundation for Research and Technology - Hellas (FORTH)

Period addressed by the study

The study concerned Citizen Science activity that took place on Zooniverse between 2015 and 2021 (August) with a particular focus in the months predating and following the declaration of the COVID-19 as a pandemic.

Main aim of the study

The study aimed at assessing the pandemic's impact on online CS participation and capturing CS project coordinators' experience of the pandemic and their actions in managing the pandemic's effects.

Research Questions

- RQ1: How did the pandemic affect participation in online CS projects?

- RQ2: How was the pandemic perceived and experienced by CS coordinators?
- RQ3: How did CS coordinators react in counterbalancing the pandemic's potential effects?

Research Context

Since the pandemic's outbreak, various CS communities and platforms reported a sudden increase in participation during March and April 2020 with some arguing that as restrictions forced people to stay indoors, citizen scientists increasingly turned to online CS projects. Despite that, the pandemic's effects on CS are not well documented, with current literature failing to provide answers about the extent to which the pandemic affected participation in online CS projects and how CS project coordinators actually experienced and managed the changes the pandemic brought.

Research Method(s) applied

The study followed a sequential mixed methods design (Teddlie & Tashakkori, 2009) consisting of two phases (strands), one quantitative and one qualitative. The first phase aimed at assessing the pandemic's impact on online CS participation (RQ1) as well as selecting representative cases (in terms of discipline, geographical distribution, participation and number of contributions per participant) that would inform data collection for the second phase of our research. The second phase aimed at obtaining a deeper understanding of the pandemic's effect on participation (RQ1); capturing CS project coordinators' experience of the pandemic (RQ2) as well as their actions in managing the pandemic's effects (RQ3) by following a multiple case study design (Yin, 2018).

Procedure(s) applied

Data for 332 Zooniverse projects which included the project's name, total number of participants, total number of contributions, discipline and launch date were gathered. Prior to any analysis, all outliers and extreme cases were identified and removed and a total of 259 projects were selected. A cluster analysis on the average number of contributions per participant resulted in a five cluster solution. Following that, coordinators from 36 Zooniverse projects were invited to participate in an interview and nine of them accepted our invitation. The interviews were semi-structured and lasted between 40 and 60 minutes. The interview protocol incorporated questions that reflected our theoretical framework, i.e., questions related to a project's organisational aspects, coordinators' perceptions and experience of the pandemic and response strategies for mitigating the pandemic's effects on the project taken. Interview data were analysed by following a Qualitative Content Analysis approach (Schreier, 2014).

Summary of results/findings

Our findings show that during the COVID-19 pandemic, Zooniverse projects witnessed an increase in the number of contributions made by citizen scientists. The analysis we undertook indicates that this increase was associated with two interrelated factors; the investment of more time by existing citizen scientists and the influx of new ones. In terms of their perceptions and experience, coordinators acknowledged the pandemic as being more an opportunity rather than a threat. More importantly, the interviews highlighted a number of challenges that have been prominent in the pre-pandemic literature but were probably enlarged due to the increased participation:

the need for recognizing citizens' contributions to science, the need for maintain engagement with citizen scientists, the role of CS in promoting scientific literacy and the importance of technology in initiating and supporting online CS projects.

Conclusion

The need for recognizing citizen scientists' efforts and contributions has been well emphasised in literature and our data point towards a specific type of recognition, that of personal recognition. Thus, the notion of a formal recognition of citizen scientists' efforts seems to be a required condition for CS projects and platforms to adopt in the near future. Likewise, given the role of CS as a method for enhancing/complementing science education in schools, implementing formal mechanisms for acknowledging (in this case) students' efforts in a CS project seems to be also vital for engaging future citizens with science. To this end, it is obvious to us that providing a formal recognition to citizen scientists facilitates engagement with CS. In relation to engagement, our results demonstrate that the increase in participation during the COVID-19 pandemic was only temporary and given that no in-person events could be held, as soon as restrictions were eased, participation decreased. As such, engaging citizen scientists through motivating means to which participants' recognition is an essential component, appears to have a positive impact on a project's management and scalability strategies.

Link to complete report

The paper/full report concerning this study is currently under review (on the date of submission of this deliverable - November 2022), no link to a repository is available. Contact the corresponding authors (m.anastasakis@uoc.gr and katerina@iacm.forth.gr) if you have interest to receive further information.

Article on eMagazine: https://cstrack.eu/format/reports/an-explorative-study-on-the-effects-of-the-recent-pandemic-on-online-citizen-science-lessons-learnt-for-improving-project-management/

References

Schreier, M. (2014). Qualitative Content Analysis. In U. Flick (Ed.), The SAGE Handbook of Qualitative Data Analysis (pp. 170-183). SAGE.

Teddlie, C., & Tashakkori, A. (2009). Foundations of Mixed Methods Research: Integrating Quantitative and Qualitative Approaches in the Social and Behavioral Sciences. SAGE Publications, Inc.

Yin, R. K. (2018). Case Study Research and Applications: Design and Methods (Sixth ed.). SAGE.

3.2 Report on a Survey among Organisers of Citizen Science Projects[1]

Authors and Research affiliation

Michael Strähle[a] & Christine Urban[a] *(alphabetical order)*

[1] Explanatory note by the authors: This text is a summary of the *Report on a survey among organisers of citizen science projects* that was published in Zenodo at https://zenodo.org/record/6865659. It is partly identical in wording with the report.

[a] Wissenschaftsladen Wien - Science Shop Vienna

Period addressed by the study

Promotion of the survey started in December 2020, was paused in January 2021 to avoid overlap with the survey in CS Track's Work Package 4 and to solve technical problems with the survey software. It was still possible to respond but since the end of January 2021 no effort was made to promote the survey. The deadline for answering the survey was 18 February 2021.

Main aim of the study

Collect data on some project characteristics which cannot be answered by visiting project websites.

Research question/s

The research questions were:

- What do organisers of citizen science know about participants in their projects? Are they confident to estimate gender, age and social situation?
- How do academic disciplines, attributed to a project, match research expertise in the team of project organisers?
- Are there other response behaviour & response patterns of interest in a survey with a very short questionnaire?

Research Context

It already became evident that essential information on citizen science activities is not easy to find. Thus, the focus lay on issues that were often missing on websites.

Research Method(s) applied

Online survey targeting project owners, resp. coordinators of citizen science projects. The survey was limited to very few questions about the respective project and those who participate(d) in it, when this information was not available online:

- The project objectives,

- the scientific disciplines involved in the project,

- the type(s) of citizen science activities,

- rough estimates on the participation of different social groups, including their gender and age distributions, and

- questions on practical issues, such as the availability of the respective project for further research.

Most questions aimed at project organisers' estimates of numbers or characteristics of participants in their projects. Citizen science project organisers were targeted without pre-selection.

Procedure(s) applied

The survey consisted of ten questions, mainly tick-box questions, this to avoid a barrier for smaller projects with little or no funding. All questions but question 1 were optional ones. The obligatory first question identified a project without doubt.

The remaining nine questions were on estimations of organisers about participants. Time resources are presumably low if there are no employees who can fill in lengthy questionnaires. Additionally, those projects that are keener on being part of citizen science networks would be more inclined to fill in a questionnaire and thus further distort the picture.

For the survey, it was decided to use LimeSurvey because of privacy issues.

The survey was promoted by Twitter messages and a blogpost on *Österreich forscht*, the online platform of *Citizen Science Network Austria*, at the beginning of January 2021. In December 2020 consortium partners promoted the survey in scientific mailing lists and by contacting research and higher education institutions by email directly.

Promotion messages contained a link to the CS Track website.

Summary of results/findings

Completed responses to the survey: 56.

Only three languages were used to answer the questionnaire: English (n=42), German (n=10) and Greek (n=4). Often English was used in spite of the availability of a language version that matched the official languages of the location of the projects. In 50 cases the language version that was accessed initially, was the English one, in six cases it was the German one. This does not indicate the language the questionnaire was finally filled in, as respondents could switch to another language version. It is interesting insofar as the questionnaire was sent out by different partners with respective links to different language versions.

On geographic regions and sites: Where the responding citizen science projects take place is not clear at first sight. Email addresses or domain names do not always localise the projects reliably. It is necessary to visit the project website. Then one finds in most cases, but not in all, a clear-cut answer. The location of the project organisation and geographic outreach of projects were researched separately.

Summary of goals (Question 3): The project organisers referred to projects showing a broad range of activities, settings, goals, involvement intensities, etc. Several projects had objectives related to biodiversity, the environment and/or a combination of both, but in different ways. This had to be expected.

Disciplines in the diverse teams (Question 4): Of the 56 analysed responses, 53 gave an answer to the question while 3 respondents skipped it. Those who answered named 1 to 5 disciplines for their project team which resulted in a total of 162 entries. Each of these entries was manually allocated to both Web of Science (WoS) subareas and the Frascati Manual (FOS) classifications as best as possible.

According to the 6 main categories of the Revised Field of Science and Technology classification in the Frascati Manual, natural scientists were most strongly represented. A first check did not show an obvious mismatch of disciplines in the organisation teams and the WoS-based classification of research areas of the projects named in the CS Track project database. But this may be more easily answered for disciplines in the technical and natural sciences than for the social sciences and the humanities. For the latter, there may be major differences between science traditions and curricula in the different parts of the world. Furthermore, some of the mentioned methods, like participatory action research, may in some contexts need some group dynamical and almost therapeutic psychological knowledge, which cannot be followed up with a questionnaire.

Number of participants roughly estimated by respondents (Questions 5 & 6): All 56 respondents gave an estimate of the numbers of participants at project start. 55 respondents gave an estimate of the numbers of participants at the time responding to the survey or when the project ended.

Asking for two estimates gives a safer impression of the project size than asking for only one as projects can change considerably over time. The answers mirror the very broad range of the size of such projects that one can find in literature as well.

Estimations of gender distribution (Question 7): 47 respondents gave a rough estimation of how many percent of the participants would be male, female or of diverse/other gender. While most of the respondents indicated a rough gender balance, there are a few projects that involve mostly men or women. The percentage of diverse/other gendered participants was estimated in 7 cases.

Estimations of age distribution (Question 8): 45 responses to this question. There are only three projects which indicate 100% for one age group, namely below 18 years. The youngest age group is also highly dominant in 4 additional projects (80% or more of the participants are estimated as being younger than 18 years old) and moderately dominant in another project (65%). The second youngest group (18 – 35 years) is seen as very present, too: They are estimated between 65% and 95% of the participants by 4 organisers. At the other side of the spectrum, we find 2 respondents who estimate that 75% of their projects' participants are older than 60 years.

Estimations of professional status of participants (Question 9): 45 respondents gave feedback to this question, and it most likely could only be answered if a project is targeted to a specific group (i. e. pupils, students) or if the project is small enough that people know each other quite well.

Response patterns: The authors had expected to see a stronger connection between the number of participants and organisers' tendency to give rough estimations of their characteristics. As expected, almost all responding organisers of projects with less than 21 participants answered Questions 7 – 9. It is plausible that in a smaller project those involved know each other personally. But we also see a surprisingly high number of estimations from very large projects (more than 1000 participants) who made a rough estimation.

The answers can neither be regarded as representative for projects that consider themselves as citizen science nor can be safely assumed that they cover the whole spectrum of possible citizen science activities.

For detailed information on the numbers, see the report.

Conclusion

Not too many organisers of citizen science answered the questionnaire. As the questionnaire was very short and would have taken only a few minutes to answer, it is safe to assume that not all non-respondents could answer the questions. There may be several reasons for this, which would merit some more research.

In line with the authors' research so far (e.g., Strähle, Urban et al., 2021), the survey showed a potential indication that many projects do not know very much about the participants, their characteristics or even their number (or not want to admit to it) and refrain from answering. In view of the benefits that several scholars, practitioners, policy makers and others claim citizen science brings with it, this would make some of them unfounded if not even implausible. Moreover, an attempt was made to investigate – in cases where academics were among the organisers - how far their expertises match(ed) the research areas of the projects. This proved exceptionally tricky because there exists no classification scheme which mirrors the broad variety of academic education in different regions.

Link to complete report

Link to the report: https://zenodo.org/record/6865659.

Article on eMagazine: https://cstrack.eu/format/reports/report-on-a-survey-among-organisers-of-citizen-science-projects/

Link to dataset:

Link to dataset: https://zenodo.org/record/7310071

References

Eitzel, M. V., et al. (2017). Citizen Science Terminology Matters: Exploring Key Terms. Citizen Science: Theory and Practice, 2(1), 1. https://doi.org/10.5334/cstp.96.

European Commission (2018). Horizon 2020: Work Programme 2018–2020: 16. Science with and for Society.

https://ec.europa.eu/research/participants/data/ref/h2020/wp/2018-2020/main/h2020-wp1820-swfs_en.pdf.

Heigl, F., & Dörler, D. (2017). Public participation: Time for a definition of citizen science. Nature 551: 168. https://doi.org/10.1038/d41586-017-05745-8.

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

OECD Working Party of National Experts on Science and Technology Indicators (2007). Revised Field of Science and Technology (FoS) Classification in the Frascati Manual.DSTI/EAS/STP/NESTI(2006)19/FINAL.
https://www.oecd.org/science/inno/38235147.pdf.

Riesch, H., & Potter, C. (2014). Citizen science as seen by scientists: Methodological, epistemological and ethical dimensions. Public Understanding of Science 23(1) 107-120.

Strähle, M., Urban, C. et al. (2021). Framework Conceptual Model D1.1. Zenodo. https://doi.org/10.5281/zenodo.5589618.

Strähle, M., Urban, C. et al. (2022). Conceptual Framework for Analytics Tools D1.2. Zenodo. https://doi.org/10.5281/zenodo.6045639.

Strähle, M., & Urban, C. (2022, forthcoming). The Activities & Dimensions Grid of Citizen Science. In: Proceedings of Science. PoS(CitSci2022)087. https://pos.sissa.it/418/087/.

3.3 Availability of information on citizen science activities, checked against the Activities & Dimensions Grid of Citizen Science on the basis of some projects

Authors and Research affiliation

Michael Strähle[a] & Christine Urban[a] *(alphabetical order)*

[a] Wissenschaftsladen Wien - Science Shop Vienna

Period addressed by the study

The list of projects has been extracted from the CS Track database, including projects that have been extracted from the corresponding CS Platforms from 2019 to 2022.

Main aim of the study

Pursuing the question if more and hidden information on citizen science activities within projects from the database can be found via extensive online research or interviewing organisers.

Research question/s

- Which information on citizen science activities is online available that matches the Activity & Dimension Grid of Citizen Science or goes beyond it?
- Is there any contradictory information?
- What can be the reason for the availability or non-availability of information about citizen science activities?
- How does/could this impact on the CS Track's recommendations?

Research Context

It was established by CS Track that public information on citizen science activities on project websites or databases often are poor and/or incomplete (Strähle, Urban et al., 2021; Calvera-Isabal et al., 2023). In a survey among organisers of citizen science signs were found that organisers might not always know much about those who participate. Apart from this, there may be several good reasons to withhold certain information. It is also possible that organisers do not think about mentioning certain aspects of the activity. For doing research on citizen science and to give recommendations it would be crucial to have extensive information on activities or at least know which information is not deliberately given for certain reasons.

Research Method(s) applied

In-depth research and analysis of citizen science activities selected from the CS Track database using the Activity & Dimensions Grid of Citizen Science developed in WP1. Contacting organisers to clarify if they consider their activities as citizen science and if there is more information available on the presented activities.

Procedure(s) applied

The projects were randomly selected from a list of 3318 projects taken from the CS Track database. Information about these projects was available in English or German for all these database entries. The list was provided by Miriam Calvera (UPF). The sample was selected by using a simple formula for random ranking in combination with exclusion criteria.

Summary of results/findings

Intense research on 6 projects showed that there was in none of the cases enough and sometimes only a fraction of the information available that can be important to answer the question, which benefit and caveats, barriers and enablers, incentives and disincentives citizen science activities may have (see Activity & Dimensions Grid of Citizen Science).

Conclusion

Giving information is a question of transparency, although there can be sensitive issues that need to be kept secret. Standards for transparent public information would be required not only for research purposes but also for those who are interested in participating.

In WP1 the Activities & Dimensions Grid of Citizen Science was developed which was based on literature research. The information found on a few random sampled projects is too thin or at least not very detailed, which makes systematic research on CS quite difficult. There can be solid reasons in research to withhold some information. Nevertheless, standards for transparency should be discussed: what information should be made public by organisers of CS in all cases (e.g., funding, conditions for participation, requirements, etc.) and which information could be kept secret if reasons are made public (e. g., data protection, location of sensitive environments, etc.)

