Crowd4SDG
Citizen Science for the Sustainable Development Goals

Deliverable 4.3

In-situ assessment report of citizen local interactions and self-reporting GEAR cycle 1

Deliverable identifier: D4.3

Due date: 30/04/2021

Justification for delay: due to exceptional activity as agreed with Project Officer

Document release date: 29/06/2021

Nature: Report

Dissemination Level: Public

Work Package: 4

Lead Beneficiary: UP

Contributing Beneficiaries: UNIGE

Document status: Final

Abstract:

In this report, we present a preliminary data-driven approach leveraging digital traces to describe the GEAR cycle 1, through participant demographic profiles, team diversity, activity and interaction dynamics. In addition, we introduce CoSo (Collaborative Sonar), a digital platform (app and web) for monitoring team collaborations and task management through self-reports and surveys within both in-situ and online contexts. These insights serve as a base for i) exhibiting the potentiality of using digital traces on online tools to derive measures related to team process, and ii) highlighting how newly developed technologies allow us to measure collaboration metrics and descriptors in the upcoming GEAR cycles.

For more information on Crowd4SDG, please check: http://www.crowd4sdg.eu/

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 872944.

Document history

	Name	Partner	Date
Authored by	Marc Santolini, Camille Masselot	UP	26/03/2021
Revised by	Jesus Cerquides Elena Proden, Madina Imaralieva	CSIC UNITAR	02/04/2021
Edited by	Marc Santolini, Camille Masselot	UP	03/05/2021
Reviewed by	Jose-Luis Fernandez-Marquez	UNIGE	30/05/2021
Edited by	Marc Santolini, Camille Masselot	UP	10/06/2021
Approved by	Jose-Luis Fernandez-Marquez	UNIGE	15/06/2021

Table of Contents

Project Partners

	Partner name	Acronym	Country
1 (COO)	Université de Genève	UNIGE	CH
2	European Organization for Nuclear Research	CERN	CH
3	Agencia Estatal Consejo Superior de Investigaciones Científicas	CSIC	ES
4	Politecnico di Milano	POLIMI	IT
5	United Nations Institute for Training and Research	UNITAR	CH
6	Université de Paris	UP	FR

Crowd4SDG in brief

The 17 Sustainable Development Goals (SDGs), launched by the UN in 2015, are underpinned by over 160 concrete targets and over 230 measurable indicators. Some of these indicators initially had no established measurement methodology. For others, many countries do not have the data collection capacity. Measuring progress towards the SDGs is thus a challenge for most national statistical offices.

The goal of the Crowd4SDG project is to research the extent to which Citizen Science (CS) can provide an essential source of non-traditional data for tracking progress towards the SDGs, as well as the ability of CS to generate social innovations that enable such progress. Based on shared expertise in crowdsourcing for disaster response, the transdisciplinary Crowd4SDG consortium of six partners is focusing on SDG 13, Climate Action, to explore new ways of applying CS for monitoring the impacts of extreme climate events and strengthening the resilience of communities to climate related disasters.

To achieve this goal, Crowd4SDG is initiating research on the applications of artificial intelligence and machine learning to enhance CS and explore the use of social media and other non-traditional data sources for more effective monitoring of SDGs by citizens. Crowd4SDG is using direct channels through consortium partner UNITAR to provide National Statistical Offices (NSOs) with recommendations on best practices for generating and exploiting CS data for tracking the SDGs.

To this end, Crowd4SDG rigorously assesses the quality of the scientific knowledge and usefulness of practical innovations occurring when teams develop new CS projects focusing on climate action. This occurs through three annual challenge based innovation events, involving online and in-person coaching. A wide range of stakeholders, from the UN, governments, the private sector, NGOs, academia, innovation incubators and maker spaces are involved in advising the project and exploiting the scientific knowledge and technical innovations that it generates.

Crowd4SDG has six work packages. Besides Project Management (UNIGE) and Dissemination & Outreach (CERN), the project features work packages on: Enhancing CS Tools (CSIC, POLIMI) with AI and social media analysis features, to improve data quality and deliberation processes in CS; New Metrics for CS (UP), to track and improve innovation in CS project coaching events; Impact Assessment of CS (UNITAR) with a focus on the requirements of NSOs as end-users of CS data for SDG monitoring. At the core of the project is Project Deployment (UNIGE) based on a novel innovation cycle called GEAR (Gather, Evaluate, Accelerate, Refine), which runs once a year.

The GEAR cycles involve online selection and coaching of citizen-generated ideas for climate action, using the UNIGE Open Seventeen Challenge (O17). The most promising projects are accelerated during a two-week in-person Challenge-Based Innovation (CBI) course. Top projects receive further support at annual SDG conferences hosted at partner sites. GEAR cycles focus on specific aspects of Climate Action connected with other SDGs like Gender Equality.

Grant Agreement description of the deliverable

The focus of Work Package 4 (WP4), led by the University of Paris, is to conduct research on Citizen Science. This encompasses the establishment of methods and the collection of data to inform the development of effective, high-quality citizen science projects. To that aim, this work package develops metrics and statistical models in order to assess the many-faceted outcomes of the citizen science projects developed within the Crowd4SDG consortium.

By leveraging the digital traces from online tools that document project progress and citizen engagement, it is possible to quantitatively monitor and analyze i) the activity of teams working within the GEAR collaborative framework, and ii) the activity and engagement patterns of citizen science participants. As such, we propose to frame the participants' behaviour, their interactions and engagement with tools used during the GEAR cycles.

This deliverable addresses the following tasks:

T4.2: Measuring analytics of citizen collaborations using new metrics/descriptors on digital traces (UP, UNIGE)

The epistemological analysis performed in Task 4.1 provides the conceptual foundation for specific new metrics and descriptors for Citizen Science. While individual learning progress can be monitored in a straightforward manner by the increase in relative level of expertise on specific keywords/topics (see, for example, the iLearn project at CRI), it remains unclear how such insights can be generalized to assess overall learning progress achieved in specific challenges/projects and the entire scope of projects initiated within the scope of Crowd4SDG. For this reason, this Task will develop and implement new metrics and descriptors that not only measure productivity and output, but also assess the overall learning and research dynamics as well as the diversity, originality, relevance, robustness, and adaptiveness of the knowledge produced in the context of each individual citizen-innovation team as well as the entire group of citizen-participants across all CS projects within Crowd4SDG. The work involves the development of specific algorithms based on a fundamental discussion of the kind of knowledge Citizen Science is expected to produce and whether/how it differs from knowledge produced by conventional scientific approaches (see Task 4.1). In the proposed project, participants will have access to and assess in a distributed, large-scale manner the local impact of the Crowd4SDG program. By monitoring the activity patterns of participants when they use the Citizen Science Solution Kit (examples: project documentations of SDG in progress or the community management solutions of CrowdBuilder) to conduct real time analyses of the collaboration within the teams, we can frame the participants' behaviour, their interactions and engagement within the O17 Challenges that form part of the project.

These digital traces will be used to explore the previously defined metrics of citizen science quality. The work in this task will build upon previous work by the partners in the context of the Open Science and the community-based Epidemium Challenge on Cancer Epidemiology . Using data from the various tools offered by the CS Solution Kit (see section 1.3.4.2), namely SDG in progress and Pybossa, we will in a similar manner quantify team diversity (skills and backgrounds), dynamics (bursts of activity), organisational structure (core-periphery network structure and leadership dynamics), as well as the influence of physical meetups on team activity.

T4.3: In-situ assessment of citizen local interactions and self-reporting (UPD, UNIGE)

In this Task a self-report smartphone application will be developed in order to be able to establish deeper insights into collaboration dynamics. The app will be able to assess social proximity through smartphone sensors (Bluetooth, wifi and GPS) and trigger context-based notifications. We intend to use this app to provide fine-grained measurements of citizen collaborative dynamics, both in terms of social network dynamics and in terms of learning experience through the self-reporting contextualisation it offers. The application will be based on an open-source platform developed by researchers at the Child Mind Institute in New York in the context of a large-scale mental health study. Their platform, Mindlogger , is a general-purpose open source data collection platform that can be used by anyone to administer surveys, quizzes or different types of tasks. Proximity-based notifications will be implemented to trigger “Ecological Momentary Assessments" asking individuals to document the type of interaction they are currently part of, so that we obtain an augmented understanding of the social context. Lastly, we will provide reports that provide insights into network structure and summary statistics. The reports will be in the form of a dashboard tool based on open source software MITeams . The results of this task will be used for assessing participant interactions during the different phases of the GEAR methodology cycle and thus feed back into WP3. The self-reporting phone app will be used for the in-person events to be carried out in the Accelerate Phase carried (see Task 3.2).

1. Purpose and scope of the deliverable

WP4 aims to develop new metrics and statistical models in order to assess the many-faceted outcomes of the citizen science projects developed within the Crowd4SDG consortium over the 3-years course of the project.

