A case of Citizen Science for Cell Biology Images Analysis

Eduardo Lostal Lanza[1,2], Fermín Serrano Sanz[1,2], Carlos Val Gascón[1,2], Francisco Sanz García[1,2], Patricia Martínez Alonso[2], J. A. Carrodeguas Villar[2]

[1] Fundación Ibercivis, Spain

[2] Instituto de Biocomputación y Física de Sistemas Complejos (BIFI) de la Universidad de Zaragoza, Spain

{eduardol,fermin,carlos.val,frasanz}@bifi.es, carrode@unizar.es

Abstract. *As a consequence of the quick advance in technology, people is more connected than ever. Researchers can take advantage of this fact to overcome problems, that some of them are facing and have become their bottleneck. Infeasibility for processing vast amount of information or the need for a wider range of inputs are some of those problems. Citizen science seems to be a proper approach to overcome them at the same time that make general public participant of the science process. A complete scenario in cell biology research field is presented where cell images analysis is carried out through crowdsourcing.*

1. Introduction

Digital Society is assuming a social and information revolution. Super, large scientific infrastructures and high performance communications have resulted in e-science and allow addressing problems that were unsolvable until a few decades ago. Similarly to those large infrastructures, we can consider that millions of people permanently interconnected compose a citizen-based science (c-science) infrastructure. This c-infrastructure can be used for distributed computing, collective talent or ubiquitous data gathering. As a result, we have a complex entity with intelligence and collective knowledge.

Researchers can take advantage of it. Paradoxically, some researchers have found a bottleneck in the advance of technology and computational resources. Analysis of the results and information produced is slowing them down to keep on with their research. In other cases, scientists have a need for a very high amount of samples that need the contribution of a large community that is not always available. That is why they can make use of those c-infrastructures to collaborate in their needs.

Citizen science is a research conducted by members of the public, not necessary scientists, who work with the latter to carry out a research project. The objective is to increase public awareness and make them participants of science development. Citizen Science removes barriers between professional researchers and amateur volunteers. Some of the tasks performed in the laboratory can be outsourced for any citizen with small or no scientific background at all required.

The paper is organized as follows. In Section 2 we introduce Citizen Science and other projects. Then cell biology and motivation behind the experiment being presented is depicted in Section 3 in addition to a description of the whole scenario. Through section 4, infrastructure is detailed. Benefits in education of these kinds of experiments are discussed in Section 5. Section 6 concludes the paper summarizing its contribution.

2. Citizen Science

Besides being a recently given name, it is not a new technique. During the last century, scientists have merged the efforts of citizen and professional scientists to obtain results in research projects that could not be obtained otherwise. Such is the case of Audubon Society's Christmas Bird Count or American Association of Variable Star Observers. The main tasks of the volunteers at those projects was data gathering. However the evolution in computation resources and the distribution of Internet changes that approach making it possible to involve volunteers in new tasks.

As a result, BOINC [Anderson 2004] was created as an infrastructure to give support for a new perspective of citizen science. In this case, participants give part of their computational resources to perform research operations, i.e. volunteer computing. Nevertheless, network resources allow citizen science to go a step forward and take the initiative to participate in a more interactive way in the science process. Therefore, other projects have emerged like the ones at Cornell University to help researchers better understand birds and their habits [Bonney 2008], Galaxy Zoo[1] for classification of galaxies that encouraged Zooniverse[2] creation later, or FoldIt[3], an online puzzle video game about protein folding. Galaxy Zoo even explored the motivations under the volunteer to get involved in the projects. The result was that a relevant percentage of users participate for the simple fact of contributing [Raddick, Jordan et al. 2010]. Those projects had such an impact that made us realize the web has increased the possibility for a deeper involvement of those citizen scientists exponentially. Also, existence of large communities of volunteers behind those projects demonstrates the feasibility of this promising research approach.

Socientize is a FP7 project[4] that aims to promote the usage of scientific infrastructures including professional and amateur scientists. The idea is to show the capabilities of open resources. In order to do that, several experiments are being developed at the same time that a white book of Citizen Science is being written. Ibercivis is leading the project and providing most of the underlying infrastructure.