Link to dataset (if applies):

https://zenodo.org/record/7376970#.Y4X87XaZOUk

Link to complete report

The paper/full report concerning this study is currently in progress (on the date of submission of this deliverable - November 2022), no link to a repository is available. Contact the corresponding author (wilawien@wissenschaftsladen.at) if you have interest to receive further information.

References

Calvera-Isabal, M., Santos, P., Hoppe, H., & Schulten, C. (2023, forthcoming). How to automate the extraction and analysis of information for educational purposes. [Cómo automatizar la extracción y análisis de información sobre ciencia ciudadana con propósitos educativos]. Comunicar, 74. https://doi.org/10.3916/C74-2023-02

Strähle, M., Urban, C. et al. (2021). Framework Conceptual Model D1.1. Zenodo. https://doi.org/10.5281/zenodo.5589618.

Strähle, M., Urban, C. et al. (2022). Conceptual Framework for Analytics Tools D1.2.

Strähle, M., & Urban, C. (2022, forthcoming). The Activities & Dimensions Grid of Citizen Science. In: Proceedings of Science. PoS(CitSci2022)087. https://pos.sissa.it/418/087/.

3.4 Is it a match? Motivations on citizen science volunteers and recruitment arguments in project descriptions

Authors and Research affiliation

Kai Nils Weeber[a], Nicolas Felipe Gutierrez Paez[b], Ohto Sabel[a], Raija Hämäläinen[a]

[a] University of Jyväskylä

[b] TIDE Research Group, Universitat Pompeu Fabra

Period addressed by the study

Although the study started at the beginning of 2022 the data needed for the analysis was collected in different periods of time:

- The CS projects' descriptions from Zooniverse were extracted in November 2021. A total of 367 projects were available on the platform at that time.
- The questionnaire data was collected over a period of seven months (January-July 2021).

Main aim of the study

This study aims at understanding how the alignment between the motivational factors of CS participants and the recruitment speech used in the projects' description is, by performing quantitative triangulation of data collected through a survey about 12 motivational factors for participating in a CS project, and the manual analysis of the projects' descriptions available in Zooniverse website.

Research question/s

How well motivational arguments in project recruitment match the motivational structure of citizen science participants?

Research Context

Since citizen science projects rely on voluntary participants, it is relevant what motivates people to engage in those projects. Most research literature before takes only participation in one specific citizen science project or a specific science field into account. Therefore, we analyse how important motivational factors among self-related and social-related gratifications (Nov et al., 2010) for citizen science volunteers in general are. In addition, previous studies rarely considered that citizen science participation underlies also recruitment communication managed by project organisers, despite its importance for a successful work with volunteers (Shields, 2009).

Previous literature shows a variety of 12 different motivational factors for citizen science participation like topic interest, social recognition, or contribution to scientific research, connected to different project topics or features (Lampi et al., 2020). These factors, in turn, can be attributed to more large-grained motivational categories regarding more social-oriented arguments like altruistic contribution, joining a community, or social interaction, as well as more self-oriented arguments like enjoyment or project reputation (Lee et al., 2018; Nov et al., 2011).

Research Method(s) applied

Data has been collected by quantitative triangulation. 1076 participants in citizen science projects answered a survey about the 12 motivational factors for participating. They had access to the survey by social media posts or email invitations sent to people in charge of projects. Data regarding motivational arguments in recruitment come from quantitative content analysis of 367 project descriptions of the website Zooniverse.

Procedure(s) applied

The procedure applied to perform the quantitative triangulation analysis is presented in figure 3.4.1.

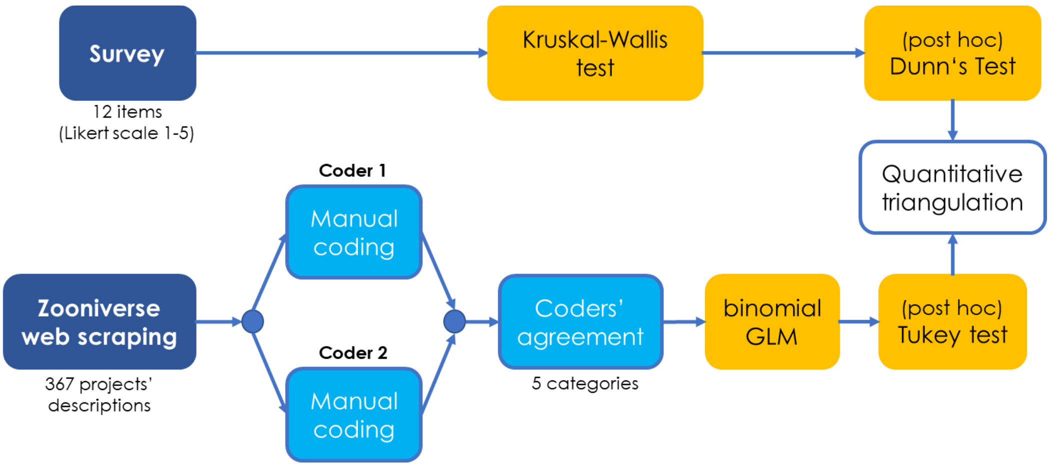

Figure 3.4.1. Procedure applied in the study.

The manual content analysis of the project descriptions was done by two coders independently. Then, both coders analysed their codings and reached consensus. The five coding categories are **desire of contributing** (The description ask for volunteers help), **joining a community** (The description invites to join to the project/community), **social interaction** (The description mentions the possibility of socially interact with other volunteers, experts, etc.), **enjoyment** (The description appeals to the enjoyment of the tasks to be performed), and **project reputation** (The description presents details on the name of entities who contribute/participate/fund the scientific research).

Data about the motivational factors for participating in CS projects was collected through 12 likert-scale items. By analysing the definition of the 12 motivational factors used in the survey, 5 were matched with the more general factors used in the descriptions' coding phase (table 3.4.1).

To triangulate data, we first filtered out the answers provided by participants in Zooniverse projects (N=17). By analysing the answers of the Zooniverse sample against the total number of responses, it was determined that they are similar, hence, the total number of survey answers (N=1074) were used for the triangulation analysis. We performed a Binomial GLM test to determine differences between the factors in the projects' descriptions, while Kruskal-Wallis test was performed to determine differences between the motivational factors of the survey respondents.

Table 3.4.1. Matching between appealing motivations and motivational aspects.

Coding category	Survey Item
Desire of contributing	Desire to help
Joining a community	Meeting new people and engaging in a community
Social interaction	Opportunities to share existing knowledge with others
Enjoyment	Fun and enjoyment
Reputation	Social reasons or recognition

Summary of results/findings

Results show that participants mainly take part in citizen science projects because of collective motives, enjoyment and a need for knowledge-gain. Analogously, enjoyment and collective ideals are also substantial arguments in citizen science project descriptions. Triangulation of both data might indicate that organisers meet volunteers' motivations in general, except for the case of social interaction / sharing opportunities, which interestingly is one of the most important motivations for CS volunteers to participate, while it is rarely mentioned in the projects' descriptions.

Conclusion

Even though our study considered different populations for the motivational aspects of the participants (it included participants of CS projects different from those hosted in Zooniverse) and the content analysis of the projects' description was limited to Zooniverse projects, our results provide a first insight of how the project organisers understand and align their projects to target engaged participants. Furthermore, the study may suggest a strategy for a wider study that includes CS projects' descriptions from other platforms and sources. It can also help to generate policy & funding guidelines for pursuing constructive motivations in CS projects.

Link to complete report:

Kai Nils, W., Gutiérrez Páez,N.F., Sabel, O. and Hämäläinen, R. (2022) "Is It a Match? Motivations on Citizen Science Volunteers and Recruitment Arguments in Project Descriptions." In Proceedings of the ECSA2022 conference: Citizen Science for Planetary Health, 69–70. https://2022.ecsa-conference.eu/files/ecsa/Bilder/ECSA2022_Conference_Proceedings.pdf

Article on eMagazine: https://cstrack.eu/format/reports/is-it-a-match-motivations-on-citizen-science-volunteers-and-recruitment-arguments/

Link to dataset:

https://doi.org/10.5281/zenodo.7310080

References

Lampi, E., Lämsä, J., & Hämäläinen, R. (2020). D4.1: Towards understanding the forms of participation and knowledge-building activities in citizen science projects.

Lee, T. K., Crowston, K., Harandi, M., Østerlund, C., & Miller, G. (2018). Appealing to different motivations in a message to recruit citizen scientists: Results of a field experiment. Journal of Science Communication, 17(1), A02. https://doi.org/10.22323/2.17010202

Nov, O., Anderson, D., & Arazy. O. (2010). Volunteer computing: A model of the factors determining contribution to community-based scientific research. Proceedings of the 19th International Conference on World Wide Web, 741 – 750. https://doi.org/10.1145/1772690.1772766

Nov, O., Arazy, O., & Anderson, D. (2011). Technology-mediated citizen science participation: A motivational model. In Proceedings of the International AAAI Conference on Web and Social Media, 5(1), 249 – 256.

Shields, P. O. (2009). Young adult volunteers: Recruitment appeals and other marketing considerations. Journal of Nonprofit & Public Sector Marketing, 21(2), 139 – 159. https://doi.org/10.1080/10495140802528658

3.5. Why keep hanging on? Semi-structured interviews with long-term CS participants in project management about motivations and perceived barriers

Authors and Research Affiliation

Paavo Räty, Kai Weeber, Ohto Sabel, Raija Hämäläinen
University of Jyväskylä

Period addressed by the study

The interviews took place in August 2022, but the discussed experiences regard a much longer timespan, back to the 1970s.

Main aim of the study

The study tried to deepen the understanding of how people get into management of CS projects. Motivations are clarified in their individual and social context. Also, the perception of inconveniences is taken into account.

Research questions

- RQ_1: Which motivators occur with the beginning of managing activities and roles in citizen science?
- RQ_2: How do motivations change for managing individuals over time in citizen science?
- RQ_3: How do managing individuals encounter barriers and inconveniences in citizen science work?

Research Context

Citizen science offers diverse options for individuals to engage in different projects as well as specific roles and activities. Thus, there is a great variety in volunteers depending on how much responsibility they take in CS projects. Also, an enduring and more active engagement in CS would sustain new projects. Looking at individuals that actually have been involved deeply for a long time might show opportunities how to promote responsible roles in CS to future volunteers.

Research Method(s) applied

The investigation is based on semi-structured, episodic interviews. The interview guide has been constructed in a way that it addresses episodic as well as semantic knowledge (Flick et al., 2000). After composing the guide, it was reviewed by other research members that had not been involved in the former process. The finalised version included three main sections: a part about the individual understanding of the term citizen science in order to better understand the familiarities with different types of projects, a part about the personal development and motivations in citizen science and, at last, a part about experienced downsides in CS engagement and personal coping strategies.

An unrelated researcher mediated with potential interview participants and considered the inclusion of heterogeneous backgrounds. In the end, 3 interviews were conducted with participants who are from different disciplines and have been engaged with different project types for varying periods of time.

Procedure(s) applied

The analysis is in progress (November 2022). Next steps include verbatim interview transcription and a hermeneutic content analysis.

Summary of results/findings

Findings have not been made yet.

Conclusion

Conclusion has not been made yet.

Link to complete report

The paper/full report concerning this study is currently in progress (on the date of submission of this deliverable - November 2022), no link to a repository is available. Once the analysis is finalized, the missing results and discussion will be available as an eMagazine article on the CS Track website. Contact the corresponding authors (kai.weeber@posteo.de and ohto.j.j.sabel@jyu.fi) if you have interest to receive further information.

References

Flick, U., Bauer, & Gaskell. (2000). Episodic Interviewing (pp. 75-92). London: Sage Publications

4. CS and its relation with the Science System

4.1 Investigating the potential of citizen science to respond to emerging challenges - The case of COVID-19

Authors and Research affiliation

Yaela Golumbic [a, b], Reuma De - Groot [a], Tslil Farchi [a] & Anne Turbe [a]

[a] The MOFET Institute
[b] The Steinhardt Museum of Natural History

Period addressed by the study

This study was conducted throughout 2021. Website content analysis was conducted between March-June, followed by interviews which took place between June-August and finally data analysis.

Main aim of the study

This study aimed to investigate the power of citizen science to respond to emerging challenges, using the case study of the COVID-19 pandemic. The two main goals were to:

- Examine how citizen science projects responded to the emerging challenges and research needs of the Covid-19 pandemic
- Investigate the scope, characteristics and development process of projects who researched COVID-19 related topics

Research question/s

- How have existing and emerging citizen science projects tackled COVID-19 as a research topic?
- What are the characteristics of projects which have done so?
- What was the level of preparedness of projects for responding to the Covid-19 research needs? and what processes were required of them?
- What are the lessons learnt from this case study, for future emerging challenges?

Research Context

The COVID-19 pandemic has challenged scientists, researchers, and industries to rapidly divert their research to better understand the COVID-19 virus spread, biology, health implications in addition to identifying medical solutions and cures. One of the avenues utilised for this cause was citizen science. Citizen science offers a huge potential to complement official responses to the Coronavirus pandemic, both in terms of facilitating scientific advances and of improving public engagement.

Research Method(s) applied

This study utilised a two-phase research approach for analysing COVID-19 related citizen science projects. Website content analysis was used to understand the outstanding characteristics of citizen science projects that endeavoured to respond to the pandemic, in terms of their geographical distribution, aims, design and characteristics of citizen engagement. Followed by interviews and detailed case studies of seven citizen science initiatives which provided in-depth understanding of the development of projects, and practices used across a range of approaches.

Procedure(s) applied

The identification and selection of projects to be included in the website content analysis was conducted the searching CS-track project database (n=13) and lists produced by citizen science associations and research institutes globally (e.g. CSA - https://www.citizenscience.org/covid-19) (n=22, with some overlaps). Following this initial compilation, projects that were not directly focused on Covid-19 (n=3 projects) were removed to a total of 25 projects for final analysis (see figure 4.1.1).

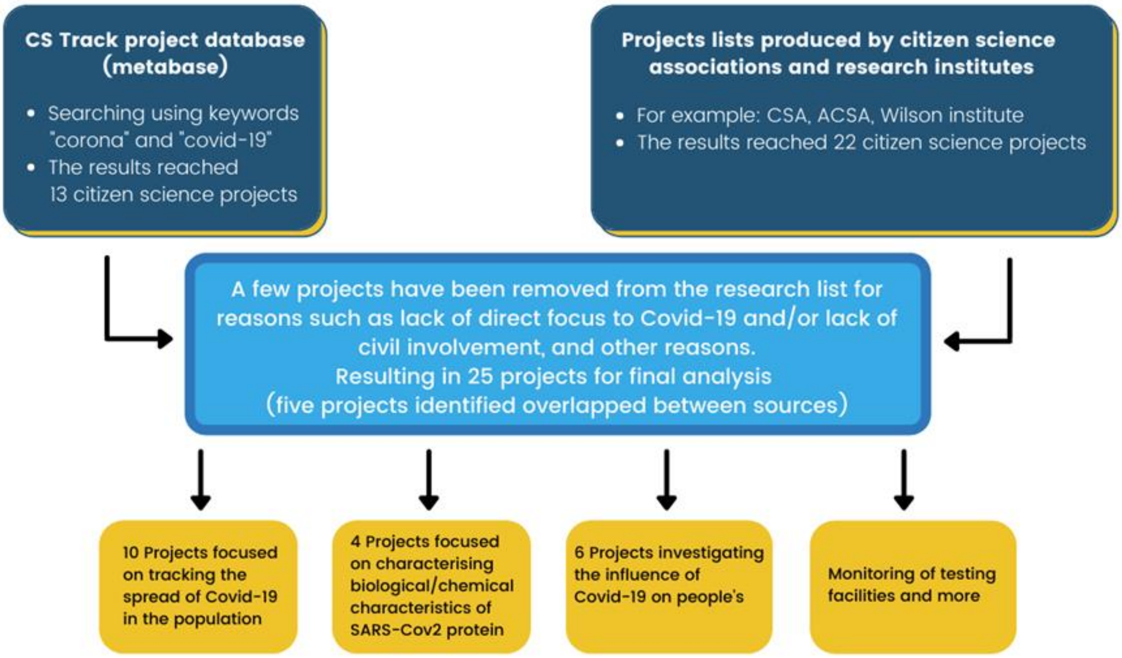

Figure 4.1.1. Selection and characterization of citizen science projects related to COVID-19 research

Website content analysis extracted characteristics related to Project focus, scope and design from project descriptions and websites. Attributes were categorized into thematic groups and validated by two independent coders.

Project analyses were contacted for further research with n=8 replying positively for conducting full case studies. These were done using semi-structured interviews, investigating 13 project attributes.

Summary of results/findings

Content analysis of projects' websites revealed projects focused on three main domains, namely tracking the spread of the pandemic in the population, investigating the influence of COVID-19 on people's wellbeing, and investigating the COVID-19 virus biology. Citizen scientists' tasks centred around responding to an online survey, self-tracking data from a wearable device and distributed computing. Overall projects were widely accessible, targeting a broad audience and requiring no special skills. Most projects required at least a moderate degree of effort from participants, asking a few types of questions, and many required frequent contributions at regular intervals.