WP4 has two specific objectives:

- Develop new standardized metrics and descriptors for measuring the diversity, originality, effectiveness, sustainability/robustness and adaptation/appropriateness of solutions and insights gained from Citizen Science projects.
- Implement the metrics and descriptors as tools to analyse the digitized records of Citizen Science collaborations and the solutions and insights they produce.

By doing so, the WP4 contributes to the following specific objectives of the Crowd4SDG project:

- SO1.2: “Create CS projects and study the mechanisms that lead to improved citizen science skills and high-quality scientific outcomes.”
- And SO2.2 “Produce economic and social outputs relevant to achieving SDGs through challenge-based CS events, with a special focus on climate change resilience”

For the first year, we present an overall picture of the GEAR cycle 1 carried out by Crowd4SDG. This includes a presentation of the available data to assess the Citizen Science collaborations and the solutions and insights they produce.

The GEAR methodology provides online coaching of Citizen Science (CS) teams in a challenge-based innovation framework for CS projects related to climate resilience through four phases. The Gather phase is a call for CS projects on a specific SDG theme. During the Evaluate phase, the selected participants take part in the Open17 (O17) 5-week coaching programme, to develop their CS ideas in virtual teams towards compelling pitches. The pitches are judged by a panel in the final week of the coaching programme. The selected projects from the O17 then move to the Accelerate phase which consists of a two-week intensive CBI workshop held by CERN. Finally, the best projects are invited to present themselves during an international event on SDGs, the Geneva Trialogue, and dedicated sessions to collect feedback from representatives of various stakeholders (UN agencies, National Statistical Offices, academic CS experts, private sector and NGO representatives).

The Task 4.3 originally included the development of a sensor-based contact tracing app to monitor *in-situ* team interactions using smartphone sensors (Bluetooth, wifi and GPS). This became both unrealistic in the context of a fully online program and infeasible in practice (novel, stricter regulations were put in place following the COVID pandemic concerning contact tracing apps since 2020). Here we discuss how we overcome this setback by 1/ leveraging the digital traces now available given the online nature of the program as well as 2/ monitoring team progress and task management by developing a stand-alone app that allows team members to report collaborative tasks and answer surveys about their team structure and organization.

In parallel to the quantitative analysis presented here, we conducted an epistemological enquiry (D4.1) complemented by a survey of citizen science experts to extract relevant metrics of quality of citizen science, both from a conceptual and a practical point of view.

These newly developed metrics will be the object of the Deliverable 4.2 (M24) and will be incorporated into the analysis of Cycles 2 and 3.

As such, this deliverable feeds the work of Task 4.2 by exploiting available digital traces in preparation for the implementation of the metrics and descriptors and of the self-report app that will result in the deliverable 4.2 in month 24. It also describes the results of the Task 4.3 after we adapted to online activities during the GEAR cycle.

This deliverable is structured as follows. Section 2 describes the overarching research questions. Section 3 describes the data collection and analysis methods. Section 4 presents the results on team diversity, dynamics and interactions. Finally, Section 5 presents discussions and perspectives for the next GEAR cycles.

2. Research questions

In this report, we provide a descriptive analysis to answer the following questions related to the GEAR cycle 1:

- What are the data sources available to describe the activity of participants and the corresponding teams?
- To what extent did the participants use the several digital platforms available to them?
- What is the demographic diversity of participants in the Crowd4SDG project?
- What is the engagement of the teams during the GEAR cycle?
- What interactions do we observe within teams, between teams, and with the coaching and organisation teams?
- Are the provided set of tools and their usage sufficient to inform team dynamics and potential success of projects?

Based on this preliminary work, we expect to address the following questions in the next two years:

- How much of the project data can be captured from the digital traces in the various Crowd4SDG platforms?
- What are the activity patterns of the citizen science participants on the different online tools?
- How do interactions with funders, researchers, and the various citizen stakeholders impact scientific knowledge and technological innovations?
- To what level do we observe CS projects to be more diverse, original, relevant, flexible, robust, sustainable, cost-effective than would be expected in traditional research settings?
- How much can digital traces about engagement, diversity and interactions inform about the local impact of the program in terms of projects producing significant scientific or socio-economic results?

3. Methods

3.1. Data collection

3.1.1. Summary of data type and availability for the GEAR1 analysis

The online tools provided to the participants over the course of the GEAR cycle consist of Goodwall[1], Slack[2], Citizen Science Project Builder[3], Decidim4CS[4], SDG in Progress[5], and Epicollect[6]. The tools that are being developed by the Crowd4SDG partners are further described in the deliverable D2.1. We present in Figure 1 an overview of the availability of data for the GEAR 1. The identification of the tools usage and data availability is a first step toward gathering digital traces and self-reports from participants at the GEAR cycle 2.

The data shared by the different platforms for the GEAR 1 analysis encompasses traffic analytics of Goodwall and SDG in Progress (processed data, pdf format) and Slack data (raw data, csv format).

	Potential data	Available data
GOODWALL	User profile User activity Project information	Screenshots of trafic analytics (pdf)
SLACK	User profile User activity User interactions	User profile User activity Channels Metadata (csv)
PROJECT BUILDER	Data sets Text and images	-
DECIDIM4CS	User profile User activity User interactions Team information	-
SDG IN PROGRESS	User profile User activity Project information	Screenshots of trafic analytics (pdf)
EPICOLLECT	-	-

Figure 1: Overview of the availability of data for the GEAR 1 analysis.

[1] https://www.goodwall.io/
[2] https://slack.com/
[3] https://lab.citizenscience.ch
[4] https://decidim4cs.iiia.csic.es/
[5] https://sdginprogress.com/
[6] https://five.epicollect.net/

The following sections present the data shared in further detail.

3.1.2. Goodwall data

Elements from Google analytics were provided to the Crowd4SDG coordination team. They contain information about user engagement in the page related to the program on the platform for the month of September. While Goodwall is rich in terms of data related to user profiles, activity and projects, no such data could be made available to the Crowd4SDG team due to difficulties to reach the person in charge of the data. While we focus in our analyses below on the fine-grain communication data obtained from Slack, we plan in the future to work with the Goodwall team prior to the Gather phase in order to avoid similar hurdles.

3.1.3. Slack data

Slack was used by teams during the different phases of the GEAR cycle as a means to communicate with other teams and with the organizing team. Given that collecting such data was not initially planned, but revealed to be a potential source of insights on team interactions, we decided to build a pipeline to extract and analyse the traces. Different Slack workspaces were used for each phase. Here, we focus on the “Evaluate” phase, consisting of coaching events and providing the largest team base (20 teams). The Slack data of the “Accelerate” was too scarce to be included in our analysis. This provides a pilot study on what observables can be extracted, and how they can inform us on team dynamics and engagement.

The extracted data consists of:

- User Metadata: This includes data from the workspace members’ profile such as the Slack ID, user name, profile description, timezone, status (admin, bot or others).
- Channel Metadata: This includes the Channel ID, description, creators, members, attributes (is private, is shared) and pinned messages.
- Messages: Each message has an ID, timestamp, sender, text, reactions (users who reacted to the message with an emoticon), and a reference ID to a parent message if it is a part of a thread.

3.1.4. SDG in Progress data

The coordination team also obtained global usage data from the “SDG in Progress” platform, aggregated per month and presented over the span of two years. However, data was not exclusive to the Crowd4SDG project participants, and we did not have any possibility to match the list of participants to the ones identified in the Slack workspace. We provide recommendations in the discussion section as to how we can better integrate the various platforms in the future to optimize the monitoring process.

3.1.5. Other source of data

Finally, the coordination team compiled demographic data of participants in the Gather and Evaluate phase. The analysis of this data is presented in the results section.

Additional data was collected via the dissemination of surveys to participants at the end of the evaluation phase. This data is presented in the D3.3 report.

3.2. Data analysis

The Slack data and the overall data provided by the coordination team were analysed using the R software. Codes and anonymized data can be found in the open source Github repository.

Using the available Slack data, we used the number of posts and number of reactions of a user as a marker of individual engagement, or team engagement when aggregated over team members. Furthermore, we built social interaction networks where a user is linked to another user if he/she mentions him/her, with a weight corresponding to the number of mentions. This allows to represent the flow of information characterizing this phase, in particular highlighting the interactions with mentors and coaches.