3. Cell Images Analysis Scenario

One of the major lines in cancer research relates to the identification of drugs to fight it. Some of these drugs could function by selective killing of tumor cells, interacting with

---

[1] www.galaxyzoo.org

[2] www.zooniverse.org

[3] fold.it

[4] www.socientize.eu

proteins or metabolic processes (chemical reactions that take place inside cells) that are modified in tumor cells but not in healthy ones.

Therefore, a first step in the search for antitumor drugs could be the identification of chemical compounds that kill tumor cells, selecting in a second step those drugs that do not kill healthy cells. There are two major types of cell death, apoptosis and necrosis. Apoptosis is a type of programmed cell death, physiologically controlled, that eliminates cells in a silent manner. Necrosis, on the other hand, causes an inflammatory response due to cell content release that can damage surrounding tissue. When we want to kill tumor cells inside a tissue we want them to die by apoptosis. A similar successful approach has been used with stem cells, which in some aspects resemble tumor cells [Conesa, Doss, Antzelevitch, Sachinidis, Sancho and Carrodeguas 2012].

These two types of cell death can be easily differentiated by microscopy in cultured cells. Advanced optical microscopy techniques allow cells to be cultured directly over the microscope in multiwell plates (cell culture flasks with several individual wells). In this way we can culture tumor cells and treat them with a different chemical (a putative pharmaceutical drug) in each well, and observe which chemicals are more efficient in cell killing, and by which mechanism (apoptosis or necrosis) by observation of changes in different cell structures. These structures can be individually labeled with specific fluorescent dyes and observed using fluorescence.

By photographing cells at different time intervals and assembling consecutive images after the experiment we can obtain videos that show cell movement and changes over time, which is known as time-lapse microscopy.

Cells are grown in each well of a 24-well plate, adding a different chemical per well, and photographed every half hour with the appropriate filter for two fluorescence colors and in bright field, obtaining three different images of the same field at each time point, in four different positions per well, in all 24 wells, over a period of 24 hours. This produces 14112 images in a single experiment. The different images obtained in each field at each time point were overlaid (bright field, blue and green fluorescence) to obtain a combined image, and were assembled chronologically to generate videos that show cell movements, cell division and cell death, if they happen, over time.

As a consequence, more than ten thousand images per experiment must be analysed. That analysis includes classification of the cells according to several parameters like cell and nucleus shape, mitochondria distribution or current cell status. All those huge data sets obtained cannot be analyzed by means of individual researchers. The optimal approach, given the quick advance of the technology, would be to face the analysis making use of computer image analysis techniques in some kind of distributed infrastructure with high computational resources. Nevertheless, computers are not as good as human brains in what concerns to pattern recognition. Namely, the classification of the previous parameters, although might be difficult even for humans, becomes harder for computers without an exhaustive training.

Aiming to address that issue, that means a bottleneck in the advance of research, citizen science stands up as a great solution. The goal for citizen science use in this line is twofold. On one hand, as can be expected, people's involvement permits dealing with large data sets, providing an amount of analyzed results that would be unreachable otherwise. On the other hand, those results intrinsically mean the creation of a large set of trainings that can be used for machine learning techniques. Keeping in mind those objectives, the diagram for the cell images analysis scenario is depicted in Figure 1.

The idea behind that scenario is to take advantage of the analyzed results from the volunteers not only as an input for the researcher, but also as trainings for a machine learning tool with the goal of improving pattern recognition software. Therefore, in parallel two software applications must be developed. The first one must be a platform to be used by volunteers to carry out the analysis. The results are gathered and delivered to the scientist behind the research. Those results are passed as inputs for the second software. This one is responsible for carrying out the automatic analysis on the computer. Through a closed-loop feedback system using the analyzed results for feeding a machine-learning algorithm, the performance of the computer keeps improving. The final objective is to have a computer infrastructure properly trained whose performance is equivalent to a human. That scenario will allow us to overcome the bottleneck of the analysis time meaning a significant contribution for the researcher who could get the results of the experiments in an extremely short time.

