

FOSTERING PRE-UNIVERSITY STUDENT PARTICIPATION IN OSGeo THROUGH THE GOOGLE CODE-IN COMPETITION

V. Rautenbach^{a,*}, M. Di Leo^b, V. Andreo^c, L. Delucchi^d, H. Kudrnovsky^e, J. McKenna^e, S. Acosta y Lara^f

^a Centre for Geoinformation Science, Department of Geography, Geoinformatics and Meteorology, University of Pretoria, South Africa - victoria.rautenbach@up.ac.za

^b European Commission, Joint Research Centre, Italy - margherita.di-leo@ec.europa.eu

^c ITC - Faculty of Geo-Information Science and Earth Observation, University of Twente, The Netherlands - veronica.andreo@utwente.nl

^d Edmund Mach Foundation, Research and Innovation Centre, Italy - luca.delucchi@fmach.it

^e OSGeo - h.kudrnovsky@gmail.com and jmkenna@osgeo.org

^f Comunidad gvSIG Uruguay - sergio.acostaylara@mtop.gub.uy

Commission IV, WG IV/4

KEY WORDS: Education, Outreach, Pre-university students, Open Source

ABSTRACT:

The Open Source Geospatial Foundation's (OSGeo) vision is to empower everyone, from pre-university students to professionals, with open source geospatial applications, tools and resources. In 2017, OSGeo decided to participate for the first time in the Code-in competition. Google Code-in (GCI) is an annual online competition aimed at introducing pre-university students (13-17 years) to open source projects, development and communities, through short 3-5 hour tasks. This is a unique opportunity to interact with pre-university students and to encourage them to become part of OSGeo. In this paper, we present OSGeo's involvement in GCI with the purpose of establishing lessons learned to improve our approach in the next editions of GCI. Over the 51 days of the competition, 279 students completed 649 OSGeo tasks. Students consistently communicated with the mentors to discuss submission and receive inputs from the wide community of developers too. During the GCI, the mentors reviewed the students' work and provided suggestions and feedback. Generally, the submissions were good and some of them are now part of the projects. As this was our first time participating in GCI these issues are seen as lessons learned and strategies to improve the process will be implemented based on the mentors' experience. It is key to encourage these students to continue contributing to the OSGeo community, as they will bring new energy and ideas into the organisation; for many of these young students, this competition is a way to introduce them to the geospatial industry.

1. INTRODUCTION

The Open Source Geospatial Foundation (OSGeo) was founded as a non-profit organization in 2006 and the number of open source software projects under its umbrella is steadily growing; the term 'open source' applies to software that is both freely distributed, and its source code is shared. The current OSGeo projects include content management systems, desktop applications, geospatial libraries, metadata catalogues, spatial databases, and web mapping. OSGeo's vision is to empower everyone, from pre-university students to professionals, with open source geospatial applications, tools and resources (OSGeo 2017). To further OSGeo's commitment to open education, the GeoForAll initiative was established in 2011 through a partnership between OSGeo and the International Cartographic Association (ICA). The importance of educational outreach and open source for the larger geospatial community was emphasized when the International Society for Photogrammetry and Remote Sensing (ISPRS), International Geographical Union (IGU), Association of Geographic Information Laboratories in Europe (AGILE), and the University Consortium for Geographic Information Science (UCGIS) joined this memorandum of understanding. At present, GeoForAll consists of 124 labs, mainly based at universities and research center world-wide. Even though there are various outreach activities at the university

level, and not only through OSGeo, the majority of open source developers are between 30-49 years (Choi and Pruet 2015). This suggests that more effort is required to engage with the younger (below 29) geospatial community and encourage their active participation.

Google has two programmes to introduce pre-university and university students to open source, namely Google Code-in (GCI) and Google Summer of Code (GSoc), respectively. GSoc was first established by Google in 2005 and has grown ever since. GSoc is an online, international program targeted to university students, that aims at fostering their participation in open source software communities. Mentoring organizations select students that will be developing software applications during 12 weeks and receiving support and feedback from mentors within the software community. Successful students are paid stipends by Google. The program aims at identifying and bringing new developers into open source software communities, as well as exposing students to real world software development. OSGeo is a veteran organization having participated in GSoc and having graduated 180 (at 2017) students from all over the world every year since 2007.