The case studies revealed the importance of early preparedness to respond to new challenges, building on existing experience, collaborations and modular software infrastructure. Important features highlighted by projects included regular and honest feedback securing the trust and engagement of the participants, and thinking big, open and collaboratively when designing a project.

Conclusion

This study provides insight as to the role citizen science projects have had in conducting COVID-19 research. While the long term contributions are not yet evident, this study showcases the response of the citizen science community to emerging challenges and the project characteristics which contributed to the success of these projects. This research has important implications for the design and management of citizen science projects, planning for a sustainable future and promoting planetary health in times of harmony and in times of crisis.

Link to complete report

URL: https://zenodo.org/record/6034585#.Yot079NBzb3

Article on eMagazine: https://cstrack.eu/format/reports/investigating-the-potential-of-citizen-science-to-respond-to-covid-19-challenges/

4.2 Mapping Sustainable Development Goals to Citizen Science projects

Authors and Research affiliation

Patricia Santos[a], Ishari Amarashinghe[a], Miriam Calvera-Isabal[a], Cleo Schulten[b], H.Ulrich Hoppe[b], David Roldán-Álvarez[c], Fernando Martínez-Martínez[c]

[a] TIDE Research Group, Universitat Pompeu Fabra

[b] RIAS Institute

[c] Universidad Rey Juan Carlos

Period addressed by the study

A total of 56 websites have been used to extract project descriptions. The list of websites (CS platforms and non CS platforms) and projects is consistently updated for the duration of the project (2019-2022). The data has been extracted from the CS Track database.

Main aim of the study

This work presents opportunities, achievements, and future challenges in using computational analytics to better understand the connection between CS and the SDGs. The work in its status does not fully cover SDGs in CS, but it evaluates and shows the potential of the text-classification techniques for identifying SDGs in CS project descriptions and for assessing trends in connection of CS and SDGs based on available data.

This study analyses different automatic classifiers by comparing the results obtained from their application in a sample of 208 CS project descriptions. The main aim is to present the benefits and limitations of these techniques (nCoder, ESA, OSDG and BERT), but also provides a discussion of the potential benefits of using data from CS projects to map the 17 SDGs. Second, this work has been extended by analysing all the project descriptions in English collected in the CS track database. The corresponding results of this analysis are presented in section 7.3 in this report.

Research question/s

Our main research question is: *How can a data analytics approach based on web-based data mining and automatic classifiers contribute to the reporting of SDGs related to CS activities and projects?*

Research Context

Previous studies have discussed how traditional data sources provide insufficient knowledge for measuring the United Nations Sustainable Development Goals (SDGs). Data related to SDGs are sourced primarily from global databases maintained by international organisations, national statistical offices and other government agencies. Recent studies show the value of using data from Citizen Science (CS) for assessing the SDGs. The online presence of CS, especially via online CS platforms provides a rich context of data. In this scenario, the role of computational data science is key. This work explores and exemplifies opportunities for combining web-data mining techniques and automatic classifiers to enhance the understanding of the interrelation between CS and the SDGs.

Research Method(s) applied

A descriptive research approach. Combining qualitative research coding (manual content analysis) with automatic classification based on the application of three different methods:. nCoder, ESA and OSDG.

Procedure(s) applied

A subset of 208 projects from 16 CS different platforms were randomly selected from the CS Track DB with the following criteria: project descriptions should be in English; platforms should contain a list of projects situated in Europe or should be projects conducted online.

The method proposed by Fraisl et al. (2020) was followed to extract and review SDGs targets and indicators metadata. The review process was done by 3 researchers to identify a list of keywords to be applied for SDG classification purposes. In this process the list published by Monash University and Australia S.D.S.N. (Kestin et al. 2017) was used. Then, manual coding was conducted by 2 researchers and 2 research assistants (n=4) of the 208 project descriptions. New keywords emerged from the manual coding process. The initial manual coding provided a ground truth against the performance of the three methods (i.e nCoder, ESA and OSDF) to evaluate the application of the selected automatic classifiers. In the corresponding full report of this study we explain the reasons for selecting these automatic classifiers. Figure 4.2.1, illustrates the process followed to classify the dataset.

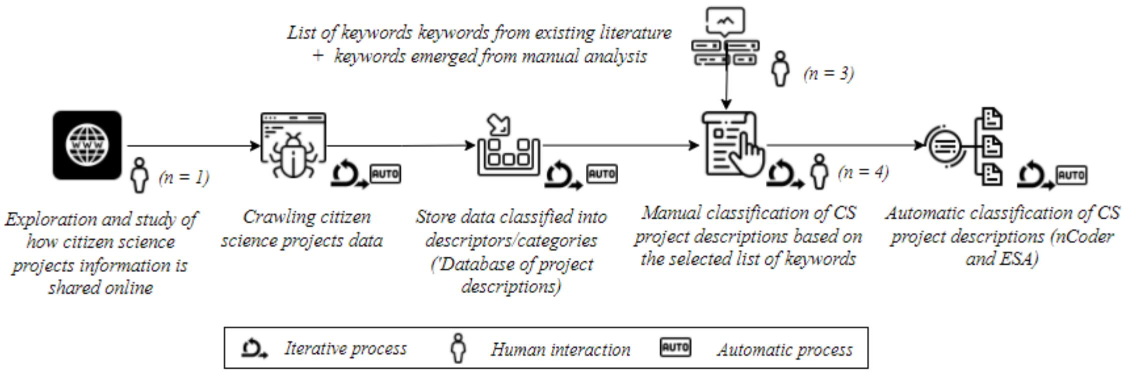

Figure 4.2.1. Dataset preparation and manual coding process

Summary of results/findings

As a main contribution, this study shows how automatic classifiers can be used to map CS data with SDGs. Additionally, we provide a discussion of the techniques covered in this study by considering their advantages and limitations when applying each technique to classify CS project descriptions with SDGs.

We observe coincidences with results from previous authors regarding which SDGs are more representative in CS. Most represented ones are: SDG#4 (Quality Education), SDG#11 (Sustainable Cities and Communities), SDG#13 (Climate Action) and SDG#15 (Life on Land). The case of SDG#10 (Reduced inequalities) is a curious case to be further investigated in the future. Similarly, in the case of SDG#4, SDG#10 seems to be a transversal SDG that can be associated with multiple disciplines.

An interesting finding shows how SDGs have dependencies among them, this is shown through a comparative analysis of SDG associations to the selected sample of projects. Most common associations are: SDG#5 with SDG#8; SDG#6 and SDG#14; SDG#4 and SDG#10; SDG#3 and SDG#10.

Conclusion

In summary, when comparing the three main techniques used in this study: (1) the results obtained using nCoder are more aligned with the results of the manual classification, as the overall process followed to train classifiers is closer to the manual classification. However, for large datasets manual coding for a number of different codes (as is the case for SDGs which is 17 different codes) is difficult and time-consuming and later using a trained classifier is not possible. (2) In the case of ESA, the main advantage is that it does not require manual coding and facilitates a fully automatic classification, hence the effort required from human coders is minimal. However, the requirement of a pre-existing source for comparison (e.g., Wikipedia articles), the quality and the original language of such external sources could create limitations. (3) When comparing the F1-Scores obtained with each technique, the ones from OSDG are lower than the ones obtained from the other techniques.
Additionally, and only in terms of discussion, we had into account the use of deep learning models such as BERT mainly because this is becoming the state-of-the-art model solution for multiple natural language processing tasks. Although obtaining satisfactory amounts of training data to train machine learning models is a challenge,

techniques such as BERT can provide more accurate results (also considering multiple languages) in the future.
Taking into consideration the lessons learned from this work, in section 7.3 of this deliverable we present an extended version of this study where ESA has been used to analyse 4849 projects with English descriptions contained in the CS Track DB.

Link to complete report:

The paper/full report concerning this study is currently under review (on the date of submission of this deliverable - November 2022), no link to a repository is available. Contact the corresponding author (patricia.santos@upf.edu) if you have interest to receive further information.

The link on Zenodo includes a presentation as part of the CS Track ECSA event (8th October 2022, Berlin): CS-Track database: a central database of CS projects in Europe that can be key to understand the connection of CS and SDGs "Understanding the nature of Citizen Science in a rapidly changing world". This work has been presented as part of the eMagazine publications: https://cstrack.eu/test/beta-report/mapping-sustainable-development-goals-to-citizen-science-projects/

Link to dataset:

https://zenodo.org/record/7310477#.Y2zornaZNPY

References

Kestin, T., van den Belt, M., Denby, L., Ross, K., Thwaites, J., & Hawkes, M. (2017). Getting started with the SDGs in universities: a guide for universities, higher education institutions, and the academic sector.

Fraisl, D., Campbell, J., See, L., Wehn, U., Wardlaw, J., Gold, M., ... & Fritz, S. (2020). Mapping citizen science contributions to the UN sustainable development goals. Sustainability Science, 15(6), 1735-1751.

4.3 Tasks in Citizen Science: proposing a hierarchical framework for categorising citizen scientists' activities in CS projects

Authors and Research Affiliation

Marinos Anastasakis[a] and Kathy Kikis-Papadakis[a]

[a] Foundation for Research and Technology - Hellas (FORTH)

Period addressed by the study

The study concerns CS activity that has been captured by CS Track's database. This includes CS projects which use English in their website and as their primary language for online communications with citizens.

Main aim of the study

The study aims at proposing in a clear and systematic manner a way of categorising and classifying the tasks that citizen scientists are engaged with in CS projects. Our efforts are not only focused on proposing a scheme that CS projects can utilise, but also suggesting a set of descriptors for the CS Track database.

Research Questions

- RQ1: In what kind of tasks are citizen scientists engaged?
- RQ2: Where in the research cycle can these tasks be positioned?
- RQ3: In what ways do current schemes categorise these tasks?

Research Context

Building upon the main conclusions drawn in D1.1 for the categorisation of CS in CS Track (activities-dimension grid for CS classification), we focus on CS activities pertaining to Area 1 (Taking part in research projects) and Area 4 (School projects with minors). In such projects, the literature has failed to provide a concise account regarding citizen scientists' active contributions, with current typologies offering a divergent and incomplete image of the landscape.

Research Method(s) applied

The CS Track's database is currently used as the main source of data. We compiled a set of projects published in English, and we chose those having information related to the tasks that citizen scientists undertake while participating in the project (included in the descriptors "Methodology" and "Activity type"). In total, 2,053 projects were selected.

Procedure(s) applied

At a theoretical level, our approach is guided by second generation Activity Theory (Leontiev, 1981), a framework which views human activities as processes having a hierarchical structure and categorisation of artefacts Engestrøm's (1990). At an analytical level, Qualitative Content Analysis (Schreier, 2014) will be used for classifying the tasks citizen scientists undertake while engaged in a project. The main codes used for our analysis are derived from the activities-dimension grid's area (Area 2: Participation in Research; Area 4: School), forms of data collection (observation, reporting, taking samples, measuring and counting, searching for artefacts, conducting interviews, and supporting data collection) and forms of data preparation and processing (classifying, characterising, describing, localising, matching, transcribing).

Summary of results/findings

This study is still in progress (November 2022). Our preliminary findings indicate that citizen scientists are predominantly engaged in tasks related to data collection, preparation and processing. But more results/findings are expected after the manual coding of 2053 projects.

Link to complete report

The paper/full report concerning this study is currently in progress (on the date of submission of this deliverable - November 2022), no link to a repository is available. Contact the corresponding authors (m.anastasakis@uoc.gr and katerina@iacm.forth.gr) if you have interest to receive further information.

References

Engeström, Y. (1990). When is a tool? Multiple meanings of artefacts in activity. In Y. Engeström (Ed.), Learning, Working and Imagining: Twelve studies in Activity Theory (Helsinki, pp. 171–195).

Leontiev, N. (1981). Problems of the development of the mind. Progress Publishers.

Schreier, M. (2014). Qualitative Content Analysis. In U. Flick (Ed.), The SAGE Handbook of Qualitative Data Analysis (pp. 170-183). SAGE.

5. CS and Education

5.1 Identifying learning dimensions in CS project descriptions

Authors and Research affiliation

Marius Oesterheld,[a] Vincent Schmid-Loertzer,[a] Miriam Calvera-Isabal,[b] Ishari Amarasinghe,[b] Patricia Santos,[b] & Yaela N Golumbic[c,d]
[a] Wissenschaft im Dialog
[b] TIDE Research Group, Universitat Pompeu Fabra
[c] The MOFET Institute
[d] The Steinhardt Museum of Natural History

Period addressed by the study

The data source used for this study was extracted from the CS database, which contains information about CS projects extracted from a total of 59 online websites. CS project information, the list of websites and projects is consistently updated for the duration of the project (2019-2022).

Main aim of the study

Whereas most existing studies investigate perceived or observed learning gains of citizen scientists (Land-Zandstra et al., 2016; Aivelo & Huovelin 2020), this study took an alternative perspective by examining learning-related aspects in textual self-representations of CS projects—namely in project descriptions posted online.

Research question/s

The research question we wanted to answer through this study was 'which dimensions of learning are the most prominent in CS project descriptions?'

Research Context

While the main objective of citizen science (CS) projects is generally to answer a scientific question, they might also have many important educational benefits for participants. In fact, previous studies show that the wish to acquire new skills and knowledge is one of the main reasons for people to join CS projects (Jennett et al., 2016). Since learning opportunities are an important motivational factor for participation in CS, and project descriptions posted on websites play a key role in attracting volunteers, we decided to examine what these texts can tell us about the educational potential of CS projects.

Research Method(s) applied

For this study, we randomly selected 94 CS projects from the whole CS Track database (4949 CS projects in total) in April 2021, applying the following criteria:

- CS projects with the description in English
- Balanced dataset between types of platforms: structured, semi-structured, non structured and manual extraction (cf. section 6.1)

Procedure(s) applied

As a theoretical and methodological starting point for our study, we chose the widely used and referenced "Framework for articulating and measuring individual learning outcomes from participation in citizen science" developed in 2018 by Tina Phillips and colleagues. Building on their framework, which proposes six main categories of learning outcomes in CS projects (Interest; Self-Efficacy; Motivation; Content, Process and Nature of Science Knowledge; Skills of Science Inquiry; and Behaviour and Stewardship), we conducted a structuring qualitative content analysis (see Figure 5.1.1) as described by Philips Mayring. In order to accommodate the material that did not fit into the model, we decided to include one additional learning dimension—Attitude Change—and two aspects related to the deliberate design of learning opportunities for participants—"Training and Didactic Materials provided by the project" and "Access to Project Results". At the same time, we chose to exclude the "Motivation" category due to a lack of relevant text in our sample.

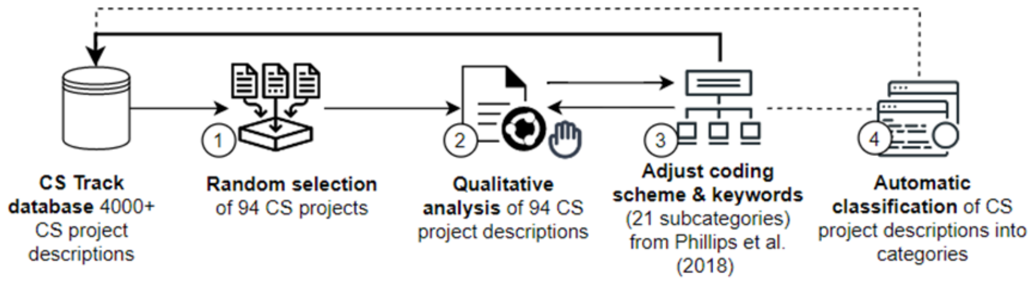

Figure 5.1.1. Process followed

After manually assigning phrases, sentences and short paragraphs to the resulting eight main categories (six learning dimensions and two categories relating to deliberate design of learning opportunities) and 21 subcategories, distinctive and frequently occurring keywords were extracted from these text snippets. The entire sample was coded independently by two members of the research team. The rate of agreement was found to be over 90% for six of the eight main categories (i.e. Skills of Science Inquiry, Self-Efficacy, Interest, Attitude Change, Training and Didactic Materials, Access to Project Results)—and between 70 and 80% for the remaining two categories (i.e. Content, Process, and Nature of Science Knowledge and Behaviour and Stewardship).

Summary of results/findings

Our results (see Figure 5.1.2) indicate that most project descriptions focus strongly on science-related learning dimensions while disregarding other personal or interpersonal benefits (such as self-efficacy, attitude or behavioural changes etc). References to Content, Process, and Nature of Science Knowledge play a minor role compared to statements relating to Skills of Science Inquiry, which clearly dominate most project descriptions.