4. Results

4.1. Overview of tool usage during the GEAR cycle 1

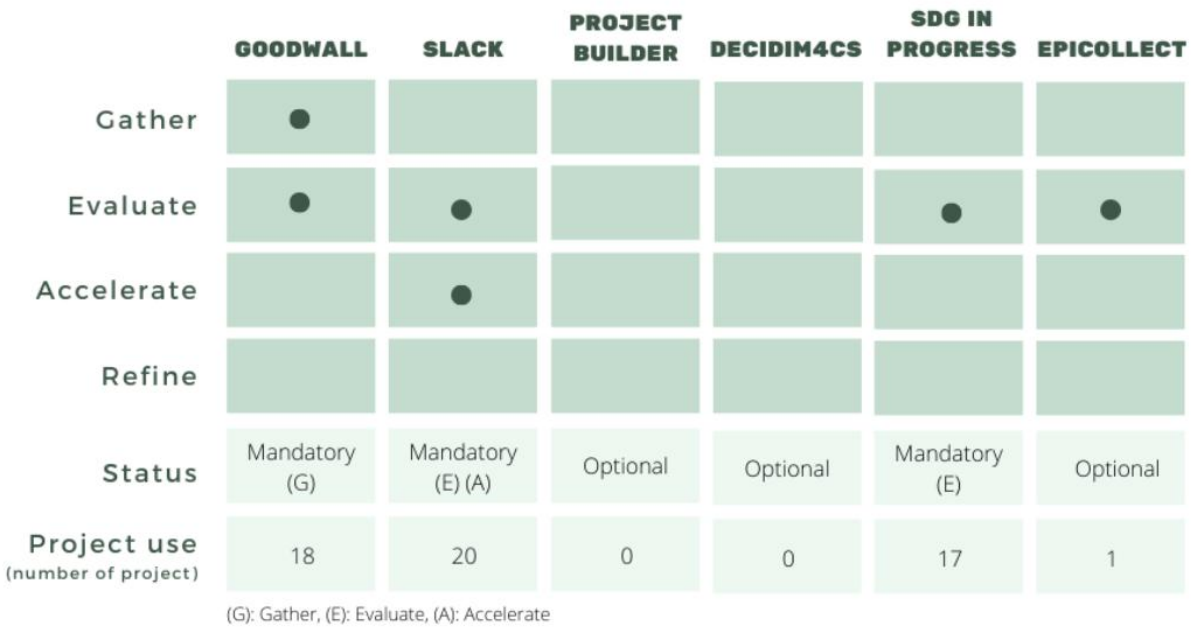

Figure 2: Overview of the tool usage during the GEAR cycle 1. Check marks indicate that a tool is made available during a given phase.

As a reminder, we show in Figure 2 the various tools proposed to the participants, along with their usage (see D3.3 for more details).

Tools with features that could benefit certain but not all projects were made optional, whereas documentation and communication tools were made mandatory. When the tool is made mandatory its use is widespread across most teams. Goodwall was the only mandatory tool for participants during the gather phase whereas Slack and SDG in Progress were mandatory during the evaluate phase. However, when made optional tools were significantly less used.

The Evaluate phase comprises the highest number of tools (4 in total). The Gather phase and the Accelerate phase only used one.

4.2. Profile of participants

A total of 48 people participated in the GEAR 1. Their age ranged between 14 and 36 years old. Half of them were below 21 years old and 56% were women (Figure 3a). The boxplot presented in Figure 3b shows that the average male participant was 22.5 years old and the average female participant was 20.2 years old. However this apparent difference of age across gender was not statistically significant (p-value = 0.11, Wilcoxon test).

a

Characteristic	N = 48[1]
gender	
Female	27 (56%)
Male	21 (44%)
age	21.0 (14.0, 36.0)

[1] Statistics presented: n (%); Median (Range)

b

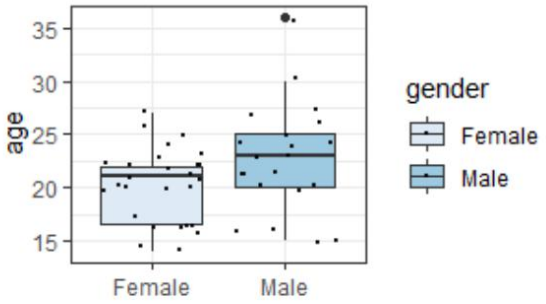

Figure 3: Age and gender of GEAR 1 participants

Around 20% of participants came from Nigeria (Figure 4). The predominant continent among participants was Asia which counted for 18 of the participants. Africa followed with 14 participants while Europe counted 10 participants, and North America and South America both counted 3. This higher number of Nigerian participants may be linked to Goodwall activities at the moment of the Gather phase. Indeed, at this time Goodwall started a collaboration with UNICEF focusing on Nigeria. As pitches circulated in Goodwall, it is probable that they were shown to these new members.

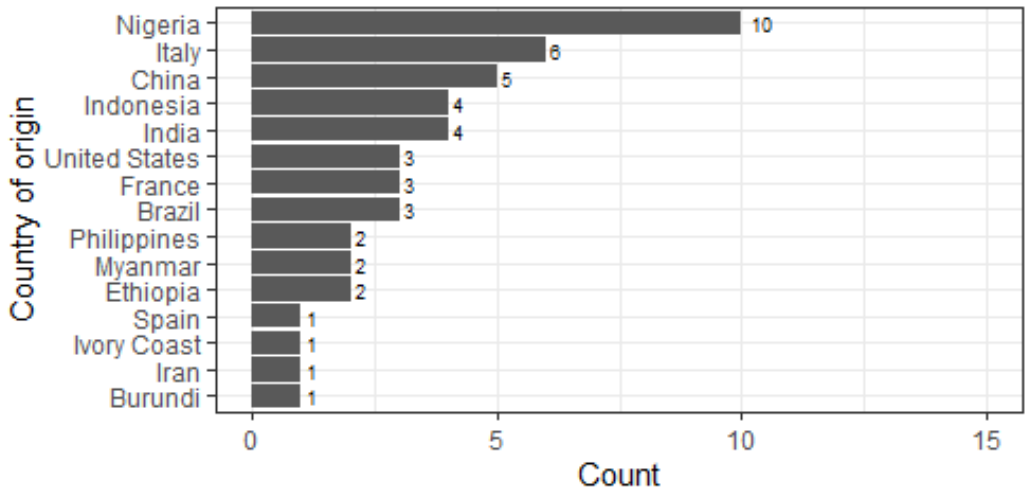

Figure 4: Number of participants per country of origin.

4.3. Diversity of teams

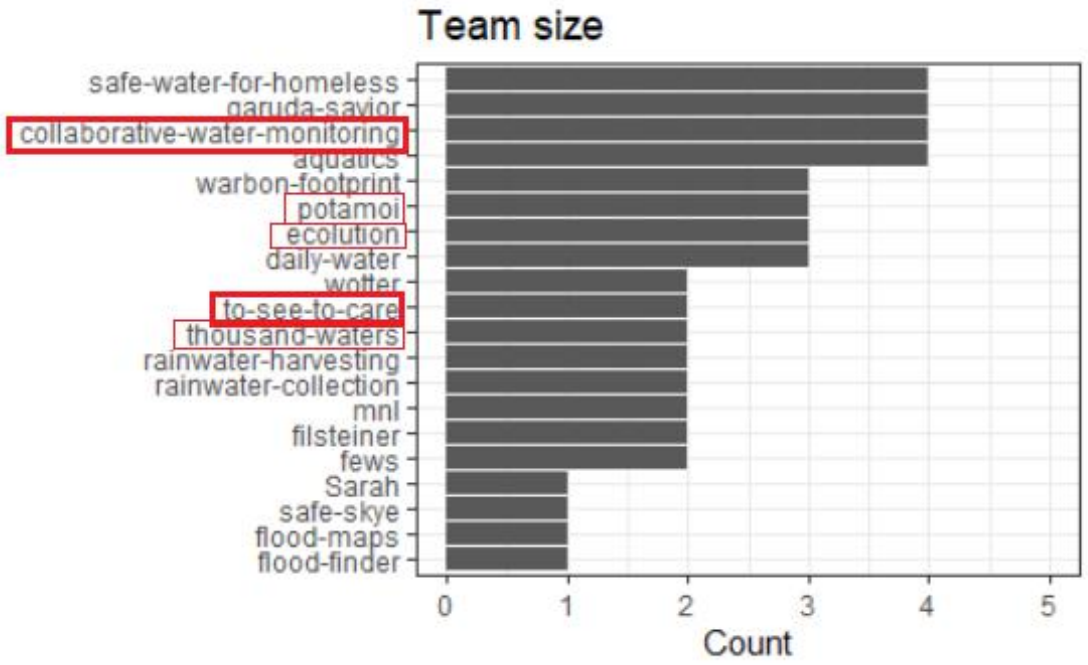

Figure 5: Number of participants per team. Red rectangles designate the teams selected at the end of the Evaluate phase for the next Accelerate phase, and bold ones designate the teams eventually selected for the final Refine phase.

The number of team members for the 20 teams registered for GEAR 1 spanned between 1 and 4 people (Figure 5). 60% of the teams had 2 or less people.

The five teams selected for the Accelerate phase all had at least two team members (Figure 5, red rectangles). The teams “Collaborative-water-monitoring” and “To-see-to-care”, selected for the “Refine” phase, had respectively four and two people.

Figure 6: Sankey diagram showing the team members nationalities (right) across teams (left). Red rectangles designate the teams selected for the Accelerate phase and bold ones designate the teams selected for the Refine phase.

Five teams were international, with members belonging to at least two countries: “Potamoi”, “Collaborative Water Monitoring”, “Daily Water”, “Wotter”, and “Rainwater Collection” (Figure 6).

Out of the five teams selected for the Accelerate phase, two teams were international teams. “Collaborative Water Monitoring”, one of the two finalist teams, gathered the highest level of geographical diversity among its members, with three continents represented, while the other spanned two continents.

