



Crowd4Sat

Executive Summary

ESA Reference	AO/1-8068/14/F/MOS
Work Package	
Deliverable reference	
Date Completed	11 April 2016
Version	1.0

Prepared by:	Suvodeep Mazumdar	Date:	11 April 2016
Checked by:		Date:	
Approved by:		Date:	
Released by:	Suvodeep Mazumdar	Date:	11 April 2016

Author(s)	Affiliation
Suvodeep Mazumdar	University of Sheffield
Stuart Wrigley	University of Sheffield
Fabio Ciravegna	University of Sheffield
Camille Pelloquin	Starlab Ltd.
Sam Chapman	The Flow Ltd.
Laura De Vendictis	e-GEOS S.p.A
Domenico Grandoni	e-GEOS S.p.A
Michele Ferri	Alto Adriatico Water Authority
Luca Bolognini	Aizoon
Project manager	Affiliation
Suvodeep Mazumdar	USFD
Verified by	Affiliation

Change Record			
Issue	Date	Changes made	Author
0.1	01 April 2016	First version prepared	Suvodeep Mazumdar
0.2	11 April 2016	Final version prepared with CS validation/quality assurance	Suvodeep Mazumdar
0.3	07 May 2016	Final version for only Executive Summary	Suvodeep Mazumdar

Distribution Record			
Number of Copies	Date	Sent to	Format

Crowd4Sat: Crowdsourcing for Observations from Satellites - Executive Summary

Suvodeep Mazumdar, Department of Computer Science, University of Sheffield

Introduction

The Crowd4Sat project explored a range of crowdsourcing methodologies and technologies, from opportunistic sourcing - the ability to extract relevant data from unrelated activities - to participatory sourcing, where citizens and authorities explicitly participate in the data collection. The crowdsourced data and information collected ranged from real time pollution data from cars (via the The Floop's technologies who have hundreds of thousands of users on four continents), to crowdsourced geotagged images from sites like Panoramio, Flickr and Twitter, to social media messages (e.g. from Facebook, Twitter, Foursquare, etc.). Observations from Satellites (OS) data was sourced from a wide range of ESA missions and products including ERS/Envisat and Sentinel- 1 and additional OS sources such as Landsat-8 and MODIS. A variety of stakeholders such as authorities, emergency responders, city councils, insurance companies, as well as single individuals and citizens associations have been involved throughout the process. The project addressed concrete scientific and societal problems through four use cases demonstration projects, targeted at key scientific and societal issues: pollution in metropolitan areas, land use, water management and snow coverage, and flood management and prevention.

A variety of activities were conducted throughout the project - conducting reviews of relevant initiatives, roadmapping activity, stakeholder analysis, requirements analysis, technology design and development, user evaluations and stakeholder feedback. These activities resulted in several findings, all of which have been documented in several reports and deliverables. This chapter summarises the findings and provides recommendations for ESA and the citizen science and crowdsourcing communities. The recommendations are in line with the ones identified by the roadmapping and state of the art activities of Work Package 1, with key findings from the demonstration projects. The following section provides a brief summary of the lessons learned in the demonstration projects.

Lessons Learned from the Demonstration Projects

Collecting crowdsourced information from citizens and communities is a straightforward, yet challenging process. A variety of crowdsourcing mechanisms can be employed -

passive sensing (as employed in DP 2 - traffic and pollution), volunteer thinking and participatory sensing (as employed in DP 1, 4 - snow cover, land use and water management), opportunistic sensing (DP 3 - flood emergency planning). The Crowd4Sat project was envisaged as a feasibility study to help understand how crowdsourcing and citizen science can practically add value to ESA space products and services. As a result, one of the main aims of the project was to understand practical limitations and issues that can arise out of engaging with citizens and communities.

One of the main findings from the demonstration projects was the difficulty in involving citizens for one demonstration project (DP1 - snow cover). There was a lot of interest among a special interest group, the hiking community and eventually resulted in several users downloading the developed application. However, no observations were recorded. Several reasons have been attributed to this, all of which are practical concerns that need to be addressed in crowdsourcing / citizen science initiatives. The findings from Work Package 1 recommended that communities would require information fed back from the mobile application in order to stay engaged. This was considered while developing the solution, however it is possible that users would have preferred to have access to more information than what was provided (hiking trails and user submitted snow reports). At the same time, communities may have been unwilling to rely on sources other than their traditional medium of information. Another reason for lack of observations could be that a decreased snowfall this year had reduced public interest and participation for reporting snow. The final reason for missing information could be that the users found reporting observations cumbersome and unfeasible while tackling difficult terrain. All the four reasons are practical issues, alleviating which requires further investigation and research. Developing solutions that, in such circumstances exploit passive crowdsourcing could provide more information from communities. Additionally, opportunistic sensing via social media could also potentially increase the amount of crowdsourced information collected.

