Citizen scientists in volunteer computing: . . .

Jose Luis Fernandez-Marquez [*], Francois Grey [*], Ioannis Charalampidis [†], Ben Segal [†] and SpMohanty [‡]

[*]University of Geneva,[†]CERN, and [‡]??

Submitted to Proceedings of the National Academy of Sciences of the United States of America

Keyword1 | Keyword2 | Keyword3

Abbreviations: SAM, self-assembled monolayer; OTS, octadecyltrichlorosilane

Introduction

- From citizen science to volunteer computing.
- what is volunteer computing?
- what do we currently know about the people participating in volunteer computing?
- Contribution of this paper. Why do we want to know who we are? - Expand volunteer computing to a wider audience, . . . (improve it depending on the analysis results)

A citizen science project is a collaborative reseach either fully or partially conducted by people without any required formal expertise.

There are different aspects motivating the citizen partipation in research projects: (1) reducing cost [2], (2) enhancing learning and creativity[1], (3) going beyond current limitation of computational techniques such as, artificial vision or text recognition, and (4) promoting research and making citizen aware about the scientific research efforts.

Over the last 15 years there were an increasing number of ways citizen participate in research projects. Thus, the need of embracing them under a common term, so called citizen science.

There are many different ways to participate in citizen science: (1) Data collection, also called participatory thinking, where citizens participate to gather data (e.g. monitoring noise pollution in a city by using mobile phones, WideNoise app [2]); (2) Analysis tasks, also called participatory thinking where citizen scientists helps to analyse data that is easy to analyse for human but very complicated or even impossible for computers. (e.g. classifying galaxies [3], Extracting information about shelters in Africa [4] and many other examples[5]); (3) Pooling of resources, also called participatory computing or volunteer computing, where citizen scientists share computational and storage resources of their computers to be used in scientific projects.

Volunteer computing appears in 1996 for first time in a project called "Great Internet Mersenne Prime Search" focused on the search for the largest known primes.

Volunteer computing reached a large audience in 1999 with the projects SETI@Home[6] focused on search for extraterrestrial intelligence, and Folding@home[7] aim to find new drugs by simulating protein folding. Both projects reached more than a hundred thousand of participants.

The term "volunteer computing" was originally mentioned by Luis F. G. Sarmenta in 2001 on his PhD thesis titled "Volunteer computing" [1].

A major contribution in volunteer computing appeared in 2002, Berkeley Open Infrastructure for Networking Computing[8] (BOINC) project at the University of California. BOINC provides a middleware for volunteer computing and grid computing allowing scientists to create volunteer computing projects. Currently BOINC has a strong community of more than two hundreds thousand volunteers and more than 40 different projects. BOINC provides a GUI that allows user to easily participate and choose those projects the user wants to participate.

Even though the increasing participation of citizen in volunteer computing over the last 15 years, little is known about who they are, thus making difficult to reach them for creating volunteer computing project, and understanding their motivation to participate in citizen science project by providing computational and storage resources.

A recent publication [2] analysed the citizen partipation in 6 different projects hosted on Zooniverse[9]. From this analysis, they observed that most of people who start participating in a project never come back, reaching up to 83% at the worst case. Visits on the project are usually quite short (7 to 25 mins.). Additionally, they found that the top 10% of contributors carried out about 80% of total effort on those projects. Moreover, the noticed that the participation was highest right after right after the first release, going progressively down until they stabilised on a given number of users, also the are some peaks of audience.

All these numbers show that citizen participation in science is still in its enfancy and it is needed a better understanding about who are the people behind volunteer computing (i.e. who is this low percentage that make most of the work, or the huge percentage that never come back?), how we reach this subset of extremely high active participants (i.e. social platforms, journals, or data hubs), and how much they contribute.