Nigerians were the most ubiquitous members among teams, with members represented in six teams. They were followed by Indians and Americans (three teams), and Brazilian, Chinese and French (two teams). Interestingly, “Ecolution” and “Rainwater Collection” illustrated intra-continental collaboration by bridging West and East Africa.

Figure 7: Sankey diagram of teams (left) and their members' time zone (right) identified through Slack data. Red rectangles designate the teams selected for the Accelerate phase and bold ones designate the teams selected for the Refine phase.

Beyond international members, we found 3 additional teams with members across at least two time zones (Figure 7). Members of "Daily Water", "Collaborative Water Monitoring", "Warbon Footprint" and "Ecolution" were spread across three distinct time zones. This level of internationality can be explained by the fact that apart from "Warbon Footprint" who was an already existing team, all other teams were formed during the Gather phase on the online Goodwall platform.

Overall, participants spanned 16 time zones. They originated from 6 different time zones in Asia, 4 time zones in Europe, and 3 in America, and 2 in Africa.

4.4. Activity of teams

In this section, we leverage time stamped Slack posts to describe the activity of the community (including participants, organisation team, and mentors) during the Evaluate phase.

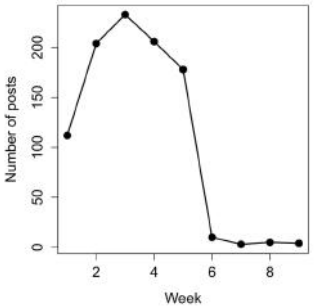

Figure 8: Total number of Slack posts per week.

We observe a sustained activity during the 5 weeks of the Evaluate phase (weeks 1-5 in Figure 8), with around 200 posts per week. The activity drops significantly after the end of the Evaluate phase (week 6). This drop can be attributed to several causes. First, the organizing team used a different Slack for the Accelerate phase. Second, several teams used their own communication tools during the GEAR cycle. Lastly, there was a delay of a few weeks between the Evaluate and Accelerate phase.

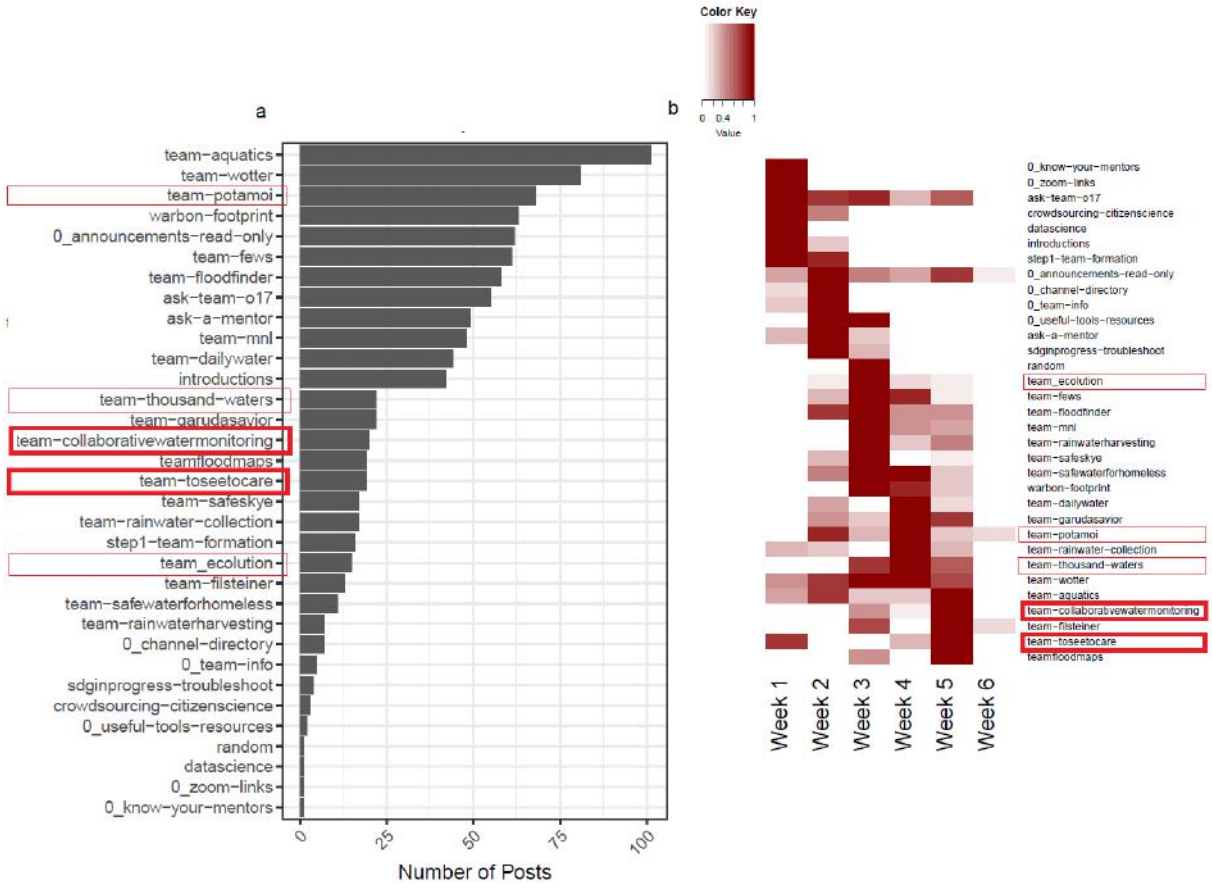

Figure 9: **a** Barplot showing the number of posts per channel. **b** Heatmap showing the temporal distribution of posts across channels. For each channel, we compute the probability that a message is posted a given week (number of posts that week divided by total number of posts in that channel). Colors range from white (low) to red (high). Red rectangles designate the teams selected for the Accelerate phase and bold ones designate the teams selected for the Refine phase.

Team “Aquatics”, “Wotter”, “Potamoi”, “Warbon Footprint” were the most active on Slack during the 5 weeks of the Evaluate phase, with activities ranging from 60 to 105 posts (Figure 9a). The other selected teams for the Accelerate phase showed less engagement, with less than 25 messages posted (Figure 9a). We find that successful teams selected for the later phase of the program show more limited activity on the Slack (red rectangles). We discuss possible explanations at the end of this section.

The channel “announcement-read-only” was the most used by the organizing team and ranked 5th in terms of activity, with 60 messages. The channels “ask-team-o17” and “ask-a-mentor”, devoted to interactions between participants and members of the organisation team, follow closely with 50 posts.

Figure 9b illustrates the activity of each channel over time. First, we see that introduction, team formation, and getting to know mentors were the focus of the first week. The second

week gathered the most activity from organisers who posted resources and content to guide participants, with some early activity across team channels. Finally, we note an intense activity of teams during week three, four and five.

We observe that the two finalist teams "Collaborative Water Monitoring" and "To See to Care" were the most active ones during the last week of the Evaluate phase, as can be seen from normalized post data in Figure 9b. The other teams selected for the Accelerate phase similarly show higher activity at weeks 3 and 4.

These results showcase the limits of the Slack digital traces to interpret team activity. Indeed, there was a larger spectrum of means used by teams to communicate internally, as reported by the organization team. For example, members of the teams "Potamoi" and "To see to care" and "Thousand Waters", who shared the same location, met in person. Two teams, "Collaborative Water Monitoring" and "Ecolution" used Whatsapp to communicate. The organization team reports that the burst of activity during the last week was attributed to the review of the team deliverables prior to their assessment. Overall, the identification of such limits points to the **need of collecting self-report and survey based data to better cross-validate the collected data.**

4.5. Team interactions on Slack during the Evaluate phase

Beyond individual post information, Slack offers the opportunity to identify interactions between individuals, through their reaction to posts of other members. This allows us to gather insights on the quantity of interactions at the intra-team, inter-team, as well as mentors/organizers team levels. To examine the interaction dynamics, we use a network approach, with nodes representing individuals (designated by their role or the name of their team) or teams (aggregated data across team members) linked by their interactions. In the following figures, node size and link thickness are proportional to the number of interactions (at the individual or pair level). The network approach allows us to gather insights on i) who are the most central individuals in the information flow and ii) how much horizontal communication across the community is observed.

In order to reconstruct interaction networks from Slack data, we need to define a unit of interaction. This is challenging, since text-based interactions differ from traditional face-to-face interactions or friendship networks used in social network studies[7]. Here, we used two approaches. First, we reconstructed a "direct interaction network" using user mentions. These correspond to explicit mentions of a user by another user, using an "@" tag, and represent specific calls for help /action (as such mentions notify another user). In addition, we reconstructed a (denser) network of reactions to posts using message replies and emoji reactions. These represent more discreet interactions that signify quick approval (similar to nodding to signal understanding).