The second practical consideration is to understand scenarios where the scale of satellite data and crowdsourced observations are unequal. Hence, augmenting satellite data with user observations at high granularity levels may not provide significant benefits - in terms of ground based road pollution data, augmenting satellite data at a larger scale such as regions of cities are more promising than compared to small sections and individual junctions. One of the possible reasons for this is that smaller sections such as roundabouts are designed by planners to have minimal elevation gradients. The Demonstration project 2 observed that conducting processes of augmenting crowdsourced observations with satellite data can provide a significant value to pollution estimates (as was also confirmed in stakeholder feedback sessions). DP2 also highlighted that the Corine data was not complete enough, and enhancing with crowdsourced mobility data promise a more accurate and fine grained dataset. However, it is essential to note in

order for crowdsourced data to be more widely acceptable, a comparative analysis with traditional high precision instruments is needed in a larger scale.

The findings from DP4 - land use and water management is closely related to the previous finding regarding mismatched scale of reference for land cover datasets (e.g. CLC2012) and crowdsourced observations. While users can easily identify categories such as urban fabric and roads, several categories such as 'discontinuous urban fabric' the Corine dataset are more difficult to identify. Hence, matching observations with the dataset was challenging in certain cases. In cases where mixed categories were involved, users found it difficult to identify the right categories and hence several discrepancies were identified. This is another practical consideration that needs to be made when designing crowdsourcing tasks - visual inspection of observable areas is prone to be misclassified if the classification is not distinct enough, inspite of users being trained experts. User observations are also open to interpretation and this can cause confusion among users.

The final set of findings related to the opportunistically crowdsourced dataset, extracted from social media platforms. The practical applicability of such data varies widely on the use case and as a result, different sources provide more contextual information than others. For example, in case of flood emergency, videos and images shared on Youtube and Pinterest can provide immense help in understanding the scenario on the ground. However, the amount of contextual information available in each source is not the same and hence, there is a need for some further analysis to extract further information. Several challenges exist in order to effectively exploit such datasets. For e.g., a very small fraction (less than 1% of analysed Tweets) of data collected from Twitter provides geolocation information and in most cases, the position of the Tweet referred to an area outside the crisis region. Hence, in order to automatically detect precise locations, further analysis is required. This is an important aspect for what concerns the usefulness of crowdsourced data in the framework of emergency mapping, where timeliness is a critical factor. Another observation from the data was the lack of useful information from media shared by citizens. Public bodies and organisations, on the other hand provide more relevant information. The final challenge in opportunistically crowdsourced data is to deal with redundant information - retweets and re-sharing content in social media may result in re-processing of already processed information and in scenarios of emergency response, this can cause critical delays. On the other hand, understanding the extent of how much a piece of information has been shared can be helpful in understanding how important or critical it is.

Recommendations and Summary

At the core of opportunities to ESA Earth observation activities from crowdsourcing and citizen science are the calibration and validation of satellite data and products as well as passing value to existing products and services. These can be achieved using a number of different levels of citizen science project ranging from volunteer computing through to participatory sensing. The identification of new applications or disruptive products could be achieved through hackathons and crowdsourced solution contests. However, the use of crowdsourcing and citizen science is not without its pitfalls. The two main areas which must be considered with care are engagement and data quality. These two topics are critical to the sustained success of an initiative both in terms of maintaining the amount of data collected or processed over time as well as the value of the final crowdsourced information.

Funding Support: The use of citizen science and crowdsourcing should be prioritised in future EO funding calls in order to support existing crowdsourcing / citizen science initiatives and foster broader uptake. Naturally, care has to be taken in order design programmes which allow different forms of citizen science to take place from civic-initiated to research organisation-led. Calls should be designed in collaboration with national and regional government and non-governmental organisations in order to more closely match citizen needs and understand existing citizen networks suitable for incorporation into satellite- data validation and augmentation initiatives. Attention should also be focused on a broadening of funding approaches to support a wider set of participants including 'true' citizen scientists: citizens who may not have formal scientific qualifications or employment but can still demonstrate exciting, beneficial and valuable uses and enhancement for existing ESA products and services. The demonstration projects highlighted several areas where there is a significant scope for research such as understanding user motivations and continued engagement, developing mechanisms for comparing with high precision measuring instruments, understand interaction mechanisms for providing user input for complex activities and so on. Such areas of research can provide a significant benefit to the crowdsourcing and citizen science communities.

Embedment: Related to the increased financial support, citizen science and crowdsourcing ought to be more strongly embedded across all funding initiatives wherever possible. Funding programmes commonly require actions to improve communications and prospective projects need to demonstrate the positive impacts the project will have if funded. In a similar vein, projects should be strongly encouraged to incorporate citizen science wherever possible. This goes beyond the inherent benefits discussed above in

terms of lower-cost data collection and analysis, wider geographical reach, etc. By 2020, more than 800,000 technology posts across Europe will be unfilled due to the skills gap, and even lower level positions will require increasing levels of STEM knowledge and competence [84]. Efforts are being made at the governmental and continental scale with policies being put in place to encourage wider participation in science, technology, engineering, maths (STEM) as well as citizen science. Endeavours by all funders – ESA included – are essential to support citizen engagement and science learning at all ages. Initiatives that can help citizens and communities relate to their own interests, personal activities and hobbies can have a greater acceptance and engagement, thereby driving sustained contribution from citizens. Understanding how citizens can sustainably contribute to crowdsourcing activities is a complicated activity which needs a thorough understanding of user needs. For e.g., the demonstration project 1 - snow cover provided mechanisms for highly relevant user communities to provide their crowdsourced information, at the same time receiving relevant information. However, it was realised that the process needs to be more inclusive, to understand how much are citizens willing to contribute and in which form.