In 2017, OSGeo decided to participate in Google Code-in (GCI) for the first time. GCI is an annual online competition aimed at

* Corresponding author

introducing pre-university students (13-17 years) to open source projects, development and communities, through short 3-5 hour tasks. As opposed to GSoC, in GCI students are not selected by the organizations, but freely pick up tasks from one or more mentoring organizations and complete them. Students qualify for different prizes (i.e. certificate, t-shirts, hoodies and the grand prize of visiting Google's main headquarters in San Francisco) depending on the number of tasks they complete. During GCI, participating organizations have a unique opportunity to interact with pre-university students and to encourage them to become part of their respective organizations. Thus, OSGeo's ultimate goal is to encourage and inspire the students to become actively involved in OSGeo after the GCI contest has ended.

In this paper, we report on our experience participating in the 2017 GCI contest and the lessons learned to improve our approach in the next editions of GCI. The remainder of the paper is structured as follows: Section 2 provides an overview of how GCI is structured; in Section 3 we briefly discuss the method followed; an overview of the OSGeo's involvement in GCI and report on the experiences of the mentors is presented in Section 4; and lastly in Section 5 the overall results and observations are discussed, and conclusions are provided.

2. GOOGLE CODE-IN

Google Code-in (GCI) is an online, international competition aimed at introducing pre-university students (13-17 years) to open source software development (Google 2017) and communities. For most of the students, GCI is their first experience with open source and thus, the competition follows a strict structure to gently introduce them to the open source world. The GCI competition generally runs over a period of seven weeks around the beginning of the calendar year. Once Google announces the program every year, organisations apply to participate in GCI and if selected, they need to create numerous tasks. The tasks should take approximately 3-5 hours to complete and they can represent different levels of experience and difficulty (i.e. from beginner to advanced). The task description also includes the mentor(s) responsible, the type of task (i.e. coding, documentation, training, outreach, research, quality assurance and user interface), links to relevant information, maximum amount of time the task can take to be completed (e.g. 3 to 7 days) and the number of instances available. The number of instances available for each task represents the number of times a certain task can be claimed by students. For their nature, some of the tasks can only have 1 instance (for example, a bug fix, once it is fixed, doesn't require another student working on it), whereas some other tasks can entail multiple instances (for example, designing a t-shirt for a code sprint event). Students can then select tasks from the organization's list, however, they can only claim and work on one task at a time. Only when the task has been approved by the mentor or abandoned, the student can claim another task.

Once a task is submitted for review, the mentor(s) review the work submitted and can either approve it or request more work, providing comments to improve the submission. Mentors have 36 hours to review a submitted task, but they are encouraged to provide feedback to students within 12 hours, because a delay in providing feedback can impair the student's performance in the competition. Students win prizes based on the number of tasks completed and the quality of their submissions. The prizes range from a digital certificate or t-shirt to a grand prize trip to Google headquarters in California, United States of America (USA).

Overall, the 2017 edition of GCI had 3,555 participating students from 78 countries completing 16,468 tasks with a record of 25 open source participating organizations (Google 2018). This was a record number of students and it represented a 265% increase in participation as compared to 2016. Unsurprisingly, almost half (47.8%) of the students are from India and a quarter (25.4%) from the USA. The southern hemisphere is under represented, probably due to GCI taking place during the summer vacation in most of these countries.

For 91% of the students, the 2017 edition was their first time competing in GCI. However, disappointingly only 17% of participants were girls. On average, most of the students were between 15-17 years old. Two thirds of the students completed three or more tasks and they earned a t-shirt. Refer to Section 4 for details on OSGeo's participation in GCI.

3. METHOD

In this paper, we present OSGeo's involvement in GCI with the purpose of establishing lessons learned to improve our approach in the next editions of GCI. To achieve this, we analysed the student submissions and collected feedback from the mentors.

Once the competition finished, we downloaded the data from all OSGeo tasks. These datasets include tasks designed and offered by the organization and instances of those tasks that had some activity (i.e., claimed, completed, abandoned). Each instance contains information, such as date the task was claimed, interactions among student and mentor, submissions, and date task was approved. Basic descriptive statistics (e.g., percentage of tasks completed, abandoned or out of time, answer time by students and mentors, days to complete different type of tasks, number of tasks completed per student, number of tasks per project, number of tasks with which mentors interacted) were estimated from the instances data and plotted. The script used for this aim is available at: https://git.osgeo.org/gitea/lucadelu/gci_analyst.