We show both networks aggregated at the team level in Figure 10. Links between two nodes indicate the number of interactions between team members of the two teams, and self-loops indicate interactions within teams. Links are directed from sender to receiver. We find that the mention network (Fig 10a) has a strong "star-shaped" centered around the organisation and mentoring team, indicating the importance of the "coaching" component in that phase,

[7] Poquet, O., Tupikina, L. & Santolini, M. Are forum networks social networks? a methodological perspective. in Proceedings of the Tenth International Conference on Learning Analytics & Knowledge 366–375 (Association for Computing Machinery, 2020). doi:10.1145/3375462.3375531.

with few horizontal interactions between teams. The interactions are bidirectional, meaning that teams both ask for and receive help from the coaching team.

When looking at the reaction network, we observe that while mentors and organisers are still central, the network is more distributed, with a higher density of interactions across teams (Figure 10b). This indicates that despite the online nature of the program and the competition between teams, participants were using the Slack space to react to each other's posts.

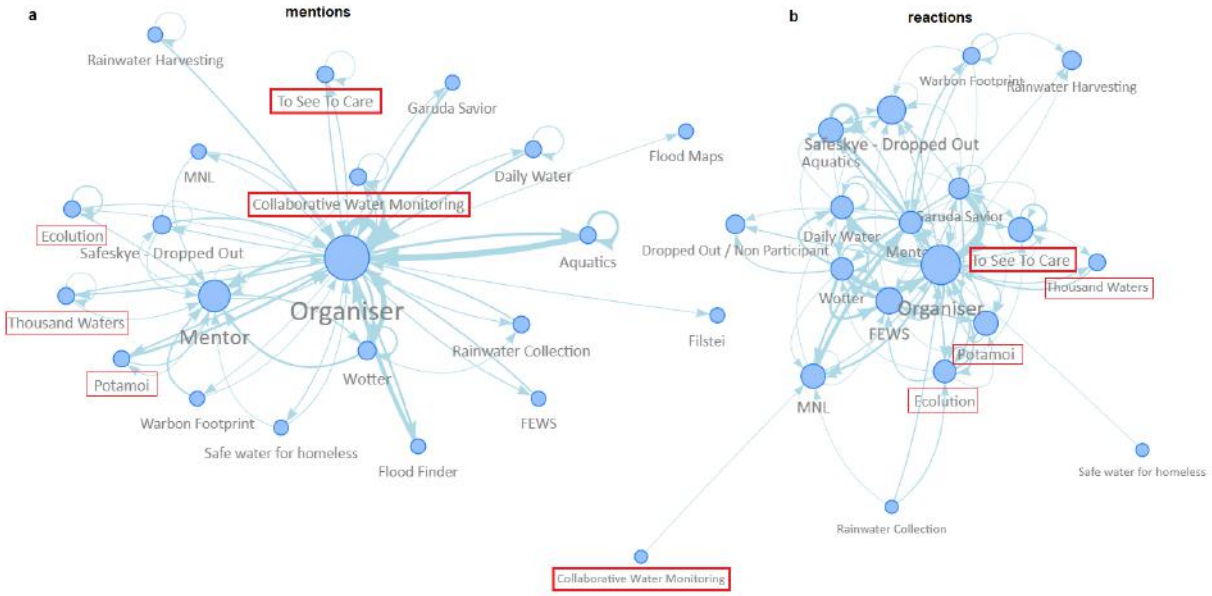

Figure 10: Communication networks from Slack during the Evaluation phase. Nodes represent aggregated data at the team level. **a** Mention network. **b** Reaction network. Teams are linked by weighted edges quantifying the number of times an individual form one team reacts to a post from another team. Red rectangles designate the teams selected for the Accelerate phase and bold ones designate the teams selected for the Refine phase.

We then analyse the interactions within various channels of the Slack, using the reaction network approach. In the following networks, each node represents a participant, with its corresponding team name. We first show channels used during the first (Figure 11) and second (Figure 12) week.

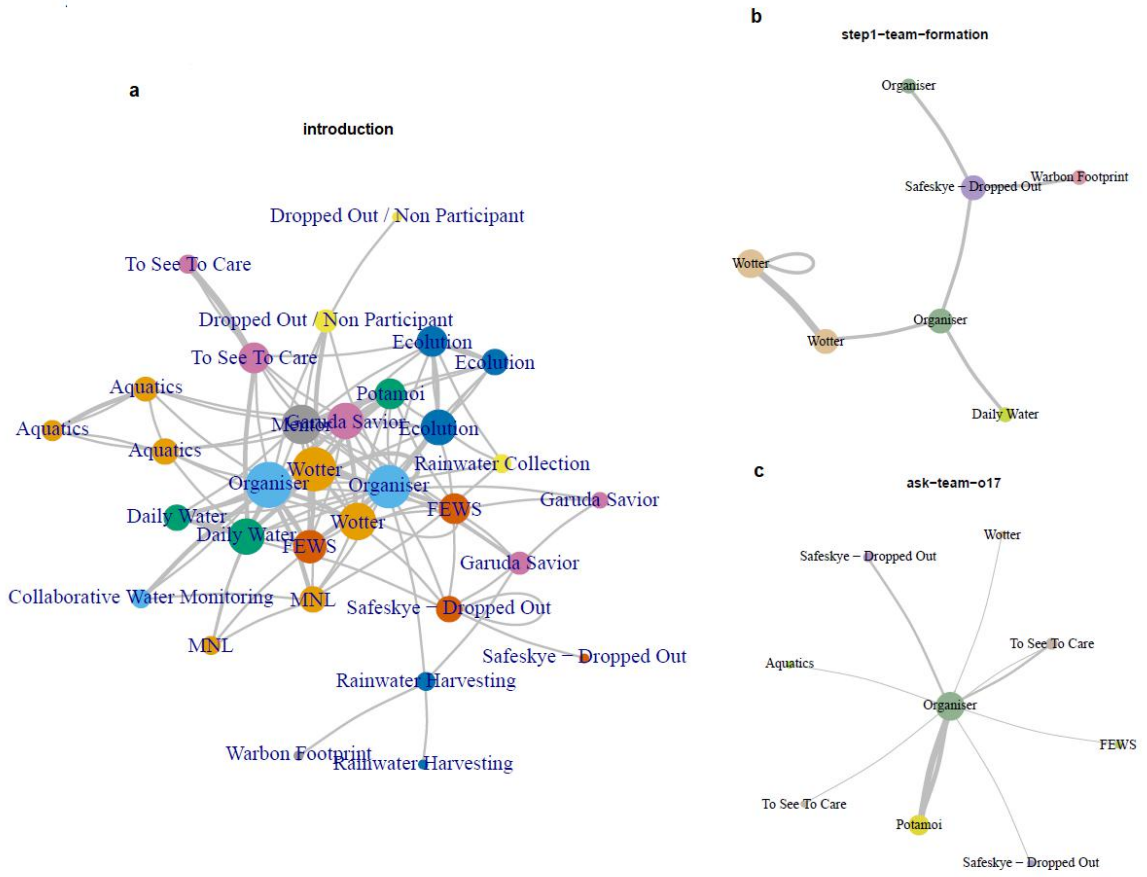

Figure 11: Reaction networks of the channels mostly used during Week 1 of the Evaluation phase (see Figure 9b). **a** Introduction channel. **b** Team formation channel. **c** Ask organizing channel. Individuals are linked by weighted edges quantifying the number of times one individual reacts to a post from the other one. Colors and labels indicate teams.

The “introduction” channel was dedicated to greetings, and invited participants to get to know each other. The high density of this network shows that participants played the game of introductions (Figure 11a). In addition, individuals cluster by team, indicating that participants who interacted more in this channel formed teams, if not formed already. This is more evident when looking at the proximity of team members in the “Aquatics”, “Ecolution” or “To See to Care” teams.

The “step1-team-formation” aimed at enabling participants to find collaborators for their team. They were advised to mention the expertise they are looking for and mention the expertise they are willing to offer. Only four teams interracted in this channel, one of which dropped out (Figure 11b).

The “ask-team-o17” was dedicated to queries and issues on logistics directed to organizers (Figure 11c). Similarly to the “ask-team-o17” channel, the channels mostly used during week 2 were devoted to direct interactions between participants and the organising team (Figure 12).