That being said, content knowledge features more prominently than the other two knowledge types. Within the category of scientific skills, data collection and submission and using technology far outstrip all other subcategories, with data analysis a distant third. This implies that, at least judging from their self-descriptions, around 88% of the projects represented in our sample seem to be contributory, rather than collaborative or co-created. Of 42 project descriptions which contain information on training and didactic materials offered to participants, only 6 (14.3%) mention interactive training formats.

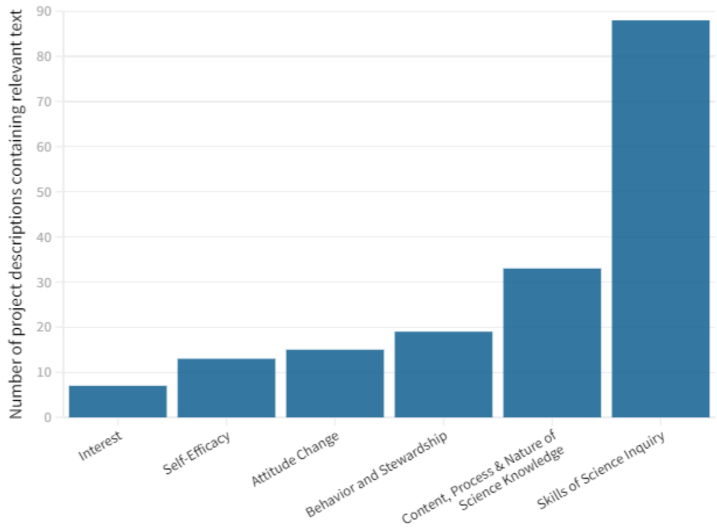

Figure 5.1.2 Distribution of learning categories across projects

(main categories only)

The keywords we extracted have since been used for other studies conducted within the CS Track consortium - e.g. to automatically classify and identify mentions skills of science inquiry in all the project descriptions contained in the CS Track database (cf. section 5.3 of this deliverable for further detail).

Conclusion

Our study revealed a very uneven representation of learning dimensions within CS project descriptions. This result suggests that project initiators and coordinators either do not devote enough attention and resources to creating the broadest possible range of learning opportunities for their volunteers, or do not communicate the educational potential of their project clearly enough in their project descriptions.

The strong focus on science-related learning that we observed seems to be a common bias in CS, as several publications have pointed out - a bias that may run counter to volunteers' actual motivations and expectations (Carson et al., 2021; Roche et al., 2022). Working with a broader definition of learning - in project descriptions, evaluation practices, and in the deliberate design of learning opportunities for volunteers - could help extend the range of benefits that CS can have on both the individual and the societal level.

Our analysis also revealed that the quality of project descriptions varies considerably and that many do not contain any information on how volunteers will benefit from participating. In light of this observation, we decided to create a set of evidence-based recommendations on how to write effective and engaging CS project descriptions (cf. section 8).

Link to complete report:

The resulting publication from this study is: M. Oesterheld, V. Schmid-Loertzer, M. Calvera-Isabal, I. Amarasinghe, P. Santos, & Y. Golumbic (2022). Identifying learning dimensions in citizen science projects. In proceedings of Engaging Citizen Science Conference 2022, PoS(CitSci2022) 070. https://pos.sissa.it/418/070/ [forthcoming]

Link to dataset:

https://zenodo.org/record/7374000#.Y4UGjnaZNPY

References

Aivelo, T., & Huovelin, S. (2020). Combining formal education and citizen science: A case study on students' perceptions of learning and interest in an urban rat project. Environmental education research, 26(3), 324-340.

Jennett, C., Kloetzer, L., Schneider, D., Iacovides, I., Cox, A., Gold, M., ... & Talsi, Y. (2016). Motivations, learning and creativity in online citizen science. Journal of Science Communication, 15(3).

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

Phillips, T., Porticella, N., Constas, M., & Bonney, R. (2018). A framework for articulating and measuring individual learning outcomes from participation in citizen science. Citizen Science: Theory and Practice, 3(2).

5.2 Learning in citizen science: a triangulation approach

Authors and Research affiliation

Miriam Calvera-Isabal [a], Marius Oesterheld [b], Fernando Martínez-Martínez [c], Aaron J. Peltoniemi [d], Patricia Santos [a], and Yaela N Golumbic [e, f]

[a]TIDE Research Group, Universitat Pompeu Fabra
[b]Wissenschaft im Dialog
[c]Universidad Rey Juan Carlos
[d] Jyväskylän yliopisto
[e]The MOFET Institute
[f]The Steinhardt Museum of Natural History

Period addressed by the study

The data needed for the analysis was collected in different periods of time:

- The information of CS projects collected online is consistently updated for the duration of the project (2019-2022). The data collected was retrieved from the database in April 2022.
- The twitter data was stored daily for two years (September 2020 - August 2022) but the one used for the analysis was retrieved from the database in August 2022.
- The questionnaire data was collected over a period of seven months (January-July 2021).

Main aim of the study

This multi-perspective study aims to investigate the alignment between (1) learning opportunities mentioned by citizen science (CS) project initiators in CS project descriptions available online, (2) learning opportunities shared in tweets from project or platform accounts and, (3) the participants' perceived learning experiences as reflected in survey responses and tweets from individual user accounts.

Research question/s

The question we wanted to answer is: *What are the main overlaps and discrepancies between learning opportunities communicated by project coordinators and citizen scientists' perceptions of learning in CS projects*?

Research Context

Participation in CS has been analysed from many different angles, ranging from the types of activities a volunteer can engage in to potential educational impacts. Development of scientific skills, the use of technology, content knowledge or science literacy are some of these educational impacts identified and discussed by the community. However, it seems reasonable to assume that the educational potential of CS projects can only be fully realised if the learning opportunities offered correspond closely with the participants' needs, interests, and expectations. Proceeding from this hypothesis, the present study uses a triangulation-based approach to conduct a multi-perspective analysis of learning in CS projects. The triangulation of different types of data allows us to examine both the way CS project coordinators communicate learning-related aspects of their projects and the way volunteers describe learning opportunities and experiences - and thus enables us to compare the project coordinators' and citizen scientists' perspectives on learning in CS projects. The ultimate aim of this comparison is to identify ways in which learning opportunities offered in CS projects can be brought into closer alignment with the participants' interests and expectations.

Research Method(s) applied

The study combines automatic methods, such as web scraping or social network analysis to extract data from online sites and social media, with manual data analysis. Educational effects envisioned by project coordinators were examined through Twitter data (from project and platform accounts) and a qualitative content analysis of CS project descriptions, while learning experiences of citizen scientists were studied using Twitter data (from individual user accounts) and an online participant survey.

Procedure(s) applied

As a first step, this study triangulated three datasets created in the research work of CS Track project descriptions from CS Track database (N=94), tweets (N = 216,786)

and survey responses (N = 610) – using a unique combination of web-based and computational analytics with traditional social science methods. We initially analysed each dataset independently before comparing the findings to identify ways in which the project coordinators' and citizen scientists' perspectives on learning in CS projects overlap and diverge. The first dataset - a qualitative content analysis of 94 project descriptions stored in the CS Track database – was created in the context of a previous study, which used the "framework for articulating and measuring individual learning outcomes from participation in citizen science" developed in 2018 by Tina Phillips et al. as a theoretical foundation (cf. section 5.1). The keywords derived from this manual coding of project descriptions were used to conduct an automated analysis of tweets, and thus form the basis of the second dataset used in this triangulation study. The third dataset consists of citizen scientists' responses to the CS Track online survey, which focused primarily on Europe and was distributed for a period of seven months (January-July 2021) through multiple channels.
In order to enable a comparison between project descriptions and tweets on the one hand, and survey results on the other, we first had to match our coding scheme to the survey questions (or response options) by identifying semantic overlaps or parallels. Since the two underlying studies - online survey and qualitative content analysis of project descriptions - were designed and conducted independently of each other, conceptual differences are inevitable and cannot be eliminated retroactively. What connects the items juxtaposed in the following tables is that they pertain to the same field of learning (use of technology, communication, data analysis).

As a second step, we conducted a case study of 11 projects, which allowed us to narrow down the three datasets (project descriptions: N=11, tweets: N=118, survey responses: N=139) and draw conclusions on the level of individual projects. The 11 CS projects were selected by applying the following criteria: (1) There must be more than five survey respondents who reported to have participated in the project as a citizen scientist (and who have completed the entire survey). (2) The projects in question must not be platforms which serve as data repositories or data submission interfaces for various different CS initiatives. (3) There must be a project description available online that actually mentions the project's CS activities.

Summary of results/findings

The results of both the general comparison and the project-level case study reveal that there is a significant discrepancy between the learning opportunities described by project coordinators and the learning experiences reported by project participants. This gap is particularly evident with regard to skills related to communication and project or research design, but also when it comes to scientific literacy and critical thinking. What our findings also show is that responses vary considerably even among volunteers who participated in the same CS project, which suggests that the citizen scientists' individual backgrounds, interests and motivations play an important role in shaping their learning experiences.

Conclusion

Our findings show that, in some respects, the way citizen scientists perceive and experience learning does not match what project coordinators communicate regarding learning opportunities in their projects. Closing (or at least narrowing) the gap between these two perspectives could potentially help increase the educational impact of CS projects. Exploring this possibility would be a worthwhile task for future research.

By drawing on the results of previous research conducted by the CS Track consortium, this triangulation study has generated additional insights - insights which have important implications for CS policy and highlight the need for further investigation into the differences between project coordinators' and citizen scientists' perspectives on learning in CS projects.

Link to complete report:

Since, at the time of submission of this deliverable - November 2022, the paper/full report concerning this study is still in progress, no link to a repository is available. Please contact the corresponding author (miriam.calvera@upf.edu) if you would like to receive further information about this research.

Link to dataset:

https://zenodo.org/record/7371616#.Y4SVC3aZOUk

References:

Phillips, T., Porticella, N., Constas, M., & Bonney, R. (2018). A framework for articulating and measuring individual learning outcomes from participation in citizen science. Citizen Science: Theory and Practice, 3(2).

5.3 Educational uses of CS data

Authors and Research affiliation

Miriam Calvera-Isabal [a] Patricia Santos [a] Davinia Hernández Leo [a]
[a] TIDE Research Group, Universitat Pompeu Fabra

Period addressed by the study

Total duration of the CS Track project (2019-2022) (more detail in section 6.1). During this time, interrelated studies of different durations have been carried out in parallel in order to answer our research questions. See (Calvera-Isaba, Santos & Hernández-Leo 2021) for more information.

Main aim of the study

The study aim is to investigate how a combination of methods (such as data analysis, computational or quantitative methods) could be applied to gather CS projects information to support teacher's practice and inspire them.

Research question/s

This study was divided into 3 case studies:

a) Proof of concept and a first analysis of the data about CS projects available on online websites (Calvera-Isabal, Varas & Santos P, 2021)

b) The analysis of how CS is communicated online and how a automatic methods can be applied to extract and store data that might be used with educational purposes (Calvera-Isabal et al., 2023; more detailed in section 6.1)
c) A co-design process of a tool that show information about CS gathered and to support teachers' practices in formal education contexts

The following research questions (Calvera-Isaba, Santos & Hernández-Leo, 2021) are the main ones derived from the studies above:

- *"How can web scraping and data mining methods be used to collect/analyse data online about citizen science projects?"*, addressed in studies a) and b).
- *"How data from CS projects can be presented/analysed in relation to their potential to support learning outcomes in formal settings?"* addressed in studies a) and b).
- *"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?"* addressed in study c).

Research Context

Today's societal challenges require citizens' awareness about societal and environmental problems, which means the development of scientific literacy and critical thinking (Siarova, Sternadel & Szőnyi, 2019). Formal education settings (along with non-formal or informal) are needed to improve student's interest and learning about sciences through the development of activities and usage of methodologies such as inquiry based learning or active-learning methods (Bryan et al., 2011; Swarat, Ortony & Revelle, 2012).

CS, which involves citizens in the scientific process, is a clear example of how activities related to science might improve science understanding, motivation to Science, Technology, Engineering and Mathematics (STEM) careers, awareness for instance to ecology, knowledge or the development of skills (Vohland et al., 2021; Strasser et al., 2019; Hiller & Kitsantas, 2014; Kobori et al., 2016; Phillips et al., 2018). Nevertheless, although learning by participating in CS might be considered informal learning (due to usually it is unintended, indirect, unguided, and not reflected (Bela et al., 2016)), in some cases it is also introduced in formal environments (such as schools, universities, etc…) through participation in a CS project activity or using materials or tools developed by CS projects (such as guides, protocols, videos or apps) (Nistor et al., 2019). Some activities require the usage of tools, reading about the scientific process, understanding scientific concepts, analysing data or collecting data. Some studies have analysed how writing and reading about science and the usage of technical terms and specific strategies indicates literacy development (Glynn, Shawn & Muth, 1994; Baram-Tsabari, Baram-Tsabari & Yarden, 2005; Ristanto et al., 2018; Suggate et al., 2018; Hong & Diamond, 2012).

Teachers train students through learning activities transforming their subject matter knowledge (SMK) alongside their pedagogical content knowledge (PCK) into concrete pedagogical actions adapting them to an educational context (Park & Oliver, 2008). When we talk about teachers' practice, we could see that they are influenced by their previous expertise, their student's interest and the environment (which includes other teachers and educational materials) (Bennett, Agostinho & Lockyer, 2015). Considering this, research explores how the information and resources about CS projects available online (see section 6.1) could inspire them to develop learning activities and support their practice. But, furthermore, we set teachers in the centre of the research, so we designed a user-centred approach (Barab & Squire 2004).

Research Method(s) applied

The full study follows a Design-Based Research (DBR) methodology (Hoadley & Campos, 2022) in combination with other methods to, iteratively, understand the connections between CS and education and how data available online could be used in formal education contexts.

Procedure(s) applied

For each individual study, apart from DBR methodology, we followed different procedures and methodologies.

For the proof of concept (study a) and the analysis of online CS communication (study b) we did an exploratory analysis of the websites that have information about CS projects (3 for the study a and 72 for study b) in order to better understand the websites' structures. Furthermore, we used computational methods so we could develop a crawler that navigates through the sites to select and extract the data. Once the data was stored in the CS Track database, algorithms such as Name Entity recognition (NER) or Natural language processing (NLP) were used to give meaning to the non-categorized data, to create new categories or to anonymize personal data.

From the study c), we combined DBR with a User Centred Design approach (UCD). Through 7 workshops, a total of 135 participants (primary and secondary) (N = 49) or pre-service teachers (N = 49) and TEL designers (N = 37)) participated in data selection and design of a tool that will show the data extracted about CS projects from online sites. We collected their needs and opinions via a questionnaire (N = 98 responses), a card sorting process (N = 14) and paper prototypes (N = 34).

Summary of results/findings

So far, the main results obtained from all the studies are aligned. Our results show that by having the data extracted from different sources, teachers can benefit from: (1) having information about scientific projects organised into categories, (2) variety of information and vocabulary related to science, (3) educational/scientific resources that could be used in or to prepare class activities (4) developed technology that that can be used in scientific inquiry activities (Calvera-Isabal, Vara & Santos, 2021; Calvera-Isabal et al., 2023). Those results are aligned with the study c (Calvera-Isabal, Santos, Hernández-Leo, Under Revision), from which we can conclude that teachers used to use new technologies such as "internet search", "blogs or forums" and "Open educational resources (OER)" to get inspiration.

From the co-design process we could identify functionalities needed for the tool to allow teachers to explore the data about CS projects. Furthermore, we identified what type of data they wanted to see in the tool. They suggested that having information about the title of the CS project, a brief description, tasks or how to participate in, learning outcomes promoted or research areas that apply to the project.

Conclusion

Once known that automatic computational methods can be used to centralise all the data available online, it has been used to understand the CS field better. Regarding the online communication of CS, there is still work to be done in connection to education. Other studies conducted by CS Track researchers explored how information about learning and education is shared online (see section 5.1 and 5.2). For instance, regarding the information needed by teachers to inspire themselves, we also requested participants to design a learning activity during the workshops to see what kind of information they are interested about. Initial results from the data analysed were shared during the workshop conducted on June 14th, 2022[2] and will be part of a future publication (Calvera-Isabal, Santos & Hernández-Leo, 2021).

Link to complete report:

Calvera-Isabal M, Santos P, Hernández-Leo D. Citizen science, data science and education: how to support teacher's inspiration during the learning activities design with technology enhance learning. Paper presented at: Doctoral Consortium of the Sixteenth European Conference on Technology Enhanced Learning (EC-TEL); 2021 Sep 20–24, Bolzano, Italy.