To answer all these question the Citizen CyberScience Center[10] (CCC) organised a volunteer computing event hosted at the European Organization for Nuclear Research (CERN). This event called CERN 60 Public Computing Challenge[11] run for 12 days in December 2014, as part of CERNs 60th anniversary celebration. Volunteers participating in this event computed simulations of particle collisions in the Large Hadron Collider (LHC) and other particle accelerators that have been active during the 60 years of CERN's history. Data analytics tools were used to gather information about the audience, user' adquisition and users' behaviour. All this large amount

Reserved for Publication Footnotes

[1] http://citizencyberlab.eu/
[2] http://cs.everyaware.eu/event/widenoise
[3]
[4]
[5] crowdcrafting, zoouniverse
[6] http://setiathome.ssl.berkeley.edu/
[7] http://folding.stanford.edu/
[8] http://boinc.berkeley.edu/
[9] https://www.zooniverse.org/
[10] http://www.citizencyberscience.net/
[11] https://test4theory.cern.ch/challenge/

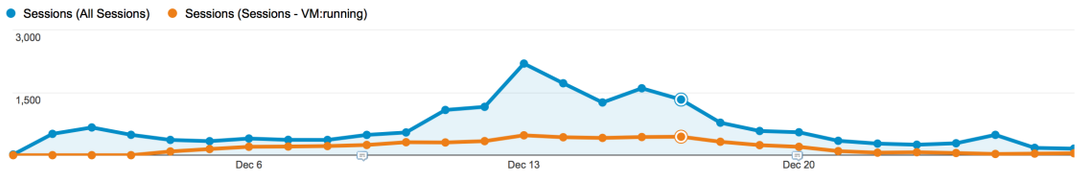

Fig. 1. Sessions versus sessions running the VM

of data was analyzed after the CERN event and used to validate our hypotesis and provide a set of suggestion about how to run successful volunteer computing projects.

As conclusion, this paper proposed a set of advises to deploy successful volunteer computing projects.

Related Work

-
- From the beginning of participatory computing to now, what we did we learn about citizen participation?

Volunteer computing.

What we already know about people behind volunteer computing.

CERN 60 public computing challenge

The CERN 60 Public Computing Challenge was a 12-day event that ran in December 2014, as part of CERNs 60th anniversary celebrations. . . . . .

Data Analysis

This section summurises the data gathered from the CERN 60 Public computing Challenge.

Analytics Framework. This section describes the analytics framework used to gathered and process data from the CERN 60 Computing Challenge event. The analytics framework based on Google Analytics provides two main features: (1) Information coming from Google Analytics about audience (i.e. age, gender, location, interest), users' adquisition (i.e. referrals,) and users' behaviour (i.e. number of sessions per user, avg. time per session, and (2) User information about their acciones performed over the course of the challenge (i.e. downloading, pausing, booting, clicking related projects, etc..).

To gather infromation related to the users, we implemented a library that is connected with Google Tag Manager. This library allows are to track user activity in an anonimous way, and send it to google analytics.

A major aspect of sensing this data to google analytics is to allow the combination with audience, or user's adquisition. E.g. we could see where people who has booted the virtual machine more than 20 times are coming (country, referrals, etc..).

- Technical description of what have been tracked and how we tracked.

Audience.

- Summary of audience data.
- is it a representative statistical sample?

Over the course of the event GA reported around 7,761 users and 13,278 sessions. Citizen scientists were coming from 109 different countries and 86 different languages.

Google Analytics was able to retrieve users information such as gender, age, and interest of around 50% of all session. From this data we know that 90% of participants were male and 10% female. The major interest of participant is technology and the common volunteers are between 18 and 33 year old.

Additionally to the information provided by google analytics we analysed different key engagement indicators to measure the users participation.

Out of the 7,761 users, 1377 start the virtual Machine (i.e. they contributed sharing computational resources) and 1,284 logged in the system.

Conclusions from the audience data. . . .

Adquisition.

- Summary of audience data.
- is it a representative statistical sample?

Behaviour + .

- Here we should focused on the engagement. . . . picture about those who has contribute over the course of the whole event.
- is it a representative statistical sample?

Conclusion

- who are behind volunteer computing?
- What did we learn? (no female participation, relavance of translating the site to gathering citizen, relevance of technology blogs and newspapers to gather new users,)
- is this work opening other rearch questions? Can we reach female participation ?
- Can this results be generalised to other projects in volunteer computing?
- How this work will impact the next CERN Challenge on April 2015?
- How this work impact in volunteer computing?

ACKNOWLEDGMENTS. This work has been supported by the EU-FP7 STREP project Citizen Cyberlab: Technology Enhanced Creative Learning in the field of Citizen Cyberscience.