The OSGeo administrators and mentors were invited to participate in a short feedback survey to collect information on: percentage of material integrated into the various projects, the number of hours spent mentoring, if any students are still actively participating in the community, and whether they would consider mentoring in the next GCI edition. Additionally, all the co-authors of the paper served as either an administrator or mentor during the 2017 GCI edition. Thus, all co-authors shared their thoughts and experience, and this was summarised in Section 4.2.

4. OSGEO'S EXPERIENCE DURING GCI

4.1. Overview of OSGeo's participation in GCI

During GCI 2017, OSGeo's team entailed 20 members from the OSGeo community (i.e., 4 admins - acting also as mentors in some cases - and 16 mentors) that created 176 tasks for GeoForAll & OSGeo, and involved 11 software projects (i.e., FOSS4G, GeoServer, GeoTools, GRASS GIS, gvSIG, MapServer, OpenLayers, OSGeoLive, pgRouting, PostGIS, and QGIS). Students consistently communicated through the GCI dashboard, IRC (Internet Relay Chat), and mailing lists with the mentors to discuss submission and receive inputs from the wide community of developers too.

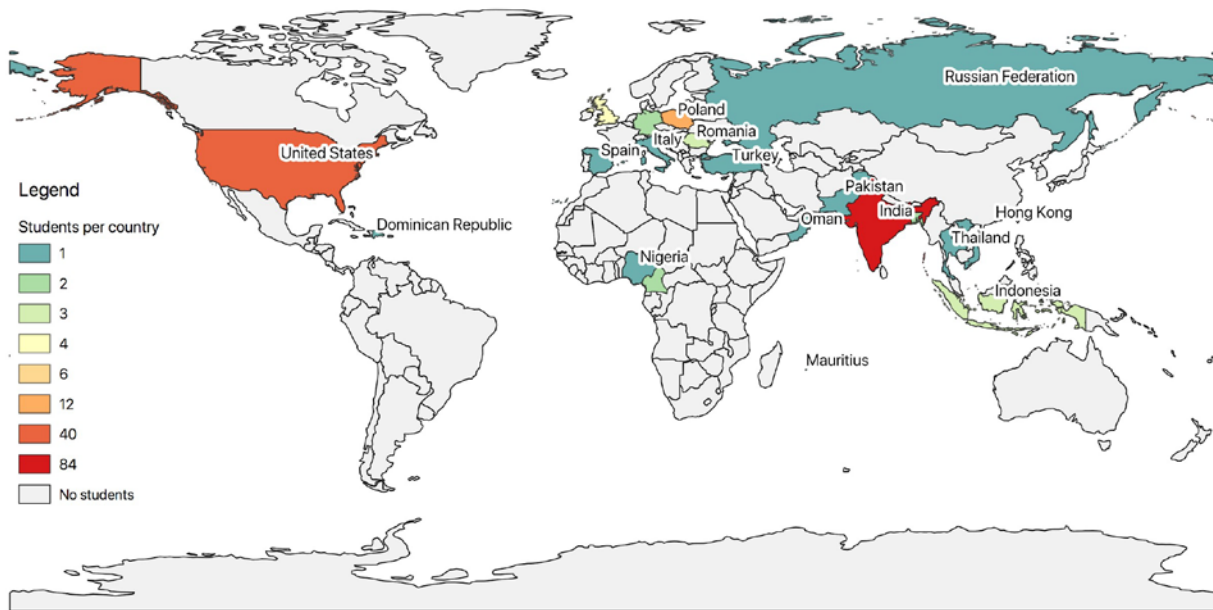


Figure 1. Overview of countries of origin of OSGeo students

Based on the data extracted from the GCI dashboard, 542 students selected an OSGeo task. The majority of the students were from India (49%) followed by the United States (24%), Poland (7%), Singapore (4%) and 18 other countries (refer to Figure 1). The distribution is based on a sample of the students that completed a task requiring them to add themselves to the OSGeo member map that 170 students completed. It should be noted that an Italian student participated in the GCI, even though Italian students were not allowed to enter. The reason for this is not known to the authors.