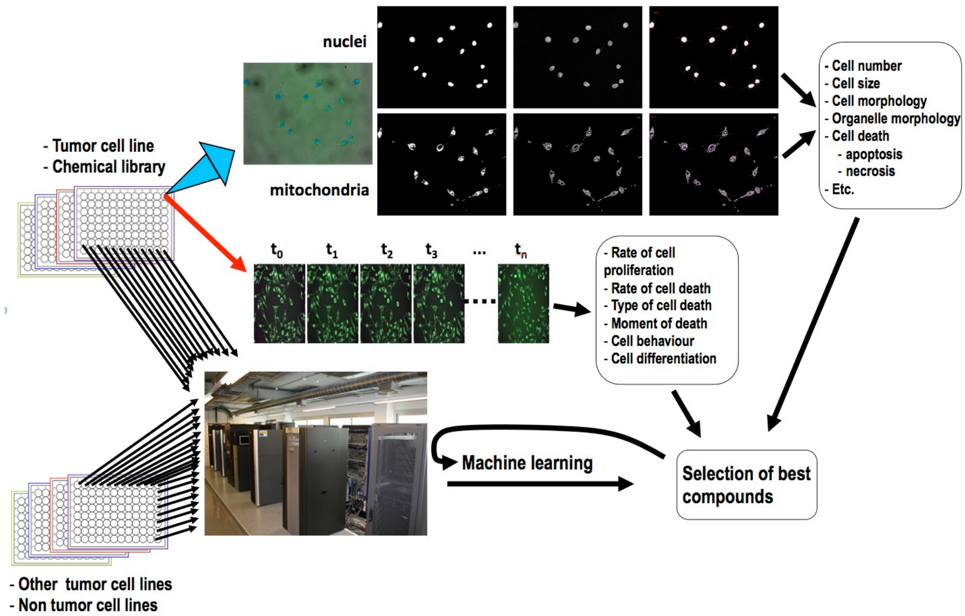

Figure1. Cell Images Analysis Scenario Diagram

3.1 Cell Images Application

The application is developed to provide a framework to the volunteers in order for them to have access to the images and carry out the analysis. Application features will be discussed next.

Usability has been one of the main concerns. The goal is to get it working for a wide spectrum of people who is not expected to be a scientist. Therefore, a key point in the development of the application is that it must be intuitive and user-friendly in order for the volunteer not to get lost with non-relevant features; user must be self-sufficient.

Application shows one image at a time to be analyzed by the citizen scientist. A snapshot can be observed at Figure 2. There are nine different parameters that made up the analysis. In order to ease the task to the user, every one of these parameters is analysed and replied in a different step, keeping the image always visible. The end-user has the chance to change the fluorescence channel to properly observe the parameter being studied at a time. Every step has several samples in order to provide the user with a close example of what he or she has to classify. A magnifying glass tool is implemented to stare at the image in a more accurate manner.

Since the application is expected to be used by any person interested in the science process, a short training must be taken to get a necessary background. That is provided through a tutorial during the first time the application is used. In addition, an endless number of tooltips and step-context specific information are available at any time.

A relevant aspect of this scenario is to get participants involved in the science process. In such a way, a remark section is provided in order for citizen scientists to be able to write anything that might have been noticed, in the last step. The idea is to encourage people to participate in that research process, what in other successful scenarios like the Galaxy Zoo one, has led to significant contributions and discoveries [Cardamone et al. 2009].

Results must be gathered and saved. The same image is sent to a number of users to be analyzed. Their replies are matched and through a statistical tool with the whole amount of responses for an image, the replicas, calculates its average, mode and standard deviation. Thus, final result to be passed to the scientist is obtained. That kind of crowdsourcing is an interesting feature of citizen science. It is the match of the whole replies what sets a more accurate and trustable method. Possible errors are corrected through the collective intelligence of the participants.

Citizen Science benefits for researchers are clear, they obtain a unique contribution from a large set of assistants to overcome the lack of resources. On the other hand, volunteers are improving their background and knowledge about the research topic they get involved with [Brossard, Lewenstein and Bonney 2005], cell biology in this case.

3.2 Current stage

The project is an ongoing work that has reached the initial phase for the deployment. On one hand, application above described is finished and deployed on a Citizen Science infrastructure. Two months later its deployment, around six hundred volunteers have been contributing completing almost three thousand tasks. Most of the volunteers come from school education where application is being used for biology courses. The goal of this phase is to obtain a proper feedback to improve the application before its final deployment. On the other hand, a first version of the computer vision software for automatic recognition has been developed and is working properly. Thus, efforts at this regard are being focused on improving the algorithms and the machine-learning techniques.