Understanding: We have seen in this report that citizen science can make considerable high quality contributions to research but this is reliant on an understanding of how citizen science initiatives work, how to ensure citizens remain engaged in the process (Section 4) and how to ensure high data quality both at the design stage and at the analysis stage (Section 5). A wider understanding of the benefits of crowdsourcing and citizen science within ESA could inspire new strategies and initiatives. For example, ESA could help provide guidelines and standards on how best to collect, analyse and re-use citizen science data (including metadata, data protection, ethics and privacy). Understanding can be further fostered with providing access to the data generated by citizens themselves. Easily accessing and visualising the data generated by contributors can provide them with an immediate observable feedback as well as provide a sense of accomplishment, with their contributions being visible. In addition to understanding the contribution of fellow citizens, it is important to share a common understanding of which information is the most relevant. The Demonstration project 4 - land use and water management highlighted that even experts can provide information that may not be most relevant or well-accepted when compared to the information expected to match satellite datasets.

Outreach: In addition to embedding citizen science as a strongly expected theme in future funding calls, citizen science could be used as a way of improved targeting of school children (i.e., before children have made decisions on subjects to pursue at university-level) to engender interest, excitement and involvement in STEM subjects: subject skills key to the continued success of ESA. For example, the Citizen Explorer 1 (CX-1) Mission (see Section 6.1) is an excellent example but initiatives need not be as

ambitious as the development of an entire satellite. The idea of incorporating citizen science and crowdsourcing themes in school curriculum could be explored with differentiating content and communication based on target age, while starting at very first classes of school. Team-building activities based on citizen science and crowdsourcing tasks could be an interesting way of teaching the values and ideas of citizen-driven data collection from an early age. In order to ease the governance overhead, such initiatives could be organised in a federated approach in which national space organisations (e.g., UK Space Agency, DLR, etc.) undertake local coordination. Such initiatives could also improve the public profile of ESA, national space organisations. At the post-school age, excellent examples of good practice are NASA's Space App Challenge and ESA's App Camp. Wider use of such hackathons can increase outreach, inspire more young people, provide novel insights into data use as well as providing an opportunity to bring disruptive ideas to market (as was the case with HackTrain1). Demonstration Project 1 - snow cover already attempted reaching communities who are relevant to special interests such as hikers. This, in the initial stages generated significant response, and resulted in initial uptake of technology. In spite of a highly positive start, no crowdsourced observations were recorded. This shows that in addition to connecting with user communities, continued engagement is a more complex and challenging task and individual communities may have individual needs and requirements.

Communication: Though closely related to Outreach, Communication is a separate activity that needs a significant amount of focus in order to ensure crowdsourcing and citizen science can be exploited to a greater extent (Section 7.1). Overall, communication is key to ensure a continuous participation from citizens and communities. This needs to be adapted to uniquely address different communities, user groups and use cases - some individuals may find it sufficient to observe and visualise their contributions to motivate them to continue contributing, whilst others may need a more personal level of communication from various actors in the initiatives such as citizen science owners or scientists to help explain how the contributions have helped the citizen science owners. Another form of communication is how such initiatives are communicated to citizens to encourage continued citizen participation. Some initiatives have had greater success in engaging citizens and communities over the years - investigating the reasons for successes and failures (appropriate advertising, marketing, media use, communication strategy, incentives, motivation etc.) could be an excellent start to understanding how larger communities can be engaged.

Widening Participation: As noted in Section 4.6, citizen science, in general, is still not representative of all of society (even in western countries) and this unequal participation in terms of gender, ethnicity, age, and socioeconomic backgrounds could have the potential for bias in the project's scientific research. Furthermore, it means a strong societal bias in

STEM engagement and a strong bias in addressing societal and community issues. Efforts are being made to increase participant diversity – for instance, one-fifth of citizen science participants in the UK Open Air Laboratories (OPAL) project were drawn from disadvantaged communities – but it is still an under-explored area. This societal imbalance is exacerbated by the exclusion of lower literacy citizens whose lower information and communication skills make it difficult to participate in 'traditional' citizen science and crowdsourcing initiatives. Although certain elements of crowdsourcing may not be suited without adaptation (e.g., gamification) their contribution would still be valuable. The latter refers back to the previous recommendation for Understanding. While having a focussed and well-targeted user group can be helpful and provide a set of highly interested user community to submit information, this also risks reducing a pool of users who could potentially contribute with observations. Demonstration project 2 - traffic and pollution, in addition to providing a passive mechanism of crowdsourcing also collected information from a large set of users. The users were not domain-specific and this could have contributed to a larger amount of crowdsourced information.