In total, the students completed 649 tasks (this includes multiple instances of the same task) but 207 tasks were abandoned, and an additional 106 tasks ran out-of-time. Most of the students completed only one task, while the two grand prize winners for OSGeo ended up with 72 and 44 completed tasks respectively across different projects (refer to Figure 2). In general, students mostly selected outreach and research tasks (52%) with documentation and training category in the second place (26%). Coding was only in the fourth place with 8% (see Figure 3).

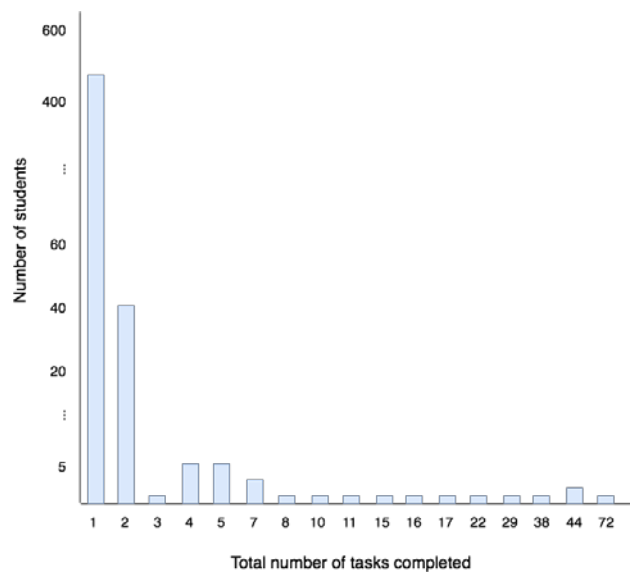


Figure 2. Overview of number of OSGeo tasks completed by students.

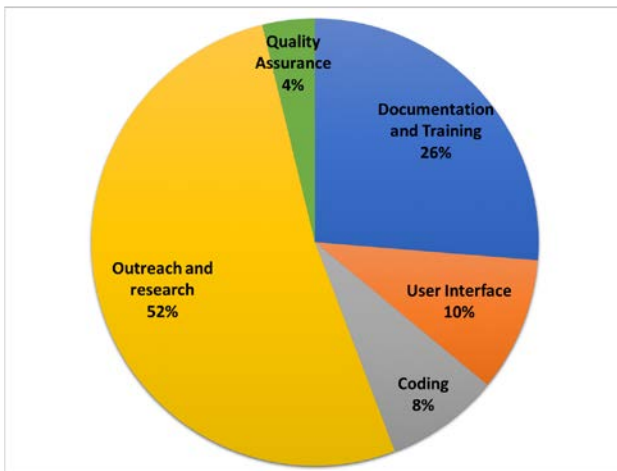


Figure 3. Type of OSGeo tasks completed

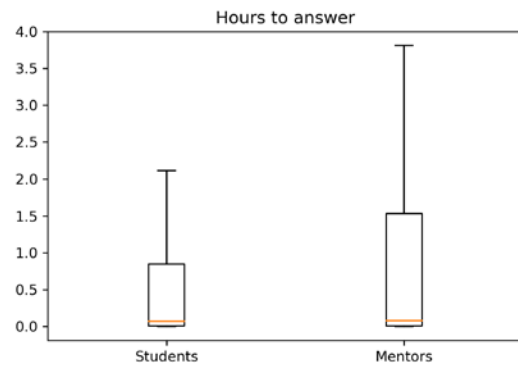


Figure 4. Boxplot showing the response time for both OSGeo mentors and students

Mentors on average took slightly longer to respond than the students, see Figure 4. This can be attributed to the fact that mentors also had their normal work responsibilities. Additionally, the mentors were located in only certain time zones and this resulted in day/night challenges.

OSGeo and pgRouting had the highest number of completed tasks, followed by OSGeo-Live and GRASS GIS. GRASS GIS also had the highest number of abandoned and out of time tasks (refer to Figure 5).