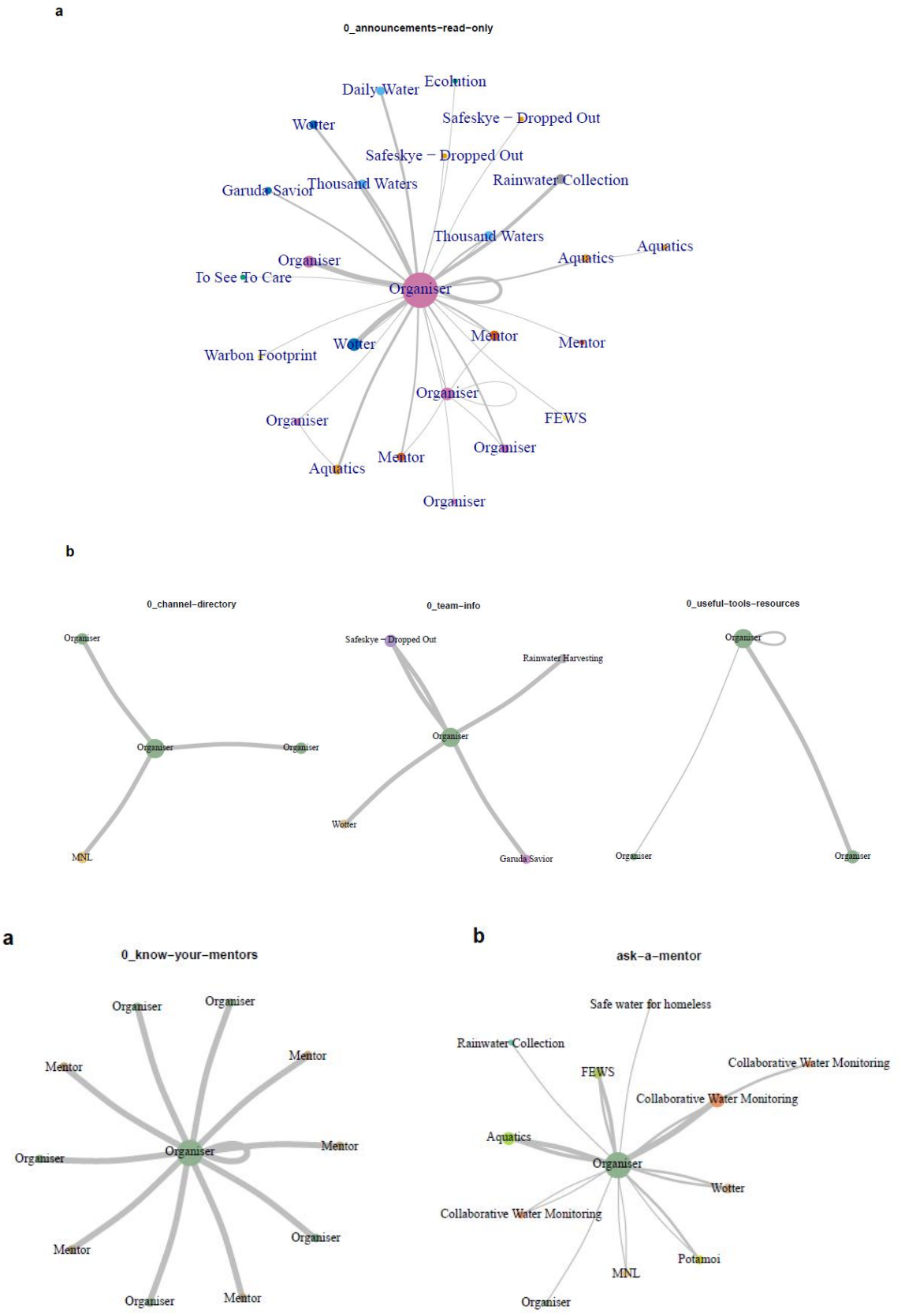

Figure 12: Reaction networks for the most active channels during the week 1 and 2 of the Evaluation phase. Individuals are linked by weighted edges quantifying the number of times one individual reacts to a post from the other one.

Finally, we show in Figure 13 the networks corresponding to the team channels. For most teams, only a few (or one) team members were interacting on the Slack. As such, these networks mostly indicate interactions between the team (represented by the members active on Slack) and the organising and mentorship team, showcasing how individual teams had a personalized coaching during the event. However, when looking at the number of such interactions per team, we find that the successful projects show a limited number of such interactions (Figure 14). This observation could be linked to the limitations exhibited in section 4.4.

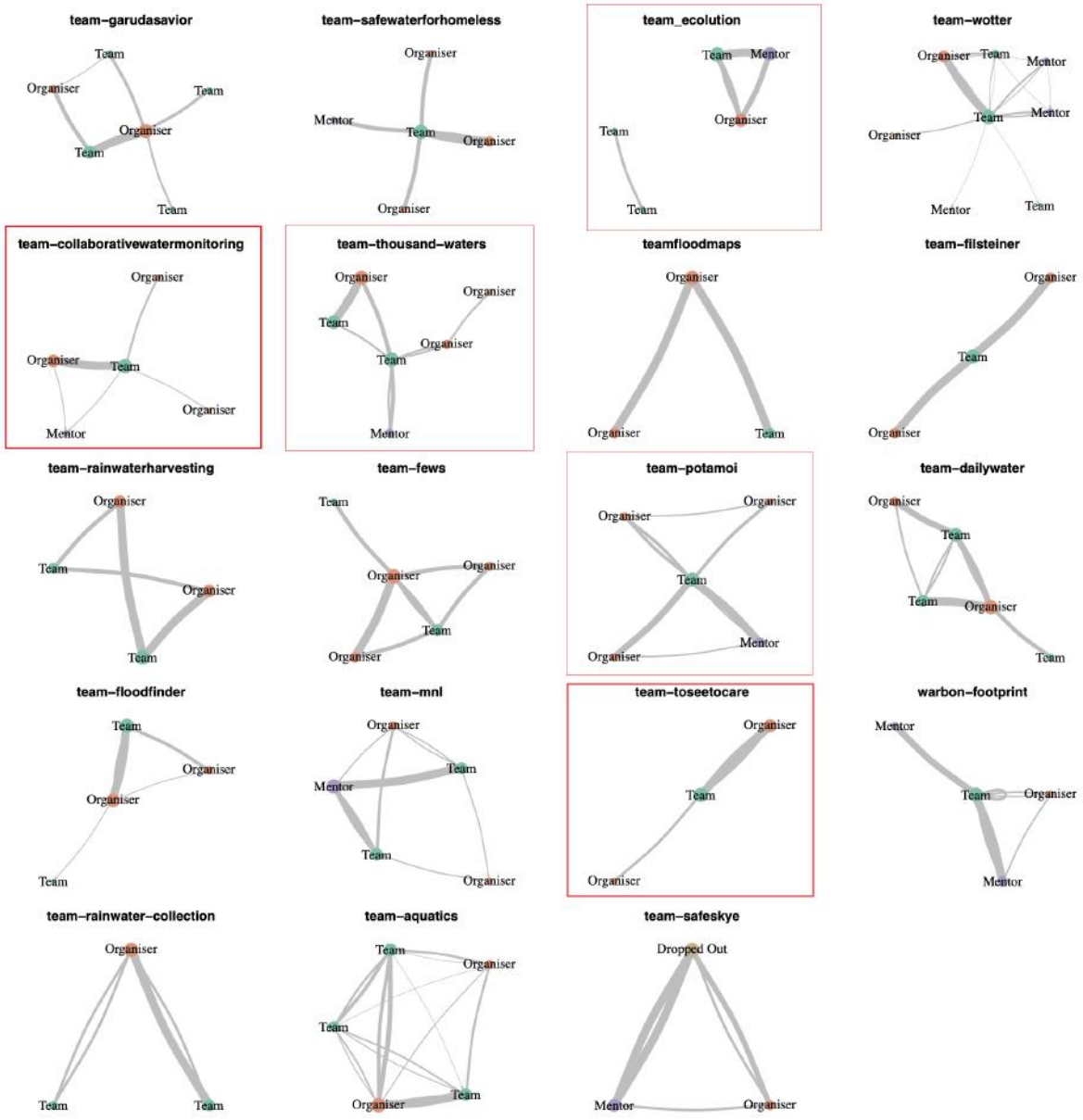

Figure 13: Channel-specific reaction networks. Colors denote team members (green), organisers (orange) and mentors (purple). Red rectangles designate the teams selected for the Accelerate phase and bold ones designate the teams selected for the Refine phase.

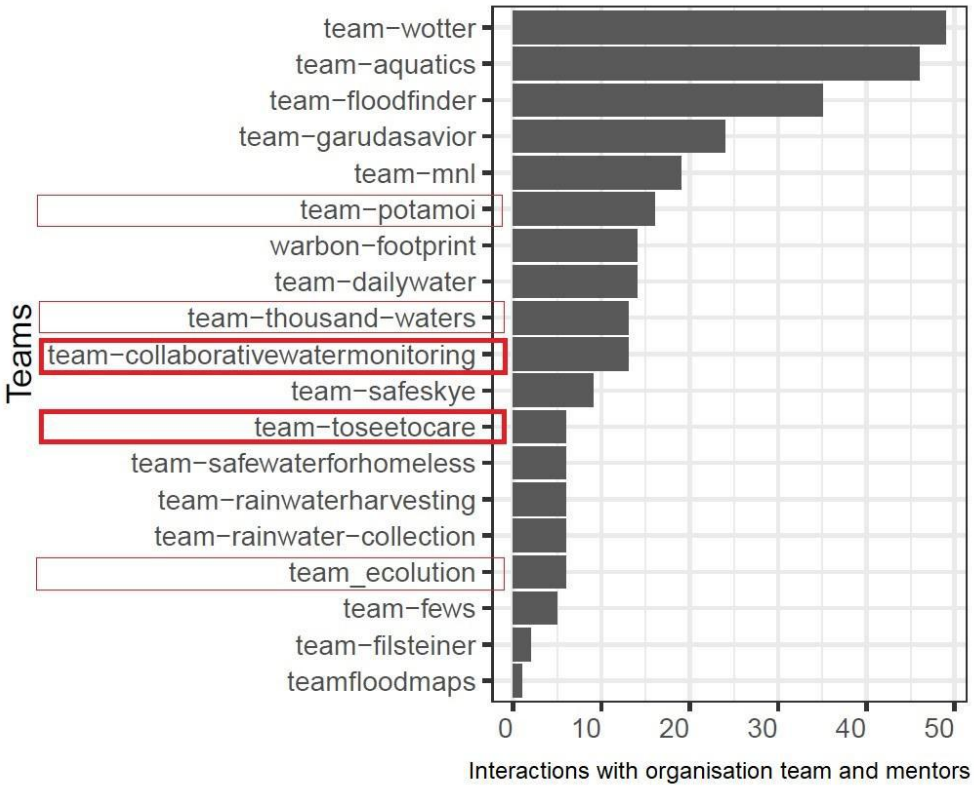

Figure 14: Number of interactions between team members and the organisation team and mentors in their respective channels.