Calvera-Isabal M, Santos P, Hernández-Leo D. Towards citizen science-inspired learning activities: the co-design of an exploration tool for teachers following a Human-Centred design approach. [Under revision - November 2022, for further information contact the corresponding author: miriam.calvera@upf.edu]

Calvera-Isabal, M., Santos, P., Hoppe, H., & Schulten, C. (2023). How to automate the extraction and analysis of information for educational purposes. [Cómo automatizar la extracción y análisis de información sobre ciencia ciudadana con propósitos educativos]. Comunicar, 74. https://doi.org/10.3916/C74-2023-02

Calvera-Isabal M, Varas N, Santos P. Computational techniques for data science applied to broaden the knowledge between citizen science and education. In: Sampson DG, Ifenthaler D, Isaías P, editors. Proceedings of the 18th International Conference on Cognition and Exploratory Learning in the Digital Age (CELDA 2021); 2021 Oct 13-15; Lisbon, Portugal. Lisbon: IADIS Press; 2021. p. 219-26. http://hdl.handle.net/10230/49216

Article on eMagazine: https://cstrack.eu/format/reports/educational-uses-of-cs-data/

Link to dataset:

Descriptors definition: https://zenodo.org/record/7310445#.Y2zmbXaZNPY

List of websites: https://zenodo.org/record/7310295#.Y2zmwXaZNPY

Questionnaires: https://zenodo.org/record/6655987#.Y2z-yHaZNPY

Cards designed for card sorting and paper prototyping activity: https://zenodo.org/record/6655972#.Y2z_LnaZNPY

Prototype 1st version - CS projects dashboard: https://zenodo.org/record/6655902#.Y2z_-naZNPY

[2] https://cstrack.eu/topic/education/citizen-science-to-inspire-educators-the-importance-of-metadata-and-open-data-online-workshop-14-june/

Prototype 2nd version - CS projects dashboard: https://zenodo.org/record/6655910#.Y20AAXaZNPY

Learning activity design canva: https://zenodo.org/record/6655958#.Y20AAXaZNPY

Citizen science to inspire educators - The importance of metadata and open data, slides presentation: https://zenodo.org/record/7350688#.Y3308XaZOUk

References

Barab, S., & Squire, K. (2004). Design-based research: Putting a stake in the ground. The Journal of the Learning Sciences, 13(1), 1–14. https://doi.org/10.1207/s15327809jls1301

Baram-Tsabari, A., & Yarden, A. (2005). Text genre as a factor in the formation of scientific literacy. Journal of Research in Science Teaching: The Official Journal of the National Association for Research in Science Teaching, 42(4), 403-428.

Bela, G., Peltola, T., Young, J. C., Balázs, B., Arpin, I., Pataki, G., ... & Bonn, A. (2016). Learning and the transformative potential of citizen science. Conservation Biology, 30(5), 990-999.

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.

Bryan, Robert R., Shawn M. Glynn, and Julie M. Kittleson. "Motivation, achievement, and advanced placement intent of high school students learning science." Science education 95.6 (2011): 1049-1065.

Glynn, Shawn M., and K. Denise Muth. "Reading and writing to learn science: Achieving scientific literacy." Journal of research in science teaching 31.9 (1994): 1057-1073.

Hoadley, C., & Campos, F. C. (2022). Design-based research: What it is and why it matters to studying online learning. Educational Psychologist, 57(3), 207-220.

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.

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.

Nistor, A., Clemente-Gallardo, J., Angelopoulos, T., Chodzinska, K., Clemente Gallardo, M., Gozdzik, A., ... & Vojinovic, M. (2019). Bringing Research into the Classroom–The Citizen Science approach in schools. Scientix Observatory.

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.

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. DOI: http://doi.org/10.5334/cstp.126

Ristanto, R. H., Zubaidah, S., Amin, M., & Rohman, F. (2018). From a reader to a scientist: developing cirgi learning to empower scientific literacy and mastery of biology concept. Biosfer: Jurnal Pendidikan Biologi, 11(2), 90-100.

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.

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.

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.

Swarat, Su, Andrew Ortony, and William Revelle. "Activity matters: Understanding student interest in school science." Journal of research in science teaching 49.4 (2012): 515-537.

Vohland, K., Land-Zandstra, A., Ceccaroni, L., Lemmens, R., Perelló, J., Ponti, M., ... & Wagenknecht, K. (2021). The science of citizen science (p. 529). Springer Nature.

6. Visibility of CS

6.1 How to automate the extraction and analysis of information for educational purposes

Authors and Research affiliation

Miriam Calvera-Isabal [a], Patricia Santos [a], H.Ulrich Hoppe[b] & Cleo Schulten[b]
[a] TIDE Research Group, Universitat Pompeu Fabra
[b] RIAS Institute

Period addressed by the study

For this study, we analysed both websites and CS projects information. From the 72 websites selected, we extracted 4949 CS projects information. The list of websites and projects is consistently updated for the duration of the project (2019-2022). The data has been extracted from the CS Track database.

Main aim of the study

In this case study we intended to reflect on how the online data about CS is shared and communicated in the websites, how could this data be extracted massively and stored in a central database to, later be analysed with different purposes. One of its, studied in this article, is the usage of all the information in educational contexts.

Research question/s

This research was focused on answering the following questions:

- How is CS communicated and promoted on online websites?

- How automatic methods such as web scraping methods and anonymization techniques can be designed, developed and used to extract data from online sites? and How could these methods be applied to comply with the GDPR?
- Is it possible and how could this data be used for educational purposes?

Research Context

CS has a wide online presence; from online platforms dedicated to local, regional or global CS practice (such as *The Citizen Science Association (CSA-North America), the European Citizen Science Association (ECSA), the Australian Citizen Science Association (ACSA), Observatorio de la ciencia ciudadana (Spain) or Bürger schaffen Wissen (Germany)*), the ones dedicated to a single CS project (such as Mosquito Alert or Cities-Health) to the ones that contains information about CS project but are not oriented to CS practice (such as the sites of a research institute, a museum or a university). These websites objectives (especially those dedicated to CS), among others, is to make CS known and promote the participation and dissemination of CS projects (Vohland et al., 2021; Veeckman et al., 2019).

The communication of science through online media might contribute to promoting informal scientific knowledge (Stocklmayer et al., 2010). Furthermore, some previous studies identify citizens' participation in CS projects might promote knowledge, development of skills, awareness of real problems addressed by projects or motivation through STEM careers (Hiller & Kitsantas, 2014; Bonney et al., 2016; Kobori et al., 2016; Vohland et al., 2021). Considering all these assumptions, this study aims to explore how online websites communicate about CS projects and how all this information available can be used in formal education contexts, for instance, to promote scientific literacy or support teachers' practice (see section 5.3 for more information about previous proofs of concepts).

In combination with automatic techniques, which have been previously used to collect and better understand the data (Diouf et al., 2019; Ponti et al., 2018), from this study we could create a database with more than 4000 projects. By centralising the data from various sites (following the European General Data Protection Regulation (GDPR)) we expected to allow us to analyse the data structures of the websites to report the data and give a response to the research questions defined.

Research Method(s) applied

For this study we applied both computational methods (web scraping) and explorative study (manual analysis of the data extracted and websites information). From the manual analysis we wanted to identify how websites share information about CS online and analyse the technical architecture to better understand to what extent they apply the metadata standard. Especially, to know if they follow the Public Participation in Scientific Research (Citizen Science) metadata standard (PPSR_core metadata standard).

Procedure(s) applied

Figure 6.1.1 shows the process followed during the analysis (Calvera-Isabal et al., 2023):

1. **Selection of websites** following the criteria of (1) Contains CS projects information, (2) those are from Europe or allow participation of european citizens and (3) allow automatic data extraction.

2. **Analysis of website's** content and characteristics. Also understand how to share the information online.
3. **Developments and execution of the crawler** to extract and store the data.
4. **Analysis of the potential usage of the data** in formal education contexts.

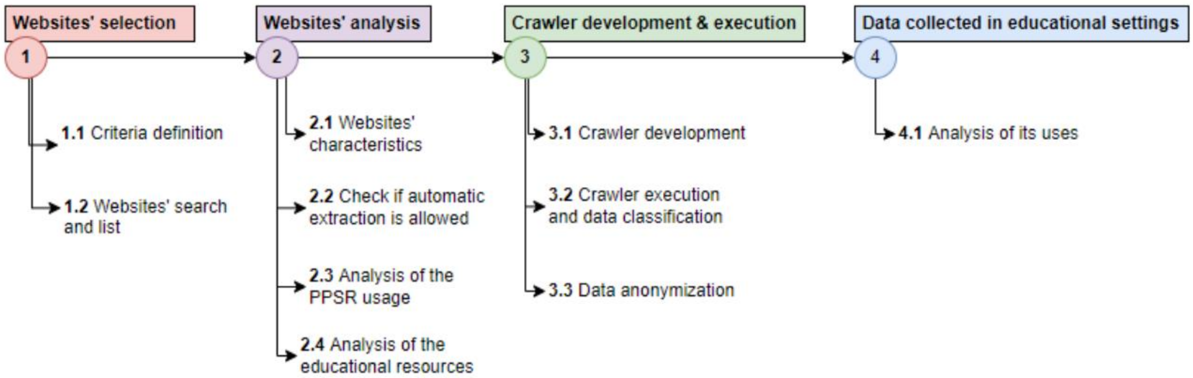

Figure 6.1.1. Process followed to extract and store the data used for the study

Summary of results/findings

After the analysis, for the CS Track database, we included 4 new additional categories to the PPSR_Core metadata standard. We observe that although the mandatory categories information is included in the 91.56% of the cases, there is still work to do from websites to take into account the PPSR_Code metadata standards. More detail about the results obtained after applying automatic methods are described in section 6.

Having access to CS massive data, online educational resources or tools developed or used by CS projects could also help teachers to create learning activities. For instance, to inspire them to create learning activities, to know more about how science is addressing real problems or allowing participation in CS projects following an educational perspective (by using the materials developed). Nevertheless, only 48.61% of websites analysed have educational material or information about learning. Although, the ones that allow online participation (such as Zooniverse) have specific educational sections. Likewise, they include information about tools used in CS projects that teachers could use in the classroom to support the student's learning process or enhance it. Finally, we expect that by exploring the data extracted and resources available teachers could improve their pedagogical skills and scientific knowledge. In the end, it might have an effect on student's knowledge and attitude toward science (Chan & Yung, 2018).

Conclusion

In order to improve the communication of CS projects or the accessibility and the analysis of the data, CS platforms might apply the PPSR more strictly. This could potentially help citizens find the key information about the CS projects and might motivate them to participate or could generate interest to know more about projects. The application of the standard would also facilitate the search and automatization of data extraction allowing algorithms such as NER to extract and classify data so it might improve the scientific knowledge of CS (e.g. SDGs (cf. section 4.2) or research areas (cf. section 6.2)).

A correct application of the standard would also help to support educational uses of CS data. Having the information structured and classified into the categories defined by the metadata standard and sharing the required information needed for teachers (see section 5.3) might help them to use this data in a formal educational context, inspire them to create learning activities or motivate them to participate in a project. This possibility was explored in other studies presented in this deliverable (see section 5.3) in which teachers explained that they use open resources, tools developed by others, their personal experiences and other teachers' practices to inspire them to create learning activities and adapt their practice.

Link to complete report:

A scientific article was published in the number 74 of the Comunicar journal: Calvera-Isabal, M., Santos, P., Hoppe, H., & Schulten, C. (2023). How to automate the extraction and analysis of information for educational purposes. [Cómo automatizar la extracción y análisis de información sobre ciencia ciudadana con propósitos educativos]. Comunicar, 74. https://doi.org/10.3916/C74-2023-02

Article on eMagazine: https://cstrack.eu/format/reports/how-to-automate-the-extraction-and-analysis-of-information-for-educational-purposes/

Link to dataset:

Database: https://zenodo.org/record/7356627#.Y39bEnaZNPY

List of descriptors: https://zenodo.org/record/7310445#.Y2zph3aZNPY

References

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. https://doi.org/10.1177/0963662515607406

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. https://doi.org/10.1007/s11165-016-9567-1

Diouf, R., Sarr, E.N., Sall, O., Birregah, B., Bousso M., & Mbaye, S.N. (2019). Web Scraping: State-of-the-Art and Areas of Application. In 2019 IEEE International Conference on Big Data (Big Data) (pp. 6040-6042). https://doi.org/10.1109/BigData47090.2019.9005594

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. https://doi.org/10.1111/ssm.12081

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. https://doi.org/10.1007/s11284-015-1314-y

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). https://doi.org/10.5334/cstp.101

Veeckman, C.M., Talboom, S., Gijsel, L., Devoghel, H., & Duerinckx, A. (2019). Communicatie bij burgerwetenschap: Een praktische handleiding voor communicatie en betrokkenheid bij citizen science. SCIVIL. https://bit.ly/3PKQz50

Vohland, K., Land-Zandstra, A., Ceccaroni, L., Lemmens, R., Perelló, J., Ponti, M., Samson, R., & Wagenknecht, K. (Eds.) (2021). The science of citizen science. Springer Nature. https://doi.org/10.1007/978-3-030-58278-4

6.2 Citizen science project descriptions as science communication texts - the good, the bad, and the ugly

Authors and Research affiliation

Yaela N Golumbic [a b], Marius Oesterheld [c] & Nicolas Felipe Gutierrez [d]

[a] The MOFET Institute
[b] The Steinhardt Museum of Natural History
[c] Wissenschaft im Dialog
[d] TIDE Research Group, Universitat Pompeu Fabra

Main aim of the study

Project descriptions are a central element of a Citizen Science project's online presence and thus play a key role in recruiting volunteers. Very often, they are the first point of contact between a project and prospective participants. As such, they need to be reader-friendly and accessible, spark interest, contain all the necessary practical information, and motivate readers to join by explaining convincingly how they will benefit from participating in the project. The purpose of this study was to examine whether the project descriptions stored in the CS Track database meet these criteria.

Research question

The questions this study aimed to answer were:

- *To what extent do CS project descriptions actually contain the kinds of information relevant to prospective participants? And is this information conveyed in a comprehensible and attractive manner?*

Research context

For the past two years, several research teams within the CS Track consortium have studied Citizen Science project descriptions stored in the CS Track database, examining for instance correlations with the SDG framework, educational aspects etc. What became apparent in the course of this work was that CS project descriptions vary greatly in terms of content, length and style. While some are so short they contain very little concrete information on the project's activities and the tasks to be completed by citizen scientists, other project descriptions provide lengthy and jargon-laden explanations of the project's scientific background. Moreover, many project descriptions fail to mention how volunteers will benefit from participating. In light of these observations, we decided to design a set of evidence-based recommendations for writing engaging project descriptions. The resulting product is an annotated template that offers general advice on length, format and style, as well as listing ten essential elements of an effective project description (cf. section 7.2). At the same time, we decided to examine the deficits described above in a more systematic and quantifiable manner.

Research Method(s) applied

To this end, we have conducted a qualitative content analysis of a random sample of 120 English-language project descriptions, using the above-mentioned ten-step template as a coding rubric.

Procedure(s) applied

For the purpose of this study, we applied two filters to the CS Track database. First, we created a dataset containing only English-language project descriptions. From this 2949-project dataset, we then excluded all descriptions which consist of less than 100 or more than 500 words. Texts of less than 100 words can not be expected to contain a significant amount of information. Project descriptions of more than 500 words are less likely to be read in their entirety than shorter texts, and thus ill-suited to the task of capturing the readers' interest and prompting them to join the project in question. This second round of filtering eliminated a staggering number of project descriptions – namely 1.666 –, leaving us with 1283 usable texts. Using this dataset, we created a random sample by applying the 'random' function of RStudio.

In total, we analysed the descriptions of 120 CS projects, which equals 9.35% of the filtered dataset and 2.42% of all project descriptions currently stored in the CS Track database.

The qualitative content analysis was performed in two consecutive steps. First, in order to ensure that the coding rubric is fit for purpose and all categories within it well-defined and demarcated, all three members of the research team independently coded 40 project descriptions. After discussing the results and making slight modifications to the coding rubric, each team member coded roughly one third of the remaining 80 descriptions.

Summary of results/findings

Preliminary results suggest that the majority of project descriptions in our sample fail to mention how citizen scientists will benefit from participating, what kind of training they will receive, how their contributions will be acknowledged, and whether they will have access to project results. Furthermore, the project's goals, its target audience, and the tasks volunteers will be expected to complete are very often not described explicitly and clearly enough. For instance, very few project descriptions contain concrete information on required skills and equipment or on the time commitment associated with participation.

Conclusion

Work is still in progress on this study. However our preliminary results suggest that work is needed in order to support project initiators in writing their project descriptions in an attractive and clear way. Project descriptions should ideally include, in addition to main goals and project impact, information on potential participants, tasks to be completed and training to assist participants in achieving these tasks. They should also include details on the benefits for participants, how they will be acknowledged and where they can access the data.

