White Rose Research Online
eprints@whiterose.ac.uk
https://eprints.whiterose.ac.uk
Universities of Leeds, Sheffield and York

Deposited via The University of Sheffield.

White Rose Research Online URL for this paper:
https://eprints.whiterose.ac.uk/id/eprint/117619/

Version: Accepted Version

**Book Section:**

Mazumdar, S., Wrigley, S., Ciravegna, F. et al. (2017) Crowdsourcing to enhance insights from satellite observations. In: Bordogna, G. and Carrara, P., (eds.) Mobile information Systems Leveraging Volunteered Geographic Information for Earth Observation. Earth Systems Data and Models, 4. Springer, pp. 35-52. ISBN: 978-3-319-70877-5.

https://doi.org/10.1007/978-3-319-70878-2_2

© 2018 Springer International Publishing AG. This is an author produced version of a chapter subsequently published in Bordogna G., Carrara P. (eds) Mobile Information Systems Leveraging Volunteered Geographic Information for Earth Observation. Earth Systems Data and Models, vol 4. Springer, Cham. Uploaded in accordance with the publisher's self-archiving policy.

**Reuse**
Items deposited in White Rose Research Online are protected by copyright, with all rights reserved unless indicated otherwise. They may be downloaded and/or printed for private study, or other acts as permitted by national copyright laws. The publisher or other rights holders may allow further reproduction and re-use of the full text version. This is indicated by the licence information on the White Rose Research Online record for the item.

**Takedown**
If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.

UNIVERSITY OF LEEDS
University of Sheffield
UNIVERSITY of York

# Crowdsourcing to enhance insights from satellite observations

**Suvodeep Mazumdar[1], Stuart N. Wrigley[1,2], Fabio Ciravegna[1], Camille Pelloquin[3], Sam Chapman[4], Laura De Vendictis[5], Domenico Grandoni[5], Michele Ferri[6], Luca Bolognini[7]**

[1] Department of Computer Science, University of Sheffield, UK

[2] Department of Automatic Control and Systems Engineering, University of Sheffield, UK

[3] Starlab Ltd., UK

[4] The Floow Ltd., UK

[5] e-GEOS S.p.A, Italy

[6] Alto Adriatico Water Authority, Italy

[7] aizoOn Technology Consulting, Italy

**Abstract** Insights from satellite observations are increasingly being used to enhance a range of domains from highly specialised scientific research through to everyday applications directly benefiting members of the public. A particular category of satellite observations - Earth Observations (EO) - are concerned with capturing information regarding the Earth's atmospheric and environmental conditions and observing human activity and its impact on the Earth's surface. A growing number of technologies and services heavily rely on EO data and the rapidly improving fidelity, coverage, timeliness and accessibility of such observations are providing significant opportunities for new applications of economic and societal benefit. With the increasing importance, relevance and size of EO datasets, it is critical to understand how the value of such data can be maximised by complementing EO with other sources of data and efficiently making complex interpretations and decisions. The wide adoption and availability of smartphones, Internet devices and increased accessibility to information has paved the way for large numbers of citizens and communities to participate in scientific, technological, societal and decision making activities. This chapter discusses the experience of the European Space Agency funded Crowd4Sat project led by the University of Sheffield that investigated different facets of how crowdsourcing and citizen science impact upon the validation, use and enhancement of Observations from Satellites products and services.