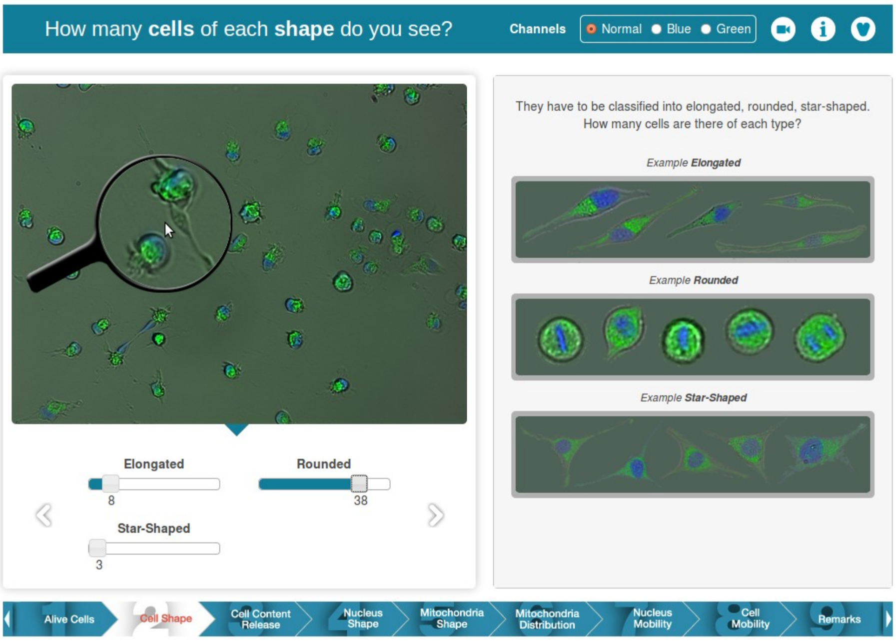

Figure2. Snapshot of the Cell Images Analysis Application

4. PyBossa as a Citizen Science Infrastructure

PyBossa is the infrastructure that stores some of the Socientize experiments (and namely, the cells one). In such a way, applications are kept together and accessible for the general public. It is an open source platform for crowdsourcing where volunteers perform tasks that require human cognition, knowledge or intelligence [CCC and OKFN 2013]. That infrastructure was inspired by the BOSSA crowdsourcing engine, which is the same than the BOINC one. Tasks are sent to the user to be carried out following some type of scheduling depending on the nature of the application at hand.

A major benefit from using PyBossa is the fact that it is a platform built for deploying crowd-sourcing and microtasking apps rather than being a crowd-sourcing application itself. Apps are written as snippets of JavaScript and HTML 5 to be deployed in any PyBossa instance. That instance will manage users, data and workflow, easing the work of the researcher. PyBossa is a python web application built through Flask micro-framework, using PostgreSQL as the main database and SQLAlchemy for object-relational mapping.

A final scheme of the application developed for cell images analysis is depicted at Figure 3. As aforementioned, cell images analysis application is added to a PyBossa instance. Users access the application through the internet. PyBossa gathers the results and saves them in the database placed at Ibercivis servers.