Link to complete report:

Since, at the time of submission of this deliverable - November 2022, the paper/full report concerning this study is still in progress, no link to a repository is available. Please contact the corresponding author (yaelago123@gmail.com) if you would like to receive further information about this research.

Article on eMagazine: https://cstrack.eu/format/reports/citizen-science-project-descriptions-as-science-communication/

7. The CS Track database: contribution of the empirical studies to enhancing understandings on/for CS

Ishari Amarasinghe [a], Miriam Calvera [a] & Patricia Santos [a]

[a] TIDE Research Group, Universitat Pompeu Fabra

As presented in D2.1 "Explorative study of CS projects in Europe, categorization and clustering to build a database of CS projects for analysis" WP2 has aimed to: (1) compile a database of CS projects (and their corresponding CS activities) in the European Union and Associated Countries; (2) to document a collection of these projects to explore their availability of data for further analysis (through WP3 and WP4) following the knowledge gaps identified by the literature review of WP1.

The implementation of the CS Track database has involved a gradual process. In section 7.1 we present the total of CS Platforms and projects collected during the total duration of the project.

Sections 7.2, 7.3 and 7.4 are focused on the analysis of three main research aspects of interest for CS Track: Research Areas in CS, Sustainable Development Goals and Skills of science inquiry represented in CS. These are three aspects that we have been able to scale (further to the empirical studies presented in previous sections) by using computational analytical methods developed in WP3 and WP2. Therefore, the results presented in these sections take into account the total number of projects collected in the DB or numerous of them (i.e. all projects with English descriptions).

In section 7.5, we present and discuss other research aspects that have the potential to be further developed to understand the characteristics of CS activities.

7.1 General overview of the CS Track database

General overview of the CS Track database (see Figure 7.1.1): number of projects, websites' countries, project description languages, number of platforms, distribution of websites types and distribution of websites' countries.

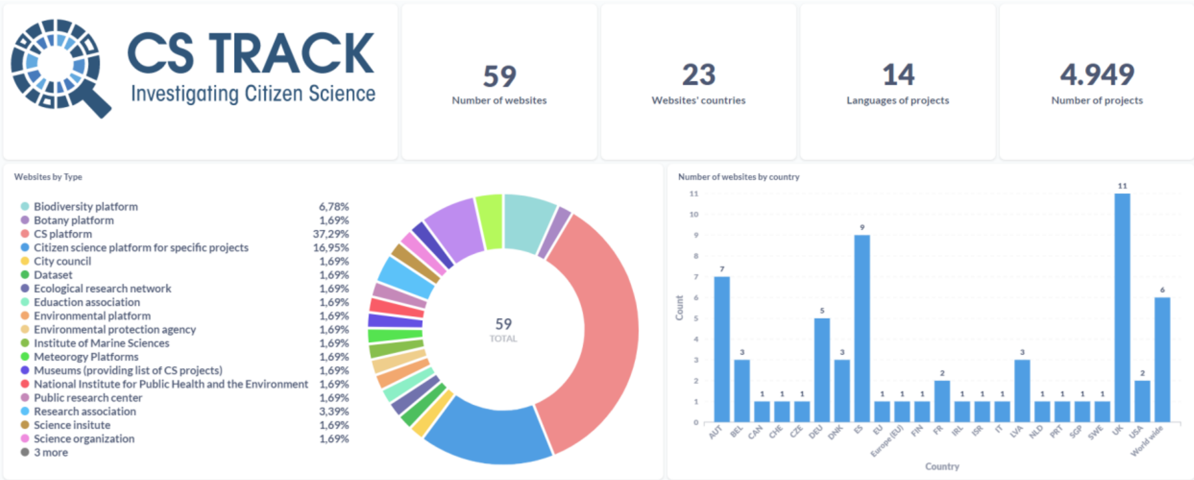

Figure 7.1.1. General overview of the Database. Number of projects, number of websites from CS projects information was extracted, websites by type and country, information languages.

Moreover, it should be noted that for some projects the related descriptions were derived from a number of platforms not from a single one. Hence, when conducting this analysis multiple platform assignments were taken into account as described in example below.

Example:
Wp2 ID (Platform ID): ["9" "89"]
Project Title: Fossilfinder
Note: In the above example, "Fossilfinder" project description has been retrieved from both platform 9 and 89 (composite assignment of platforms). In this case, when conducting further analysis at the level of the platforms the project "Fossilfinder" was considered to be derived from both platform 9 and 89.

Note: In total, there were 94 projects in which the descriptions were retrieved from more than one platform.

Note: Moreover there were 5 CS projects in the database namely: 1) Community Based System Dynamics (CBSD); 2) You + ME Registry and Biobank; 3) STEM+A@Astronomy; 4) SOCIÉTÉ FRANÇAISE POUR L'ÉTUDE ET LA PROTECTION DES MAMMIFÈRES (SFEPM); 5) Where? Where? Wedgie!) without a platform assignment. Those projects were not considered for the following analysis.

7.2 Research Areas in Citizen Science

Note: The following analysis was conducted using data retrieved from the CSTrack database on 2022/09/15. At this point the database consisted of 4949 CS Project records (this includes English and non-English descriptions).

CS projects are classified (following the algorithm to classify CS information into research areas proposed in the context of WP3) considering the following 5 main research areas:

- Arts & Humanities
- Life Sciences & Biomedicine
- Physical Sciences
- Social Sciences
- Technology

Each of the 5 main research areas consist of a number of related sub research areas (More details of the taxonomy can be found here: https://images.webofknowledge.com/images/help/WOS/hp_research_areas_easca.html). More detail can be found in D3.2 'Web Analytics Toolset and Workbench' - ESA backend.

For instance, at present in the CS Track database the research area assignment for each project is indicated as follows:

Example 1:
Project Title: Penn State Astrobiology Citizen Science Project
Project Description: ["We want to study the biogeography of microorganisms by taking water samples from domestic water heaters. Participants will acquire a water sample from their kitchen tap and answer 20 questions. The process will take ~30 minutes. We are recruiting 2-3 households per state. By looking at the genetic differences from isolates of similar microbes from across the globe, researchers are currently trying to understand the degree to which populations of microbes are isolated and whether this isolation suggests an allopatric speciation model for prokaryotes. We are still looking for participants in: AL, AK, DE, DC, KS, KY, ME, MA, NH, NM, ND, RI, SC, SD, TN, VT." "Sign up to participate: http://www.scienceforcitizens.net/PSARC"]
Research Areas: ["Physical Sciences, Water Resources, 0.6778448864490314"]
Interpretation: In the above example, Penn State Astrobiology project has been assigned a single main research area which is "Physical Sciences" and a sub research area called "Water Resources". The similarity score for this assignment is given as 0.67.

Example 2:
Project Title: Great Lakes Worm Watch
Project Description: ["The Great Lakes Worm Watch needs citizen scientists to conduct earthworm surveys in forests and other habitats anywhere in North America." "The project website provides instructions and data sheets for conducting your own earthworm, habitat, and soil surveys in the "Conduct your Own Surveys" section: http://greatlakeswormwatch.org/team/conduct.html" "If you feel you

need more help in designing a study, you can contact the project coordinators with particular questions at:"]
Research Areas: ["Social Sciences, Archaeology, 0.35307771700172963" "Life Sciences & Biomedicine, Limnology, 0.2888119357350136"]
Interpretation: In the above example, the Great Lakes Worm Watch project has been assigned two research areas and sub-research areas. However, "Social Sciences" [main research area] "Archaeology" [sub research area] received a higher similarity score of 0.35 when compared to the other assignment "Life Sciences & Biomedicine" [main research area] and "Limnology" [sub research area] which received a score of 0.29. * It should be noted that in the following sections when presenting the results of the research areas allocation, we only considered the highest similarity assignment.

In this section, the research area classification results are reported considering the following three questions:

Q1: What is the distribution of research areas at the project level? (considering the 5 main research areas listed above)

Q2. In each research area what is the most common sub research area?

Q3. What is the distribution of research areas at the platform level?

The data was preprocessed in order to answer the aforementioned questions. It was noted that 100 records consisted of missing values in "Research Areas". Therefore, the following analysis ultimately considered 4849 records.

All the results presented in the following under (Q1, Q2 and Q3) are given in Zenodo.

Zenodo URL: https://zenodo.org/record/7310341#.Y2zhgXaZNPY

Link to github: https://github.com/CS-Track-Code/project-categorization/blob/main/research_areas_assignment.ipynb

Q1: What is the distribution of research areas at the project level? (considering the 5 main research areas listed above)

In answering Q1, Figure 7.2.1 below indicates the research area assignment considering 4849 projects.

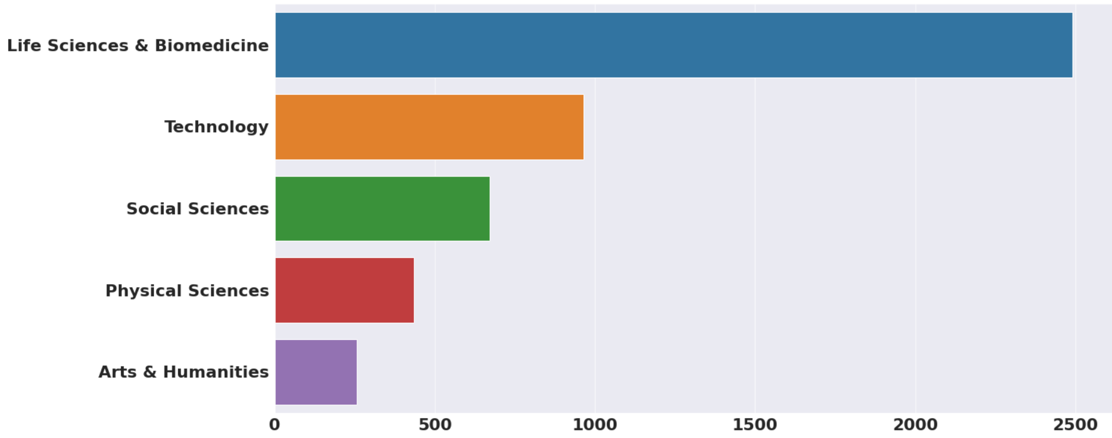

Figure. 7.2.1. Research Area assignment

As it can be seen in Figure 7.2 the majority of projects have been assigned to the "Life Sciences & Biomedicine" category (2492 projects), followed by the "Technology" category (965 projects) and the "social sciences" category (671 projects). Several projects have also been assigned to the "Physical Sciences" category (434 projects) and "Arts & Humanities" category (257 projects). There are also 30 projects that have not been assigned to any of the 5 main research areas in the dataset analysed (and was indicated using []).

Q2. In each research area what is the most common sub research area?

In answering Q2 we extracted the sub research area with the highest similarity score (See example 2 above) for each of the 5 main research areas. Due to the high number of sub research areas associated with each research area in this section we only provide the top 3 sub research areas related to each research area.

As it can be seen in Table 7.2.1 when considering the "Life Sciences & Biomedicine" research area a large number of projects were seen to relate with the "Biodiversity & Conservation" type (682 projects). In the "Technology" category most projects were related to the "Remote Sensing" (393 projects) and in "Social Sciences" a high number of projects are related to the "Education & Educational Research" sub type (121 projects). When considering the "Physical Sciences" a large number of projects were seen to related to the "Water Resources" sub research area (165 projects) and finally in the "Arts & Humanities" research area a high number of projects were identified as related to the "History & Philosophy of Science" sub research area (122 projects).

Table 7.2.1. Top sub research areas

Main Research Area	Sub Research Area	Count
Life Sciences & Biomedicine	Biodiversity & Conservation	682
	Environmental Sciences & Ecology	240
	Ornithology	233

Technology	Remote Sensing	393
	Construction & Building Technology	210
	Telecommunication	63
Social Sciences	Education & Educational Research	121
	Archaeology	82
	Public Administration	62
Physical Sciences	Water Resources	165
	Astronomy & Astrophysics	144
	Sustainability Science	23
Arts & Humanities	History & Philosophy of Science	122
	History	52
	Literature	34

Q3. What is the distribution of research areas at the CS platform level?

In the following we report the percentage of research area allocation considering the platforms. It should be noted that in total the CS project descriptions were derived from 59 CS platforms. Hence, we chose to report the results considering a selected list of 5 platforms as shown in Table 7.2.2 the criteria for the selection was:

- CS Platforms that allow European citizen to participate online
- CS Platforms that cover Europe area as a whole
- CS platforms for specific European countries
- CS platforms for specific European regions
- CS platforms that are involved actively in the promotion of CS (to measure it, we explored how often they actualize the content)

Table 7.2.2. Selected list of platforms

Name of the Platform	Platform URL
Zooniverse	https://www.zooniverse.org/projects
EU citizen science	https://eu-citizen.science/projects
Citizen Science Vlaanderen	https://www.scivil.be/en/projects
Ciencia Ciudadana España	https://ciencia-ciudadana.es/proyecto-cc/
Schweiz forscht	https://www.schweiz-forscht.ch/de/citizen-science-projekte

In the following we present the results of the research area assignment to CS projects considering the 5 platforms listed in Table 7.2.2 As presented below in Table 7.2.3 it can be observed that all 5 platforms consist of a high number of projects that are related to the Life Sciences & Biomedicine category. In general, the platforms consist of a smaller number of projects related to Physical sciences and Arts and Humanities categories.

Table 7.2.3. Research area assignment considering a list of selected platforms

Name of the platform	No. of projects related to Life Sciences & Biomedicine	No. of projects related to Technology	No. of projects related to Social Sciences	No. of projects related to Physical Sciences	No. of projects related to Arts & Humanities
Zooniverse	221	71	15	47	11
EU citizen science	90	34	24	12	12
Citizen Science Vlaanderen	10	4	2	2	1
Ciencia Ciudadana España	44	66	46	12	16
Schweiz forscht	32	21	9	2	2

7.3 Sustainable Development Goals (SDGs) covered by CS projects

Note: The following analysis was conducted based on data retrieved from the CSTRack database on 2022/09/15.

CS projects are classified (following the algorithm to classify CS information into Sustainable Development Goals proposed in the context of WP3) considering the 17 SDGs (see https://sdgs.un.org and Figure 7.3.1). More detail can be found in D3.2 'Web Analytics Toolset and Workbench' - ESA backend.

Figure 7.3.1. Sustainable Development Goals

(image source: https://www.researchgate.net/publication/333982248_Water_in_the_2030_Agenda_for_Sustainable_Development_How_can_Europe_act/figures?lo=1&utm_source=google&utm_medium=organic)

At present in the CSTrack database the SDG assignment for each project is indicated as follows:

Example 1:
Project Title: Volunteer Water Quality Monitoring Program
Project Description: ["All Missourians rely on streams in one way or another and many of our streams could use a little help. They need teams of people who love clean water, good fishing and health habitat to take care of them, year after year. That's why the Missouri Department of Conservation, the Department of Natural Resources and the Conservation Federation of Missouri joined to develop the Stream Team Program in 1989." "To learn more and to join the Missouri Stream Team Program visit our website at www.mostreamteam.org." "We have many different events and activities going on all through our the year so be sure to check out our calendar of events." "Our Volunteer Water Quality Monitoring portion of the program is one of the most popular activities to participate in and is also the only activity that requires training. So watch for information on our training workshops as well."]
SDGs: ["SDG, SDG #6, 0.3616005970498504" "SDG, SDG #15, 0.3243498682077864" "SDG, SDG #8, 0.3175175955755153" "SDG, SDG #14, 0.3063678875541241" "SDG, SDG #3, 0.2933202469723035" "SDG, SDG #4, 0.2709218003634529" "SDG, SDG #1, 0.26135416081346163" "SDG, SDG #11, 0.25928983372858355" "SDG, SDG #10, 0.25744076387086257"]
Interpretation: In the above example, the Volunteer Water Quality Monitoring

Program has been assigned a number of related SDGs. However, "SDG#6" received a higher similarity score of 0.36 when compared to the other assignments.

* It should be noted that in the following sections when presenting the results of the SDG assignments we only considered the highest similarity assignments.

We report the SDG assignment to CS projects considering the following three questions:

Q1: What is the distribution of SDGs at the project level? (considering the 17 SDGs)

Q2. What is the distribution of SDGs at the platform level?

Before answering the questions we pre-processed the data. It was noted that 100 records consisted of missing values. Therefore, the following analysis ultimately considered 4849 records.

All the results presented in the following under (Q1 and Q2) are given in Zenodo.

Zenodo URL: https://zenodo.org/record/7310353#.Y4STY3aZOUk

Link to github: https://github.com/CS-Track-Code/project-categorization/blob/main/SDGs_assignment.ipynb

Q1. What is the distribution of SDGs at the project level? (considering the 17 SDGs)

In answering Q1, Figure 7.3.2 below indicates the results of the SDG assignment to CS projects.