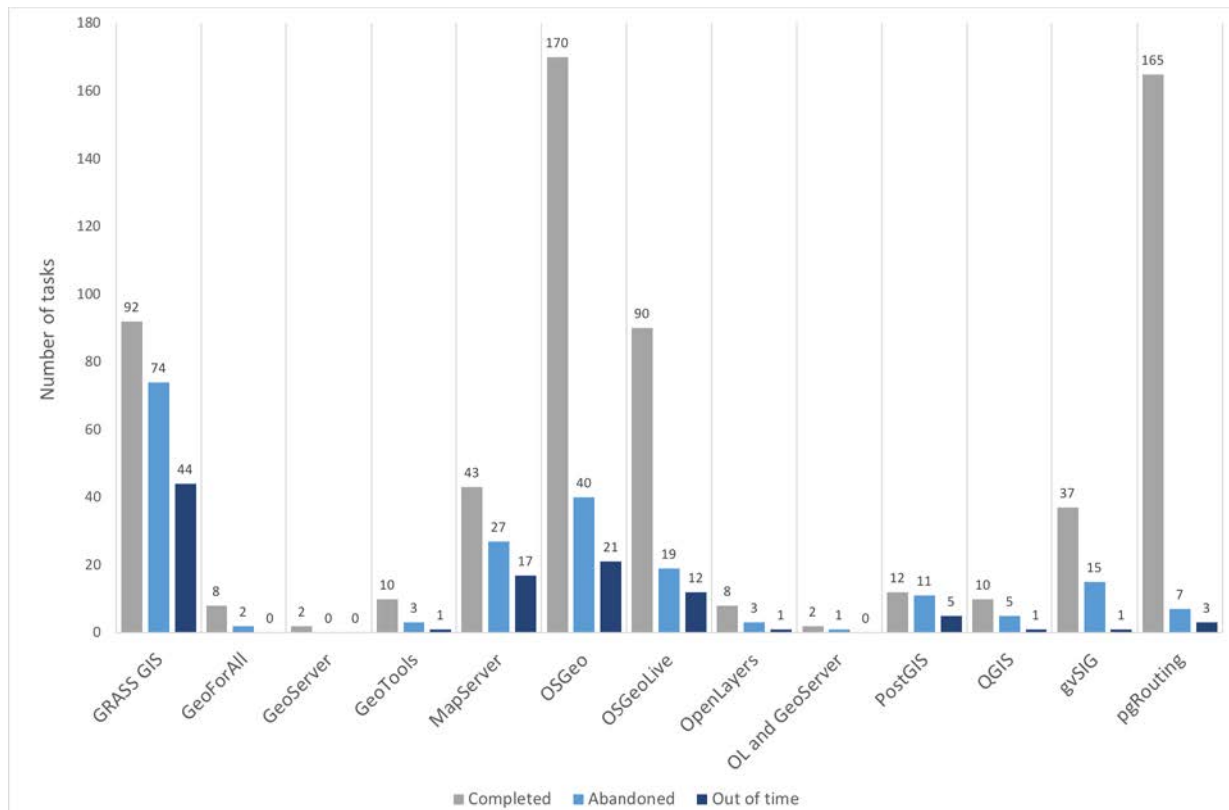


Figure 5. Number of tasks per OSGeo project

4.2. Mentor feedback and experience

OSGeo’s first participation in GCI was genuinely driven by curiosity and the enthusiasm to interact with a young generation of students. At the start of the competition, many questions came up, such as, “*What to expect from such young students?*”, “*Are they capable to contribute something worthwhile for the project?*”. We adjusted our tasks and expectations throughout the competition, but once the GCI was complete, we circulated a short feedback survey among the mentors, to gather their

impressions and whether at the end of the day, the result was positive and the effort worthwhile. The survey covered the following questions: 1) were the materials produced integrated into the various projects, 2) average time spent mentoring, 3) the continued (active) participation of the students after GCI, and 4) was it worthwhile for your project to participate in GCI.

OSGeo created various tasks (i.e., coding, documentation, training, outreach, research, quality assurance and user interface)

with the hope that some of the output created by the students could potentially be **integrated into the respective projects**. The results have been quite satisfactory, and relevant parts of the produced output was integrated into the projects. This includes, code, documentation, unit tests, tutorials, examples in manual pages, to name a few. One example related to GRASS GIS, during GCI the documentation of 12 modules was improved with examples and/or figures. Moreover, tests for the test suite of 11 GRASS GIS modules were implemented by students and added to the source code. However, in other cases, further work is required before it can be integrated into the project. Referring to Figure 3, the majority of the output were for outreach activities in the form of blog posts, and designs for t-shirts and logos that cannot directly be integrated into the various projects.