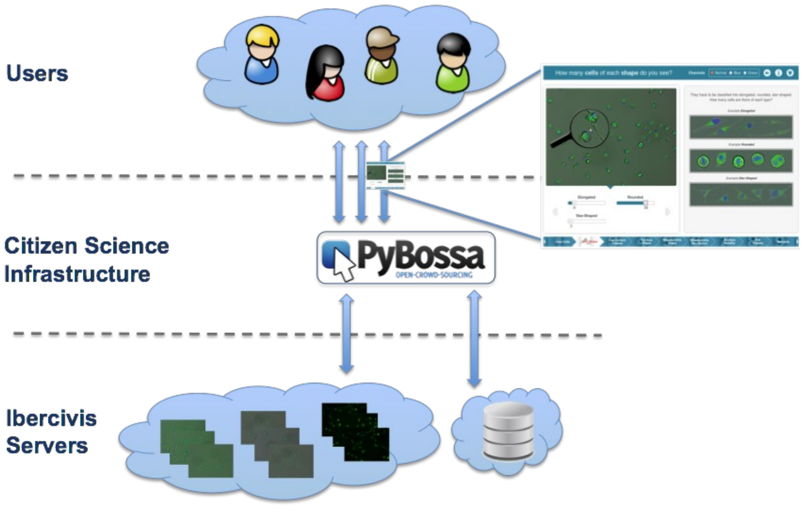

Figure3. Cell Images Analysis Infrastructure Scheme

5. Citizen Science and Education

Benefits of Citizen Science have been aforementioned. Nevertheless, in what concerns to education, we consider benefits for the volunteer increase significantly. The use of new technologies that students are familiar with can deal with the increasing lack of interest from their side. Step by step, technologies are being integrated into the education system. However, education still needs to move to more practical methods, not only in what concerns to new technology. Namely, in the science field, educators start to recognize the positive influence in education of the participation in authentic scientific experiments [Michaels, Shouse and Schweingruber 2008]. They develop understanding through practical experience. That is the idea behind the statement: "what you hear, your forget; what you see, you remember; what you do, you understand". Engaging in that process, students may also have an increase in their scientific thinking thanks to the feeling of carrying out authentic science [Trumbull, Bonney, Bascom and Cabral 2000].

Summarizing, we consider that citizen science in education brings along the integration of two relevant concepts: use of new technologies and science process participation. By using tools and technologies they are familiar with they can work on real research feeling themselves a key part of the scientific process and that they are doing a significant contribution [Lostal, Serrano, Carrodeguas, Martínez, Sanz and Val 2013]. They also develop other relevant aspects such as collaborative working either by the interaction with the scientist or with other students, at the same time that they start to create a scientific curriculum. However, it is the fact of getting involved in the project and that their contribution is truly significant what is expected to motivate the students.

6. Conclusion

Citizen Science is becoming a promising approach to help researchers who have an increasing need for higher resources or contribution of a significant amount of volunteers to continue with their research. In this work, we have addressed the first case presenting a scenario where Citizen Science provides a short and long-term solution in the field of Cell Biology. Easiness to outsource that solution to a wide range of similar problems, make us think Citizen Science is a feasible and exceptional solution. Benefits of its application in the education field have been also discussed.

Acknowledgements

A special thank you goes to Daniel Lombraña for his great work on PyBossa, the rest of the Socientize team and those volunteers who keep on helping with their contributions.

References

Anderson, D. (2004). “Boinc: A system for public-resource computing and storage”, 5th IEEE/ACM International Workshop on Grid Computing, p. 4-10.

Bonney, R. (2008). “Citizen science at the Cornell lab of ornithology”, Exemplary science in informal education settings, p. 213-229.

Raddick, M. Jordan et al. (2010). “Galaxy Zoo: Exploring the Motivations of Citizen Science Volunteers”, Astronomy Education Review 9(1). American Astronomical Society.

Michaels, S., A. W. Shouse, and H. A. Schweingruber (2008). “Ready, set, science: Putting Research to Work in K-8 Science Classrooms”, National Academies Press.

Trumbull, D. J., R. Bonney, D. Bascom, and A. Cabral (2000). “Thinking Scientifically during Participation in a Citizen-Science Project”, Science Education 84(2), p. 265-275.

Brossard, D., B. Lewenstein, and R. Bonney (2005). “Scientific knowledge and attitude change. The impact of a citizen science project”, International Journal of Science Education 27(9), p. 1099-1121.

Conesa, C., M. X. Doss, C. Antzelevitch, A. Sachinidis, J. Sancho, and J. A. Carrodeguas (2012). “Identification of Specific Pluripotent Stem Cell Death-Inducing Small Molecules by Chemical Screening”, Stem Cell Reviews and Reports 8(1), p. 116-127.

Cardamone, C. et al. (2009). “Galaxy Zoo Green Peas: Discovery of A Class of Compact Extremely Star Forming Galaxies”, Monthly Notices of the Royal Astronomical Society 399(3), p. 1191-1205.

Lostal, E., F. Serrano, J. A. Carrodeguas, P. Martínez, F. Sanz, C. Val (2013), “Cell Images Analysis as a Case of Citizen Science for Advance Education: Laboratory and School, Back and Forth”, In Proceedings of the 7th International Technology, Education and Development Conference (INTED 2013), p. 2489-2496.

Citizen Cyberscience Centre (CCC) and Open Knowledge Foundation (OKFN) (2013), “PyBossa Documentation”, Release 0.1.