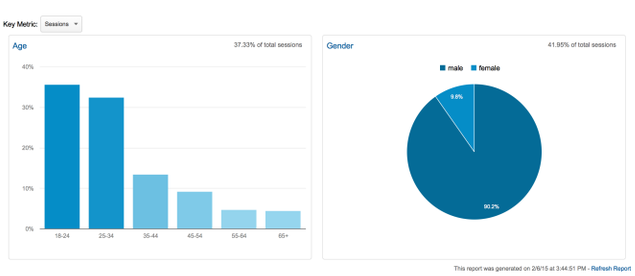

Fig. 2. Participant ages and genders

	Sessions	% New Sessions	New Users
	13,278 % of Total: 100.00% (13,278)	55.81% Avg for View: 55.46% (0.64%)	7,411 % of Total: 100.64% (7,364)
1. Direct	7,795 (58.71%)	49.89%	3,889 (52.48%)
2. Social	4,510 (33.97%)	70.29%	3,170 (42.77%)
3. Referral	853 (6.42%)	39.98%	341 (4.60%)
4. Organic Search	118 (0.89%)	9.32%	11 (0.15%)
5. (Other)	2 (0.02%)	0.00%	0 (0.00%)

Fig. 3. Acquisition

	Sessions	% New Sessions	New Users
	4,510 % of Total: 33.97% (13,278)	70.29% Avg for View: 55.46% (26.74%)	3,170 % of Total: 43.05% (7,364)
1. reddit	3,387 (75.10%)	75.35%	2,552 (80.50%)
2. Facebook	686 (15.21%)	65.16%	447 (14.10%)
3. Disqus	192 (4.26%)	3.65%	7 (0.22%)
4. Twitter	191 (4.24%)	63.87%	122 (3.85%)
5. Hacker News	25 (0.55%)	96.00%	24 (0.76%)
6. VKontakte	20 (0.44%)	70.00%	14 (0.44%)
7. Google+	5 (0.11%)	20.00%	1 (0.03%)
8. LinkedIn	2 (0.04%)	50.00%	1 (0.03%)
9. Ning	1 (0.02%)	100.00%	1 (0.03%)
10. Pocket	1 (0.02%)	100.00%	1 (0.03%)

Fig. 4. Acquisition - Referrals

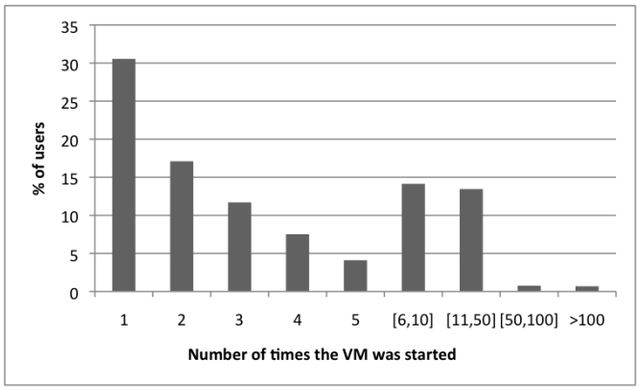

Fig. 5. Engagement - number of times users are running the VM

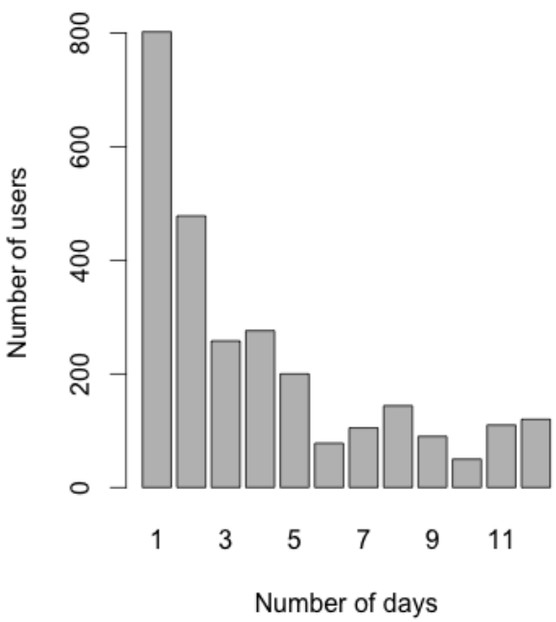

Fig. 6. Engagement - number of users/number of days

1. Luis F. G. Sarmenta. Volunteer Computing. PhD thesis, Massachusetts Institute of Technology, 2001.

2. Henry Sauermann and Chiara Franzoni. Crowd science user contribution patterns and their implications. Proceedings of the National Academy of Sciences, 2015.