The next important aspect to investigate was the **average number of hours mentors spent per week**. Mentors indicated that they spent a few hours up to peaks of 20 or 30 hours per week. This parameter, however, is strictly dependent on the popularity of a certain task respect to others as well as mentor's time availability. In fact, if a task is very popular and allows several instances, many students will claim it and the amount of time required for the evaluation will increase dramatically.

Lastly, OSGeo participated in GCI 2017 with the intention to encourage the pre-university students to become **active participants** in the OSGeo community and its various projects. The mentors were thus asked if any of the students decided to keep contributing after the competition. In only very rare occurrences did a student continue contributing to an OSGeo project. Therefore, that result might appear as a paradox to the next question, whether it was **worthwhile for their project to participate** in GCI, but actually all OSGeo mentors replied "Yes". The mentors even added that it was worthwhile experience even though it took a lot of time and effort and took place over the Christmas holiday.

You might ask, if GCI is not fostering pre-university students' participation in OSGeo, why do the mentors consider the GCI contest to be worthwhile? Firstly, the human aspect in interaction with enthusiastic young students. It is fulfilling to help eager young students to understand, learn and do their best. On the other hand, the students provided a fresh perspective on many OSGeo projects. The aim of the project is suddenly shifted, the user, smart and quick-witted, wants to obtain the result quickly and without a previous, extensive knowledge of the software, and, especially in the case of GIS and scientific software, this could be deranging for the developer's viewpoint. But during GCI everything happens quickly, and everyone needs to be pragmatic, and in most cases this "destabilization" resulted into positive, quick actions: enhancing the software project's documentation, and improving the user interface with a fresh perspective that will also benefit the average user in the OSGeo community.

For some OSGeo software projects, the participation of the pre-university students to the open source community brought a sense of activity that sometimes feels like the spring after the winter, raising curiosity and posing new questions and challenges. It is a great opportunity to be able to see what we know, with fresh eyes and a new perspective. Considering this, the response to the last question is then not so unsurprising.

The OSGeo mentors also identified some problems and undesirable student's behaviours. One of these problems was plagiarism, i.e., students submitting non-original material. This is completely against the GCI contest rules that are signed when

entering, however, due to their age and lack of experience the students did not have a clear understanding of the concepts (e.g., licensing and intellectual property). Moreover, a couple of students' behaviour could be defined as non-collaborative, meaning their objective was not to learn, but rather to complete the most number of tasks without too much effort. A student would choose a task that they believe would be simple to complete, but when receiving feedback, he/she would be inclined to not follow the suggestions and rather abandon the task. Some other students seek immediate feedback when submitting a task, as the review time kept the student from starting a new task (mentors have up to 24 hours to review student's tasks). Lastly, the tasks descriptions were not specific enough in some cases, and students would use proprietary software they are familiar with instead of an open source alternative. For example, students tend to use proprietary design software they are familiar with for logo design tasks instead of learning open source alternatives, such as Inkscape or Gimp.

4.3. Lessons learned

As the 2017 GCI was our first experience with this type of competition, we needed to make adjustments to the tasks and expectations throughout the competition. Below are some lessons learned and strategies identified to improve the process for the 2018 edition of GCI:

i) Managing the mentors' workload

As mentioned, some mentors spent up to 30 hours a week mentoring, and this can become quite overwhelming. However, there are some methods to manage the workload. If there is a specific task that is attracting numerous students, an option is to change the number of instances to 0, in order to put a certain task on hold. This will allow the mentors to create a similar task with slightly different (or more difficult) requirements.

ii) Clear and well documented directives and criteria for each task

To enable the admins and additional mentors to assist with mentoring and evaluating popular tasks, clear notes and directives are required. This allows fellows mentors and admins to step in and help with the evaluation, and it is also particularly useful when shifts are needed during holidays.

iii) Detailed task descriptions

A clearly defined task also reduces the amount of time required to evaluate the submissions, as the students are clear on what is expected and where to find additional resources. The mentors should keep in mind what seems obvious for them, might not be for the students and the additional details in the task description might assist students to overcome unnecessary barriers.