5. Elaboration of a self-report app for collaboration assessment

As shown in the previous sections, we can leverage digital traces to document social interactions and communications during the GEAR cycle. Yet, such traces give a limited qualitative account on the nature of these interactions, requiring additional information gathering methods.

We originally devised T4.3 to build upon an existing app (MindLogger) in order to use smartphone sensors to document on-site team interactions, and trigger Ecological Momentary assessments contextualizing these interactions. In addition, we were aiming to utilize the platform MITeams to provide a web dashboard to visualize the collected data and provide insights into network structure and summary statistics.

This plan had to be adapted due to several unforeseen challenges. First, the development of the MindLogger platform (not part of the consortium) was significantly delayed, making it difficult to assess whether it would be available for the GEAR cycles according to our initial timeline. Similarly, the MITeams platform (not part of the consortium), which was later renamed “OpenTeams”, stopped being maintained in 2020 and proved to be difficult to implement and maintain due to lacking documentation. Finally, novel restrictions following the COVID19 pandemic limited the use of contact-tracing apps using Bluetooth sensors, forcing us to develop alternative solutions for the monitoring of collaborations. Following these setbacks, we adapted our strategy by developing CoSo (Collaborative Sonar) a stand alone platform for the active monitoring and contextualization of team interactions using self-reports and surveys (Figure 15).

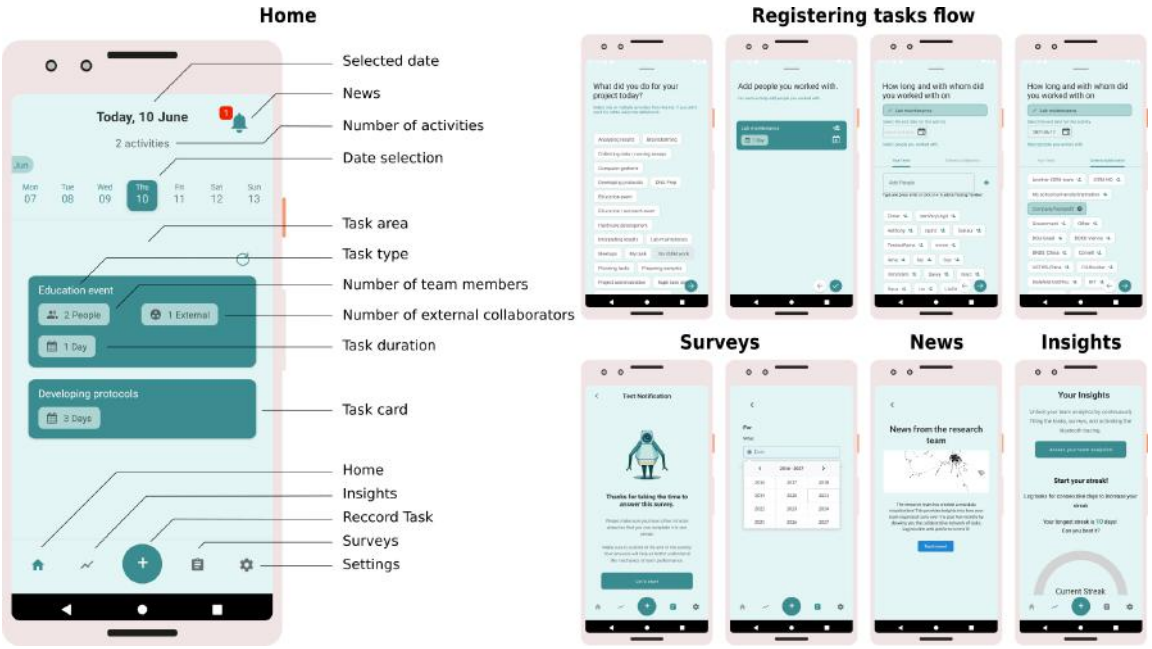

Figure 15: Presentation of the CoSo app and its main flows. The Home screen allows for activity reporting and is the main entry point towards all other screens. The users see the activities registered for a particular day and can edit the information contained within them. On the top is a bell icon displaying the number of unread “News” (communications by the research team). An example is shown on the right. At the bottom are buttons to navigate between the different screens. The first is the home, followed by the Insight tab containing analytics, the Surveys screen allowing the user to answer custom-crafted surveys by the research team, and the settings screen. Finally, prominently located at the center is the button to launch an Activity Registration (or activity recording). This button leads to the activity registration flow displayed on the right. The user is presented with several types of

actions and can select any number of them. Then, for each task, the user is invited to add team collaborators and external collaborators, as well as select an end date for the task.

CoSo is composed of three main applications. First, a backend gathers all user data into a structured database and provides an API, allowing for a simplified management and data extraction from the database of GEAR participants. Second is the core of the platform: a mobile application for Android and iOS allowing users to journal tasks, receive notifications, fill surveys and gather immediate insights about their logged data (Figure 15). This encompasses the main desired features from MindLogger, and provides an alternative pathway for contact-tracing, replacing passive bluetooth interaction information, limited to on-site interactions and yielding quantitative yet non-contextualized insights, by active self-reports on participant interactions, generalized to both on-site and online interactions and contextualized around shared tasks. Finally, a frontend web application allows for data visualization, team management, survey and communication creation by the research team as well as survey filling by users (Figure 16). This encompasses features of MITeams, and will be further elaborated in D4.6 (M33).

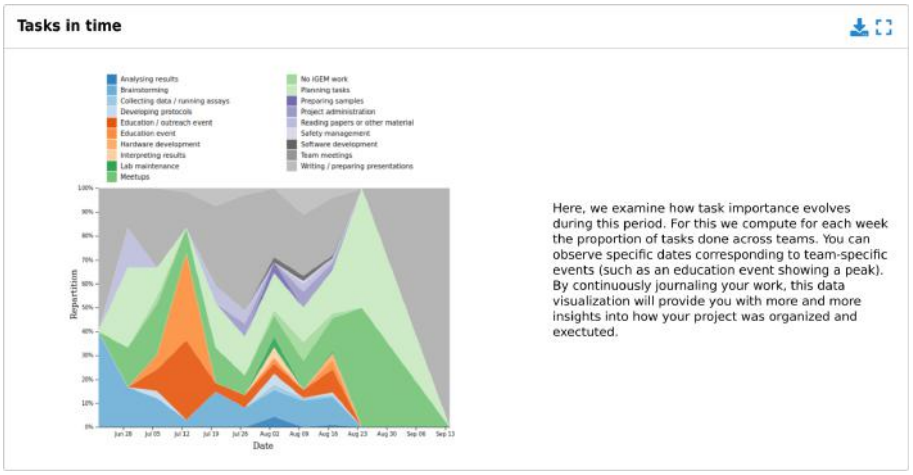

Figure 16: Example data visualization from the CoSo web platform, showing the proportion of self-reports for various tasks in time. These data visualizations are crafted by the research team to provide insights into the team's collaborative work. Each card has a title, a descriptive text, as well as an interactive data visualization.

The CoSo app has been tested in its alpha version in 2020 on a pilot team of 19 users participating in a challenge-based learning event over a period of two months, showing its ability to collect fine-grain temporal data on user interactions, revealing team substructure and task allocation structure (Figure 17). These offer promising insights into how CoSo can be leveraged to provide meta analyses of the interaction dynamics of GEAR participants.

Overall, CoSo allows to conduct the interaction measurements, contextualisation, and visualisations from Task 4.3 and will be deployed for GEARs 2 and 3.