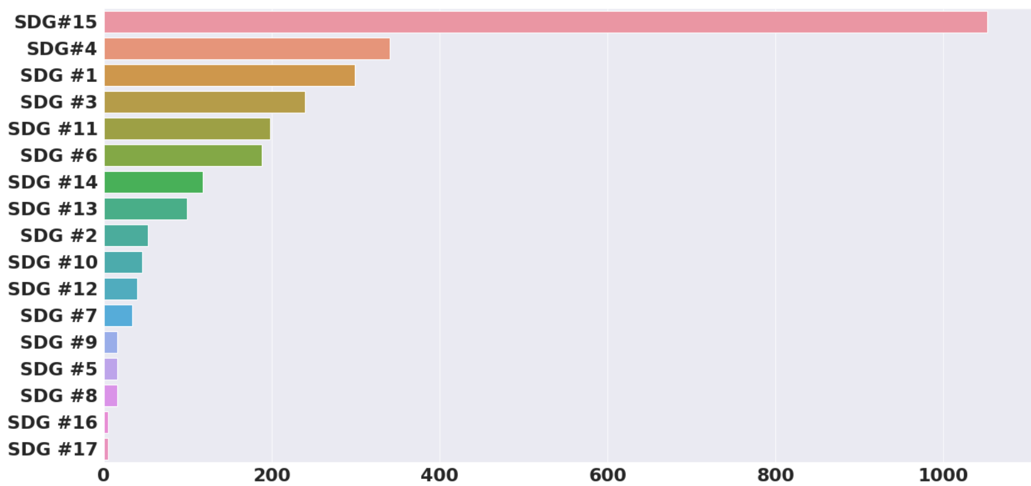

Figure 7.3.2 SDG assignment

As it can be seen in Figure 7.4 notably the majority of projects have been assigned to the "SDG#15":Life on land (1052 projects), followed by the "SDG#4": Quality Education (341 projects) and the "SDG#1":No Poverty (299 projects).

Several projects have also been assigned "SDG#3" (240 projects), "SDG#11" (198 projects), "SDG#6" (188 projects) and "SDG#14" (118 projects). SDG#13, #2, #10, #12, #7, #9, #5, #8, #16, #17 were assigned to less than 100 projects.

Notably "SDG#16" and "SDG#17" have been assigned to only 5 projects each.

Q2. What is the distribution of SDGs at the platform level?

In the following table 7.3.1 we present the results of the SDG assignment to CS projects considering the platforms listed in Table 7.2.3. Details of all assignments are given in Zenodo. Based on the results presented in Table 7.4 it can be observed that a high number of projects hosted in Zooniverse, EU Citizen Science and Citizen Science Vlaanderen platforms are related to SDG#15 (Life on land). On the contrary, a high number of projects hosted in Ciencia Ciudadana España platform are related to SDG#4 (Quality Education) and SDG#1 (No Poverty). Schweiz forscht platform also consisted a high number of projects related to SDG#1 (No Poverty). These two platforms may host projects that are of interest to the specific country (Spain and Germany). Schweiz forscht platform also consist relatively a high number of projects related to SDG#15 (life on Land).

Table 7.3.1 SDG assignment considering a list of selected platforms

Name of the platform	#1	#2	#3	#4	#5	#6	#7	#8	#9	#10	#11	#12	#13	#14	#15	#16	#17
Zooniverse	27	2	10	9	2	2	12	1	0	2	7	0	16	8	106	1	0
EU citizen science	21	6	13	24	0	3	1	1	0	2	12	2	3	3	29	0	0
Citizen Science Vlaanderen	1	0	1	3	0	1	1	0	0	0	1	0	1	0	4	0	0
Ciencia Ciudadana España	26	3	7	30	1	7	4	1	2	2	19	5	3	16	0	0	0
Schweiz forscht	14	0	2	2	0	1	1	0	2	1	5	2	3	1	13	0	0

As it can be observed in this overview analysis, in general the results obtained are aligned with the ones summarised in section 4.2 in this report.

The most popular SDG in CS, according to our results is SDG#15 (Life on Land).

We observe that SDGs such as: SDG#4 (Quality Education) and in this case (where more projects have been analysed) SDG#1(No poverty) and SDG#3 (Good Health and well-being) emerge as the most represented ones in the list (see Figure 7.3.2). We think the main reason for this is that these SDGs cover transversal topics (i.e. health, poverty, education) that can be associated with multiple disciplines. This results also show why the connection between CS and Education is important (as discussed in different sections of the deliverable).

The results shown in figure 7.4 shows other important topics covered in CS are related to: SDG#11 (Sustainable Cities and Communities), SDG#6 (Clean Water), SDG#14 (Life below water) and SDG#13 (Climate Action). These results allow us to better understand the relationship between the most popular Research Areas shown in figure 7.2.1, where we observe that Life Sciences and biomedicine is the most important research area.

7.4 Skills of science inquiry in projects' descriptions

Note: The following analysis was conducted based on data retrieved from the CSTRack database on 2022/09/15.

In order to classify project descriptions based on the presence of skills of science inquiry we used supervised machine learning algorithms. Details related to the analysis procedure are provided in a paper which is currently under review (on the date of submission of this deliverable - November 2022). Contact the main authors if you have interest to receive further information.

We report the skills classification considering the following two questions:

Q1: What types of skills of science inquiry CS projects promote?

Q2: What is the distribution of skills in CS Platforms?

All the results presented in the following under (Q1 and Q2) are given in Zenodo.

Zenodo URL: https://zenodo.org/record/7332112#.Y3agYi1Q1hE

Q1: What types of skills of science inquiry CS projects promote?

First, in order to answer Q1, we created a labelled dataset in which two raters annotated 178 CS projects as positive considering 20 skills of science inquiry. The results are shown in Figure 7.4.1. Based on the results it was seen that the majority of the selected projects promoted skills of science inquiry such as "observe", *"search"*, *"collect"*, *"find"*, *"analyse"*, *"discuss"*, *"record"*, *"locate"* and *"share"*. Moreover, skills such as *"insert"*, *"comment"*, *"measurements"*, *"fill out"*, *"come up with"*, *"identify"*, *"count"*, *"enter"*, *"note"*, *"transcribe"*, *"answer"* appeared in less than 5% of the projects.

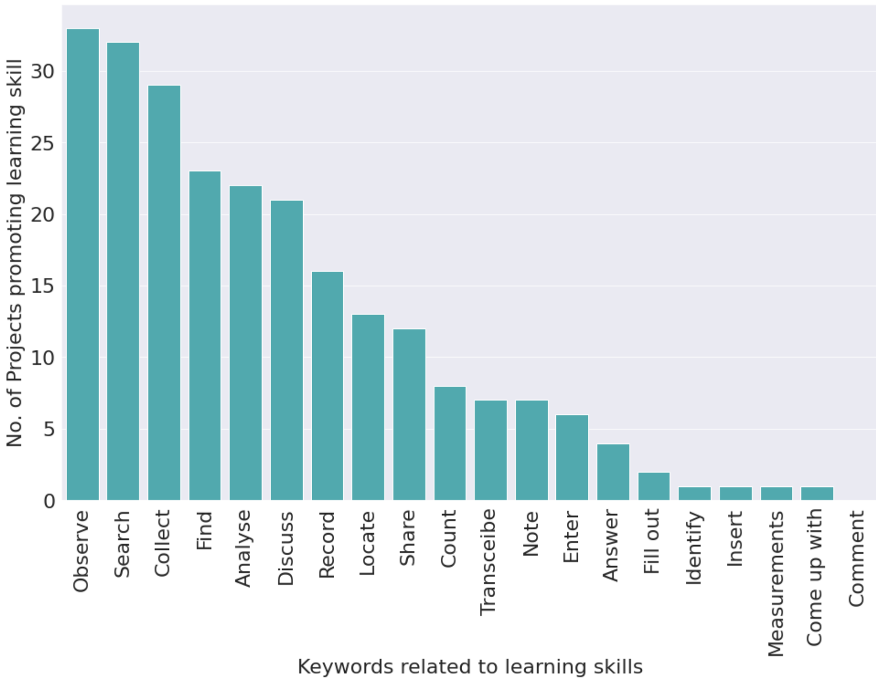

Figure 7.4.1. Presence of skills of science inquiry in 178 projects (Training Dataset)

We then used the annotated dataset to train machine learning models for the presence/absence of the skills of science inquiry. A fine-tuned language model (BERT) outperformed classical approaches (SVMs, RFs, MNBs) when used to predict skills of science inquiry in CS project descriptions using the annotated dataset (F1 score of 0.84).

The model was then used to predict 2939 project descriptions in english. Based on the prediction results 438 projects (around 15%) were labelled as '1' indicating the presence of one or more skills of science inquiry. Those project description texts were preprocessed and then counted to retrieve how many projects have mentioned skills of science inquiry. The results are presented in Figure 7.4.2 below. It should be noted that the following skills *"Collect"*, *"Search"*, *"Observe"* and *"Find"* have a strong presence in both training (see Figure 7.4.1) and prediction datasets (see Figure 7.4.2). This not only indicates that many projects promote such skills, but also the machine learning model was able to apply what was learned using the training data. It was also noted that several projects promote skills such as *Count, Enter, Answer and Note,* that were not strongly present in the manually annotated training dataset. The high presence of "Collect", "Search" etc. could be due to the contributory nature of the projects which is out of the scope of this analysis and will be considered in the future.

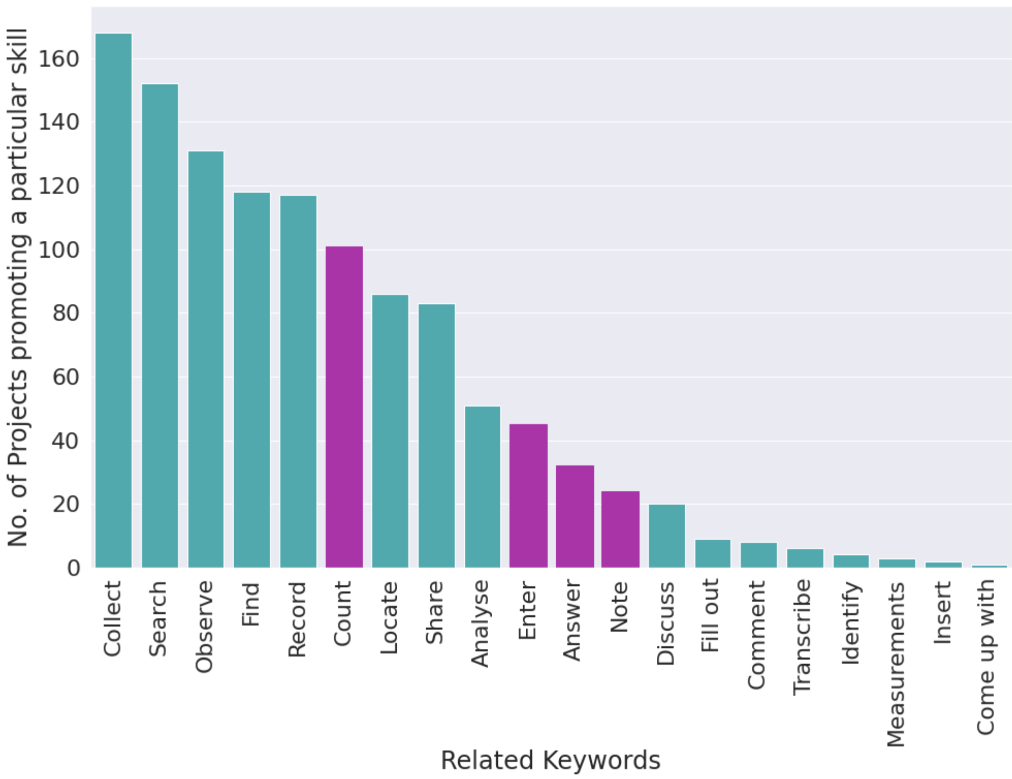

Fig. 7.4.2. Presence of skills of science inquiry in 438 CS projects

Q2: What is the distribution of skills in CS Platforms?

In the following we present the results of the skills analysis considering the platforms listed in Table 7.4.1.

As presented below in Table 7.5 it can be observed that *Zooniverse, EU citizen science* and *Citizen Science Vlaanderen* platforms consist of projects that promote science inquiry skills. Our prediction results did not include any projects retrieved from *Ciencia Ciudadana España* and *Schweiz forscht* indicating a lack of science inquiry skill promotion in projects hosted in those two platforms.

Table 7.4.1. Science inquiry skills assignments considering a list of selected platforms

Skill	No. of projects retrieved from Zooniverse promoting skills	No. of projects retrieved from EU citizen science promoting skills	No. of projects retrieved from Citizen Science Vlaanderen promoting skills
1. Collect	9	21	0
2. Search	14	14	1
3. Record	7	11	0
4. Identify	1	0	0
5.Observe (or observations)	8	15	0
6. Find	12	4	0
7. Count	9	5	0
8. Enter	3	4	0
9. Note	6	2	0
10. Transcribe	2	2	0
11. Share	4	10	0
12. Answer	7	2	1
13. Locate	6	9	0
14. Discuss	2	2	1
15. Analyse	6	4	1
16. Insert	0	1	0
17. Comment	1	0	0
18. Measurements	0	0	0
19. Fill out	0	0	0
20. Come up with	0	0	0

7.5 Other descriptors to automatically categorise CS activities

This section contains a study performed with exploratory purposes with a limited number of projects from the CS Track database. The main aim was to better understand the potential of text analysis techniques (such as NER) to classify information from project descriptions and other descriptors typically reported by CS projects. This study is a continuation of the analysis carried out in the "D2.1: Explorative study of CS projects in Europe, categorization and clustering to build a database of CS projects for analysis". As explained in this deliverable based on our previous work , analysing the PPSR metadata standard and how different CS platforms organize CS data from projects, we identified a list of descriptors associated with CS projects. However, in general, some of these descriptors (e.g. Project objective, Geographical location… see more examples below) are typically empty in the corresponding CS platform. In this study we wanted to understand if this information can be extracted from the corresponding textual description of the project, or from other associated descriptors.

The objectives of this study were:

- Identify which aspects of interest can be further understood based on the information contained in the CS Track database (including project descriptions and other descriptors)
- Analyse which computational techniques such as text processing techniques, machine learning algorithms and neural networks are more efficient to extract information and create new descriptors.
- Analyse the results obtained, identify advantages and limitations of the techniques used.

This section presents an analysis that explores the potential to better understand certain aspects described in CS projects. In order to do this, we randomly selected 45 projects with the information stored in English from the three types of platform structures (N = 15 for each one) (see section 6.1 and their references for more information). We identified 6 descriptors that could be completed with information extracted automatically with the techniques selected:

- Project objective
- Geographical location
- Main program or person in charge
- Start date, end date, status and duration
- Participants profile

Below, the analysis and results of each descriptor.

7.5.1 Project objective

We consider the project objective as: *the purpose of the project, as well as what initiative is being promoted* (see section 5.20 of the D2.1 for more information). After the exploratory analysis of the N = 45 CS project descriptions we can conclude that this information is available in the texts for almost all the projects analysed. Furthermore, in general, it is described in a single sentence.

From the texts analysed, we selected manually a list of keywords that are contained in the project objectives' sentences: *"goal", "purpose", "objective", "intention", "ambition", "promote", "dedicate"*. From this list, we included synonyms (Finally, 66 synonyms) (using python *nltk* library and function *synsets()*). Nevertheless, in order to minimise the error of selecting sentences that doesn't fit with the definition of Project objective defined, we calculated the similarity (with *SpaCy* python library and *similarity()* function) and selected a threshold (a value of 0.7) to select or reject the sentences based on that value. In order to obtain more accurate results, the selection of projects might be refined and could be selected based on a pair revision of Cs projects descriptions. Figure 7.1.5 shows one objective extracted from the CS project description.

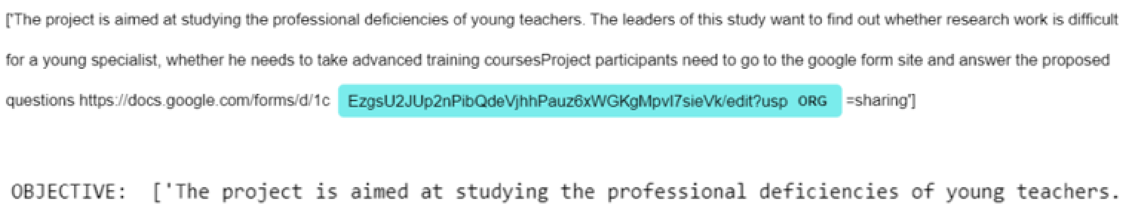

Figure 7.5.1. Objective extracted from a CS project description.

For further information and results of the classification, see [section 7.5.6](#).