This lesson can be implemented on a larger scale when it comes to writing users documentation or designing a graphical user interface (GUI). In some cases, questions posed by the students helped developers realize what could be a potential barrier for a first time user in the GUI or when using a certain function.

iv) Following up with students when they abandon a task or are about to run out of time

Students are often shy and do not ask for assistance when they encounter a barrier, especially when it is their first task. Thus, when they struggle, the students tend to abandon the task or give up until they run out of time. Once a student abandons a task or there is only a couple of hours left for the task, the mentor should send the student a message via the GCI Dashboard to follow-up

and ask if he/she has any questions. This might encourage the student to ask questions and complete the task.

v) Preventing plagiarism

Plagiarism is against the competition rules and if a student is caught plagiarizing they are immediately disqualified from the competition. In some cases, the plagiarism is due to a lack of experience and mentors should require submissions to be of such a nature as to deter the students from easily copying work or code from the internet. For example, asking the student to submit an additional screenshot of the logo designed with the terminal open displaying the student's name.

5. CONCLUSION

In this paper, we reported on our experience participating in the 2017 GCI contest and the lessons learned to improve our approach in the next editions of GCI. We provided an overview of the students that participated in the competition and the OSGeo tasks completed, and also summarised the experience and feedback of the OSGeo administrators and mentors. We encountered a number of non-desirable and difficult to deal with issues, such as plagiarism, managing the mentors workload, a non-collaborative attitude of some students, and seeking immediate feedback. As this was our first time participating in GCI these issues are seen as lessons learned and strategies to improve the process.

Newcomers often find it challenging to get involved in open source communities. Steinmacher et al. (2015) performed a systematic review and identified various barriers faced by, such as technical hurdles, too much or unclear documentation, previous knowledge of the newcomer and that it is difficult to find an appropriate task to start with. Google acknowledges the importance of open source software and open source communities, and to promote active participation in open source development they started the Google Code-in and Google Summer of Code programmes. These programmes provide pre-university and university students with a unique opportunity to get involved in open source and overcome these potential barriers newcomers face.

It is key to encourage these students to continue to contribute to the OSGeo community (the winner of the competition asked for and submitted more work even after GCI has finished!), as they will bring new energy and ideas into the organization; for many of these young students, this competition is a way to introduce them to the now 400 billion USD geospatial industry (2017). For the students, the exposure to coding and open source will be beneficial if they intend to enrol for tertiary education, especially in computing (Hagan and Markham 2000). Lastly, the mentors' experience during the 2017 Google Code-in could contribute to the outreach plan of OSGeo and provide guidelines on how to encourage students and young professionals to get involved and contribute to OSGeo.

6. ACKNOWLEDGEMENTS

The authors would like to thank all the students, mentors, administrators and OSGeo who have participated in the 2017 Google Code-in described in the paper. We would also like to thank Google and the Google Code-in administrators for hosting and coordinating the competition.

7. REFERENCES

- AlphaBeta, 2017. *The Economic Impact of Geospatial Services, Value of the Web*, <https://www.valueoftheweb.com/reports/the-economic-impact-of-geospatial-services/> (7 February 2018)
- Choi, N., and Pruett, J.A., 2015. The characteristics and motivations of library open source software developers: An empirical study. *Library and Information Science Research*, 37(2), pp. 109-117.
- Google, 2017. 2017 Google Code-in website, <https://codein.withgoogle.com/archive/> (27 April 2018)
- Google, 2018, Google Code-in 2017: more is merrier!, <https://opensource.googleblog.com/2018/02/google-code-in-2017-more-is-merrier.html> (27 April 2018)
- Hagan, D., and Markham, S., 2000. Does it help to have some programming experience before beginning a computing degree programme. *Proceedings of the Innovation and Technology in Computer Science Education*, July 11-13, 2000, Helsinki, Finland.
- Open Source Geospatial Foundation (OSGeo), 2018, OSGeo website, <https://www.osgeo.org> (7 February 2018)
- Steinmacher, I., Silva, M.A.G., Gerosa, M.A., and Redmiles, D.F., 2015. A systematic literature review on the barriers faced by newcomers to open source software projects. *Information and Software Technology*, 59, pp. 67-85.