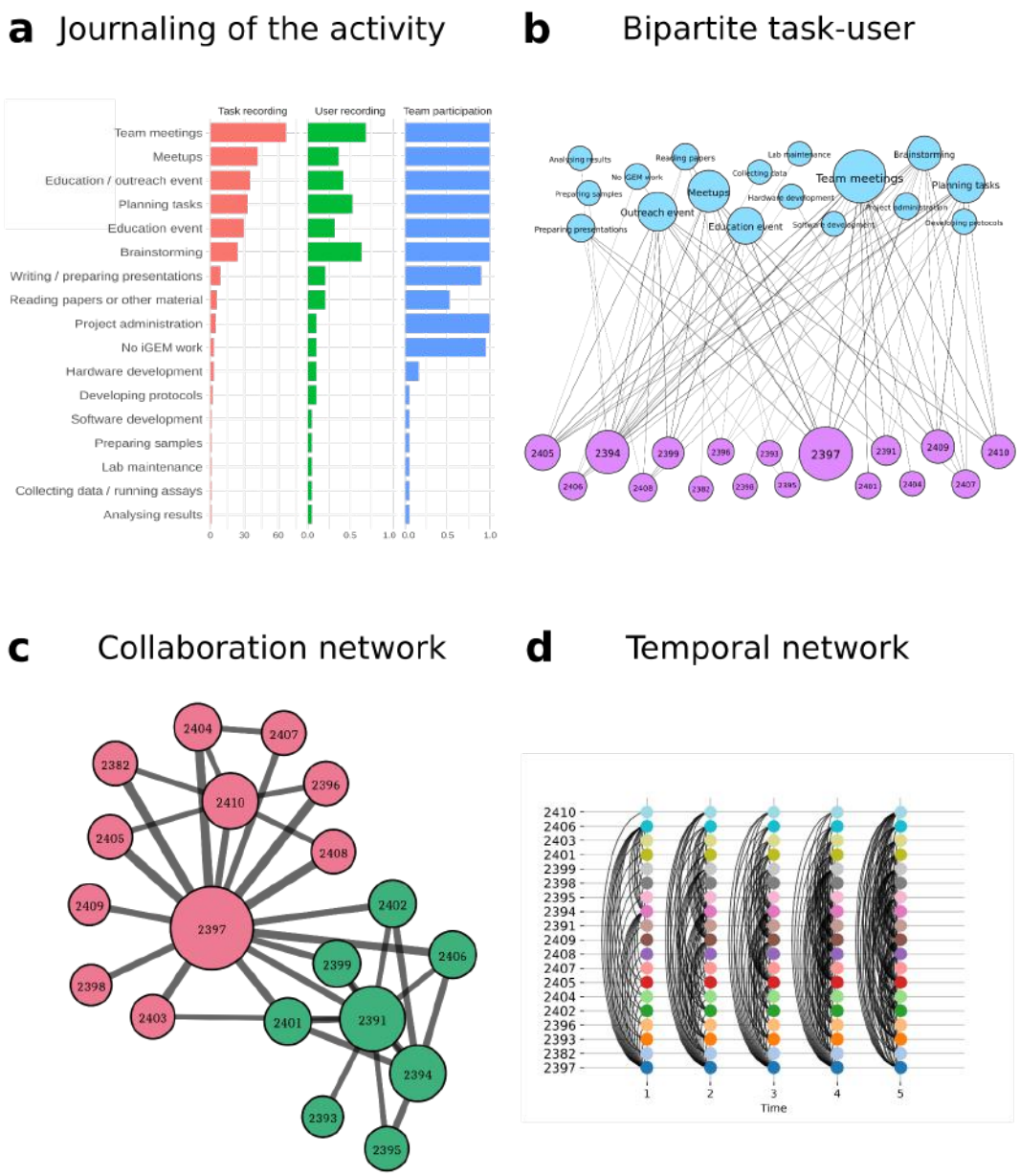

Figure 17: Example data generated with the CoSo app. Example case study from a pilot team during alpha testing. (a) summary statistics of the activity reports: number of times a category was reported, proportion of team members who reported a category, and proportion of members who were mentioned as collaborators in a category. (b) bipartite network between activity categories and users. (c) Collaboration network between users. Links depict collaborations repeated at least 5 times. (d) Temporal network of the team, showing interactions across 5 periods of 12 days each.

6. Discussion and perspectives

The Work Package 4 aims to develop and monitor new metrics and develop statistical models of team engagement and collaboration that contribute to the many-faceted outcomes of the citizen science projects developed within the Crowd4SDG consortium. In this report, we presented a preliminary data-driven approach to describe the GEAR cycle 1, through participant demographic profiles, team diversity, activity and interaction dynamics. In addition, we introduced the CoSo platform for collecting self-reported data on collaborations and task allocation structure of participating teams for both in-situ and online interactions. These findings serve as a basis for i) exhibiting the potentiality of using digital traces on online tools and smartphone apps to derive measures related to team process, ii) highlighting perspectives for measuring metrics and descriptors in the next GEAR cycles and suggest corrective actions. Hence, this work results from the efforts put into the Task 4.2 and 4.3 and will be completed by the Deliverable 4.2.

While our initial hope was to follow the *in-situ* dynamics of teams participating in the program, the unexpected shift to a fully online program led us to shift gears, focusing on digital traces from the team coordination tools, as well as building a smartphone application to facilitate the reporting of collaborative activities. The insights generated by the data collection from the Slack communication channels showed the importance of building a unified framework for collecting an homogenous dataset across all phases of the GEAR cycle, as well as the need to foster participant engagement on a few select platforms.

By leveraging a single sign-in system and the ability to run all future surveys in one place using the CoSo app, we expect to collect comprehensive profile information about participants along with temporal insights on their collaborative activities, avoiding extensive manual curation from the organizing team and ensuring analysis-ready data.

We showcased how messages from the discussion application Slack can quantify individual and team interaction and engagement. We found that teams were defined early on, and a significant group cohesion was observed in terms of inter-team reactions. Interestingly, we found that successful teams were relatively autonomous (less interactions with mentors). Further conclusions on the success of teams selected for the Refine phase can be drawn if other Slack workspaces are used during the Accelerate phase, highlighting other traces of mentorship. This showcases the importance of future homogenous communication data collection throughout the phases.

Several actions will be collaboratively carried out with consortium partners to improve the data collection process, the elaboration and the implementation of new metrics to assess both the process and the outcomes of teams participating in the Crowd4SDG project for the next GEAR cycles.

First, we will work with consortium partners to unify the tools used by the participants during the various phases and discuss their mandatory status in order to access consistent, homogenous data for the monitoring and draw a more complete picture of the full dynamics.

Second, we will use the stand-alone CoSo platform to complement the digital traces with surveys and temporal self-reported data related to team interactions and task allocation structure. By working hand in hand with consortium partners on the elaboration of these surveys we will ensure that the data analysis serves both our research purposes and the overall improvement of the GEAR cycles.

Third, we will use the epistemological report findings (D4.1) to shed light on the assessment practices performed in the GEAR cycle. As a result the new metrics and descriptors will be discussed among the consortium to refine the current project assessment grid and the project assessment process for the GEAR cycle 2 and be presented in the dedicated deliverable at month 24. Decisions made based on the evidence provided by this WP will be implemented for the next GEAR cycles.

7. Conclusion

We provided here an informative account of the implementation of the GEAR cycle, that showcases how digital data sources can provide a “meta” picture of the projects.

We showcased the gender balance and international nature of the projects. Further information such as the level of education of the participants was asked in the evaluate phase satisfaction survey, and was presented in the Deliverable 3.3. We see an opportunity to collect further participant information by either extracting raw data from Goodwall or by systematically surveying participants at different stages of the GEAR cycle. We also highlighted how posting activity on Slack could inform on i) the formation and engagement of teams and ii) the interactions of team members, both within participants and with mentors and organizers. Finally, we developed a stand-alone interaction data collection and visualisation platform, CoSo, enabling a comprehensive temporal study on team collaboration dynamics that will be used in upcoming GEAR cycles.

We conclude that such a quantitative description of the diversity of participants and teams, and the monitoring of their activity via the use of digital traces and self-reports will allow the Crowd4SDG consortium to take evidence-based decisions for the next GEAR cycles and frame the improvement of citizen science skills and outcomes of the developed projects. The tools employed within the Crowd4SDG project complemented by self-reported and surveyed data have the potential to generate even more detailed digital traces, offering an opportunity to operationalize metrics and descriptors underlying the originality, effectiveness, sustainability/robustness and adaptation/appropriateness of citizen science projects. The WP4 will continue to work around these metrics, descriptors and their elaboration process by including the epistemological considerations of the Deliverable 4.1 on diversity, inclusion and deliberation. Our contribution extends beyond the Crowd4SDG project to the general evaluation of citizen science by informing project leaders, citizen scientists and decision makers on what can be assessed online to perform high-quality citizen science.

Annex: list of abbreviations

Abbreviation	Description
AI	Artificial Intelligence
CBI	Challenge-based Innovation (in-person coaching)
CBIx	Challenge-based Innovation (remote location)
CoSo	Collaborative Sonar
CS	Citizen Science
CSSK	Citizen Science Solution Kit
GEAR	Gather, Evaluate, Accelerate, Refine
NSO	National Statistical Office
O17	Open Seventeen Challenge (online coaching)
SDG	Sustainable Development Goal
WP	Work Package