7.5.2 Geographical Location

The geographical location refers to where the CS project is carried out or has been carried out (see section 5.16 of the D2.1 for more information). After the exploratory analysis of the N = 45 CS project descriptions we can conclude that this information is available in the texts for almost all the projects analysed. Nevertheless, there is not a common criteria because in some cases there is information of a country, a city or the whole container. In some cases, participation is online so the geographical location information might not correspond to the CS project geographical location but other related information.

In this analysis we applied the Name Entity Recognition (NER) technique (using *SpaCy* python library and *nlp()* function) which identifies location by assigning the labels 'LOC' and 'GPE' (Geopolitical Entities) to the text. Figure 7.5.2 shows the location extracted from the CS project description among other categories.

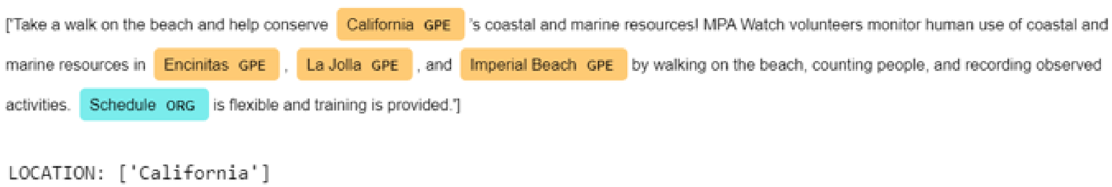

Fig. 7.5.2. Geographical location extracted from a CS project description.

7.5.3 Main program or person in charge

The main program or person in charge category contains information about the project responsible, coordinator team, organisation, association or ONG in charge or which finance the project (see section 5.14 of the deliverable 2.1 for more information). After the exploratory analysis of the N = 45 CS project descriptions we can conclude that this information is available in the texts but for some cases it is not sufficiently clear the relation of the person with the project. We used NER so it also identifies the organisation with the label 'ORG'. We have discarded the ones that correspond to the CS project name. Figure 7.5.3 shows the results of the identified categories after applying NER technique.

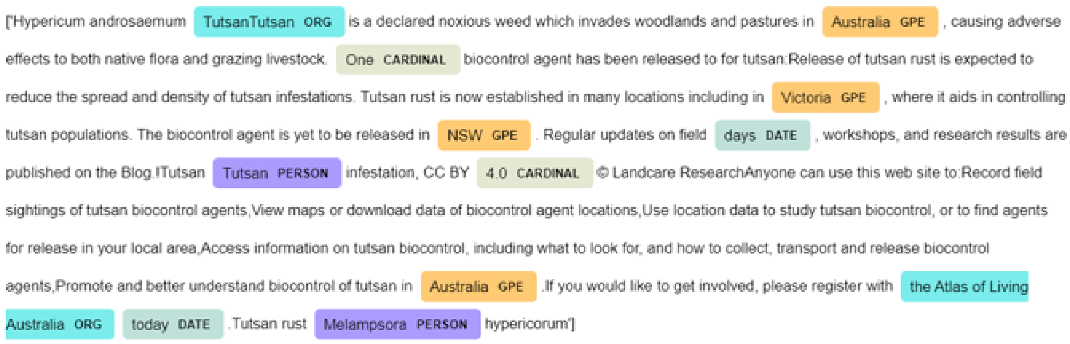

Fig. 7.5.3. Results after applying NER to a CS project description.

7.5.4 Start date, end date, status and duration

Based on the project descriptions selected, the result was that there was not much information about the start and end date of projects. We couldn't know if the project was ongoing, if they were looking for participants or if the project had finished. We think the information associated with dates is important in order to better understand the matureness of a project. We applied NER to identify the dates, informed by the 'DATE' label (see Figure 7.5.3 for more information).

In this case we conclude that for this information (i.e. start date, end date, status and duration) we can not obtain data from the project descriptions. In general, it is not informed the start and end date.

All the information stored in the CS Track database (see sections 5.3, 5.17 & 5.18 of D2.1) for these categories was extracted automatically from structured or semi-structured websites (see section 6.1 for more information).

7.5.5 Participants' profile

In some cases, there is information of the participant's profile in the CS project descriptions. For instance, if the research requires a specific population group or if they create a call for participation for school. For this study of the exploratory analysis of the N = 45 CS project description, we identified eight keywords: *'anyone'*, *'adults'*, *'students'*, *'university students'*, *'kids'*, *'area community'*, *'18 years'*, *'group of'* and *'not specified'*. In cases where we do not find a relation with the selected keywords, we assign the category *Anyone*. Nevertheless, it is important to refine this assignment because for most of the cases, not having information about the participants doesn't mean anyone can participate.

7.5.6 Discussion and conclusion

After the preliminary study (N = 45), we applied the algorithms to the CS projects in the database with the descriptions in English (N = 2637). From this, we analysed manually the results obtained for each descriptor and identified the correct and incorrect assignments (true positive, true negative, false positive and false negative). With this results, we calculated an error of the 17% of the projects (N = 150) for each category, and we have obtained:

Table 7.5.1 - Error calculated for the 17% of the projects

Descriptor	Error (%)

Project objective	20%
Geographical location	16,67%
Main program or person in charge	18,67%
Participants	26,67%

The error is higher in cases where the codebook should contain a higher list of keywords (e.g. project objective, participants). This error could be mitigated in the future by refining the list of keywords and synonyms selected and trying to identify new ones that fit better with the definition of each category. As the selection of the projects was randomly and balanced between the types of projects, we can conclude that being a significant sample of the database, the error calculation could be extrapolated to all the descriptions of the database in English. Therefore also to the translations into English of the texts that are not originally in that language.

Some preliminary results from the global analysis of CS project descriptions in English:

Fig. 7.5.4. Word Cloud of the most words used in the Project objective sentences extracted.

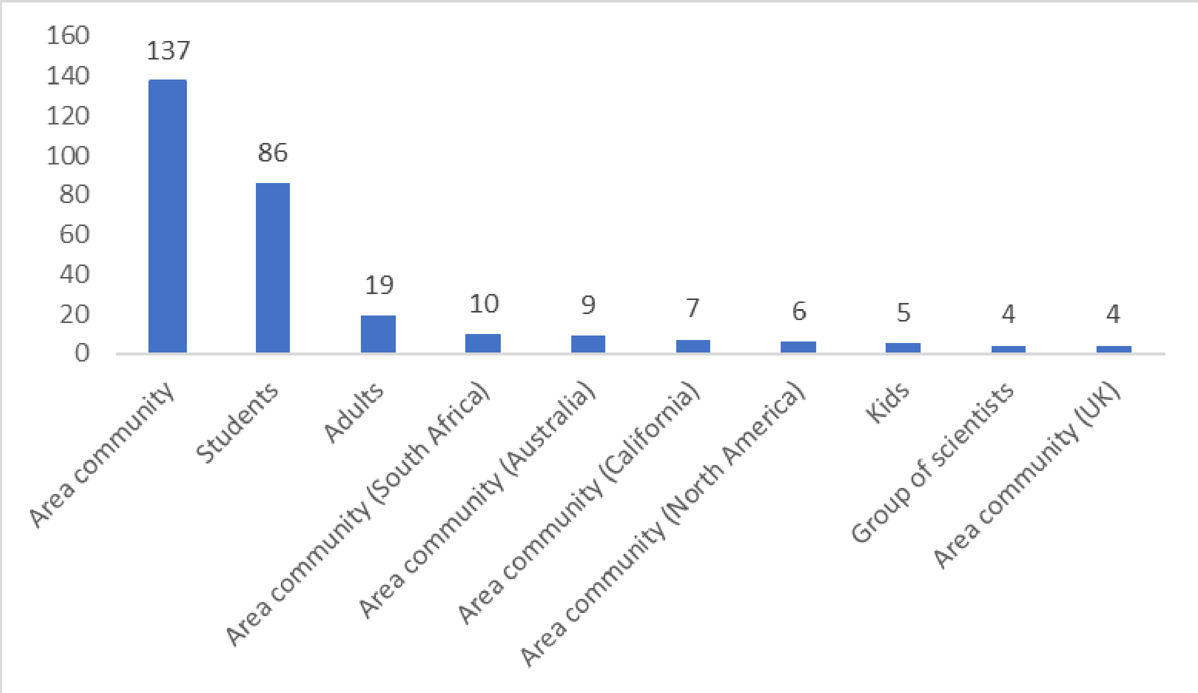

Fig. 7.5.5. Top 10 Participants' profile data obtained without the category Anyone' which had 1755 values.

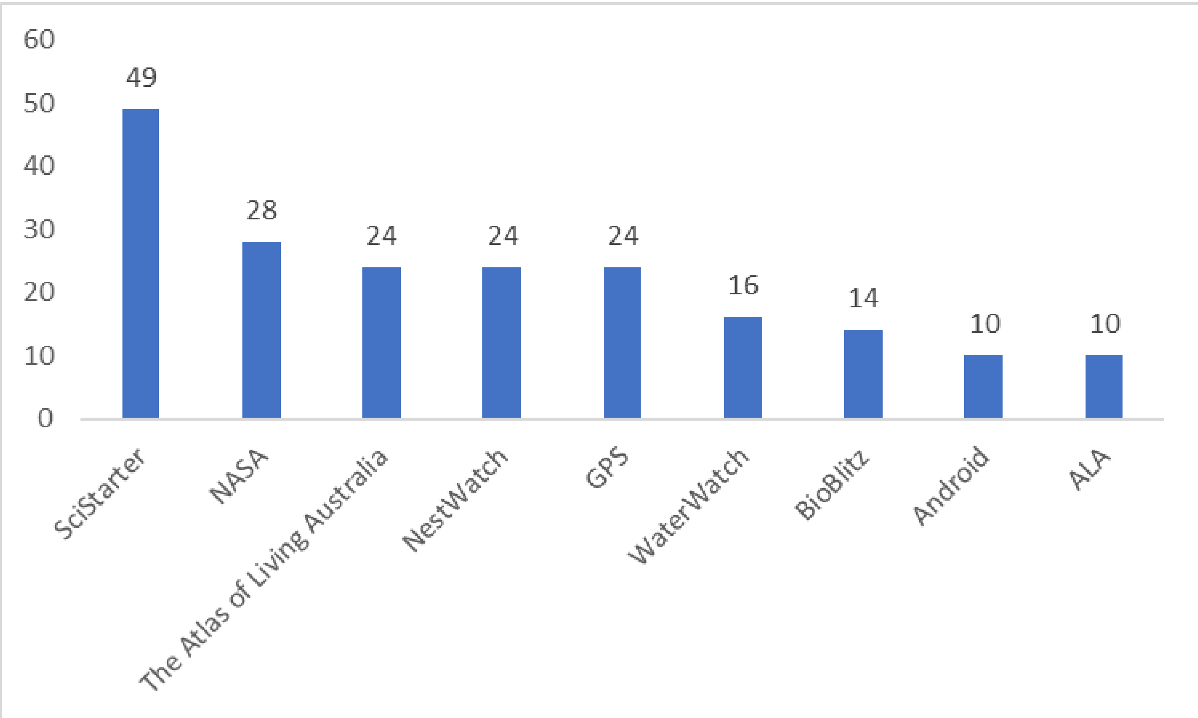

Fig. 7.5.6. Top 10 Main program or person in charge data obtained.

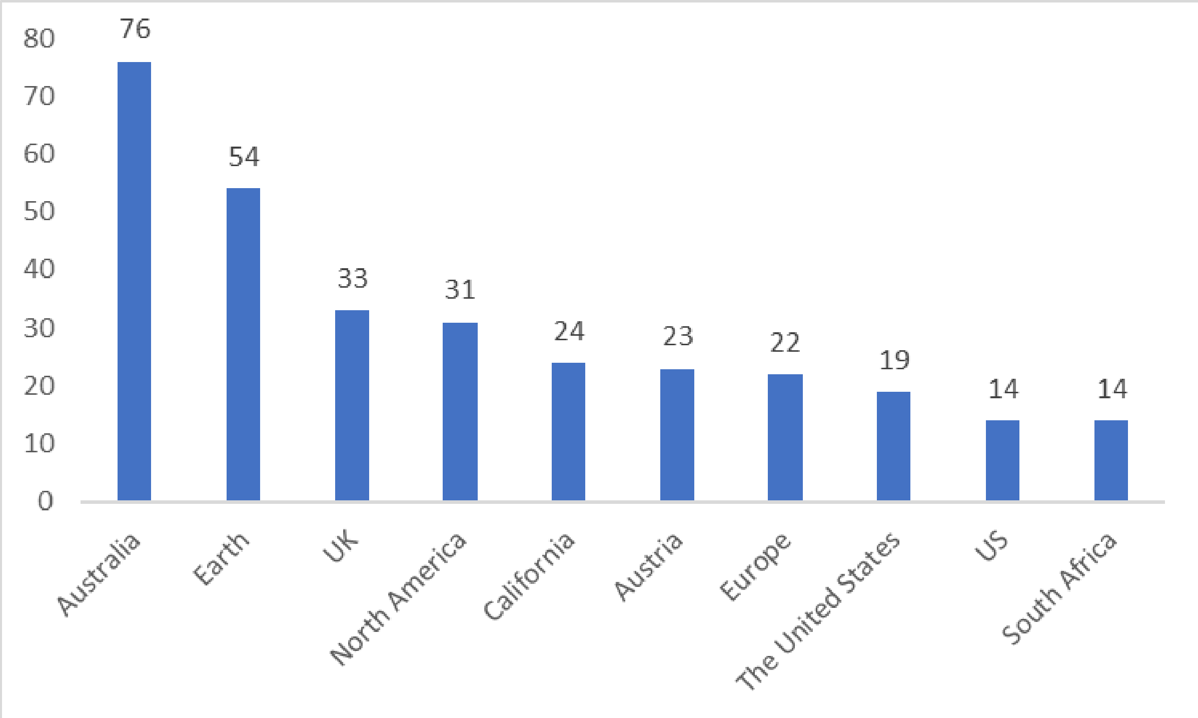

Fig. 7.5.7. Top 10 Geographical location data obtained.

The algorithms developed as part of this study were applied only to a limited sample of projects in order to understand their advantages but also to identify their limitations. According to the results obtained, some of the tasks done need further refinement in order to be able to apply them with a larger dataset of projects. However, the preliminary results show the potential of these techniques to automatically classify CS data within categories of interest.

Link to dataset:

Descriptors definition: https://zenodo.org/record/7310445#.Y2zmbXaZNPY

References

Martínez Almansa, Mara. Aplicación de técnicas computacionales de extracción y análisis de datos para entender mejor la ciencia ciudadana. 2022 http://hdl.handle.net/10230/54792

8. Guidelines for documenting CS projects

Authors: Yaela N Golumbic [a b], Marius Oesterheld [c] & Nicolas Felipe Gutierrez [d]

[a] The MOFET Institute
[b] The Steinhardt Museum of Natural History
[c] Wissenschaft im Dialog
[d] TIDE Research Group, Universitat Pompeu Fabra

Much of the work described in this deliverable is based on analyses of CS project descriptions stored in the CS Track database. Over the past 18 months these have been studied from different perspectives, focusing for instance on research area, correlation with the SDG framework, motivational factors, educational aspects etc.

What has become apparent over the course of our research is that CS project descriptions are extremely heterogeneous – both in terms of format, length, and style, and regarding their contents. This chapter will discuss what we have learned in the process and suggest guidelines for writing clear and informative project descriptions.

8.1 Learning about CS projects from project descriptions

As mentioned above, we found CS project descriptions posted on CS platforms to vary greatly in terms of their content, length and style. Some are so short that they hardly contain any information on the basic characteristics of the project (location, timeframe, technology used etc.), on the specific tasks to be completed by the citizen scientists, or on the skills required to perform these tasks. Other project descriptions provide lengthy and excessively technical explanations of the project's scientific background that are difficult to understand for non-experts. Additionally, many project descriptions fail to explain how volunteers will benefit from participating. As a result, some of these texts are not particularly well-suited to the task of sparking interest and prompting readers to join the projects they represent.

This is quite unfortunate since project descriptions are often the first point of contact between a CS project and prospective participants and thus play a crucial role in recruiting volunteers. At the same time, project descriptions that hardly contain any information make it difficult for researchers to investigate CS on a larger scale.

8.2 Guidelines for writing project descriptions that spark interest and attract volunteers

Having read hundreds of project descriptions from the CS Track database and observed the deficits discussed above, we decided to design an evidence-based template for writing engaging project descriptions. In addition to providing general advice on length, style and presentation, this template guides the reader through the process of writing a project description using ten simple steps. The template provides guidelines regarding the main topics and types of information to include in descriptions, offers explanations and suggestions for items to consider in each step and provides examples from two hypothetical project descriptions. The annotated template is available as a PDF download on the CS Track Zenodo page [https://zenodo.org/record/7004061#.YyAvTexBzRM].