## 1 Introduction

The role of citizen science (CS) and crowdsourcing is vital to a wide range of applications, spanning a large number of fields such as science, governance, public policy, environmental studies and decision making. Citizens have been employed in scientific studies and decision making processes over the years and several excellent examples have showcased how citizen generated data can provide high quality data. Although concerns exist regarding the assessment of quality and reliability of CS data [1,2], several domains such as knowledge bases, mapping, classification has demonstrated high quality achieved through the rigor of CS combined with multiple independent reviews to check reliability. It has also been reported that such data can be more detailed and higher quality than provided by official institutions [3,4,5,6]. Several large organisations such as Amazon, Trip Advisor, Twitter and Facebook also rely on crowdsourcing as primary sources of information, comprising a critical aspect of their entire business model. WikiPedia and OpenStreetMap, on the other hand serve as long standing testament to the provision of open data which is created, maintained and enriched by the public. With the potential of engaging with citizens, it is important to study how the value of their contribution can be maximised most effectively. The domain of Earth Observation (EO) is increasingly employing CS and crowdsourcing for tasks such as calibration and validation of data as demonstrated by the growing number of publications in the field [7]. While the potential of CS applied in EO can be immense, it is important to understand various factors in engaging citizens and exploiting their contributions in an operational context. There is also a need to study different approaches of employing citizens in different settings and tasks in order to develop techniques and mechanisms for effectively using crowdsourced data. Finally, it is important to understand the aspects needed to be considered while developing crowdsourcing solutions for EO, from the perspective of different stakeholders such as authorities, decision makers, researchers, industry, etc.

The Crowd4Sat project[1] led by the University of Sheffield was a study funded by ESA (European Space Agency) and informed by demonstration projects which investigated how CS and crowdsourcing could contribute to the enhancement, use and validation of satellite observation and products. The project explored a range of crowdsourcing methodologies and technologies, from opportunistic sourcing (the ability to extract relevant data from unrelated activities) to participatory sourcing (citizens and authorities explicitly participate in data collection). The crowdsourced data and information collected ranged from realtime vehicle mobility telemetry (via the The Floow's technologies used by hundreds of thousands of users on four continents) which was analysed to estimate geospatial vehicular pollution, to crowdsourced geotagged images (from sites such as Panoramio, Flicker and Twitter) and 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[2]/Envisat[3] and Sentinel-1[4] and addition-

---

[1] http://www.crowd4sat.eu

[2] https://earth.esa.int/web/guest/missions/esa-operational-eo-missions/ers

[3] https://earth.esa.int/web/guest/missions/esa-operational-eo-missions/envisat

[4] https://earth.esa.int/web/guest/missions/esa-operational-eo-missions/sentinel-1

al OS sources such as Landsat-8[5] and MODIS[6]. A variety of stakeholders such as authorities, emergency responders, city councils, insurance companies, as well as individuals and citizen associations have been involved throughout the process. The project addressed concrete scientific and societal problems through four use cases demonstration projects, targeting key scientific and societal issues: pollution in metropolitan areas; land use; water management and snow coverage; and flood management and prevention. The Crowd4Sat project kicked off on February 2015, and over the duration of 14 months comprised of two main strands: strategic roadmapping, understanding the state of the art, and demonstration projects. A variety of activities were conducted throughout the project including conducting reviews of relevant initiatives, roadmapping activity, stakeholder analysis, requirements analysis, technology design and development, user evaluations and stakeholder feedback. The work assessed the feasibility of adding value to ESA space products and services by using crowdsourcing and citizen science by understanding the practical limitations and issues that can arise out of engaging with citizens and communities. Several recommendations were identified for ESA and citizen science communities.

This chapter introduces the demonstration projects and discusses the findings from the project, to provide recommendations for developing tools and technologies for crowdsourcing for EO.

## 2 Citizen Engagement and Participation in Citizen Science

Several definitions of ‘citizen science’ have been proposed, the earliest being by Irwin [8], describing how citizens accumulate knowledge in order to learn and respond to environmental threats. [9,10] refer to ‘citizen science’ as a form of research collaboration to address real world problems. [1] defined citizen science as scientific activities that non-professional scientists volunteer to participate in data collection, analysis and dissemination of scientific projects. While being only recently coined as a term in a variety of nuances as “a collaboration” [10], “research tool” [11], “genre of (mobile) computing” [12], “activities” [1] and a “trend” [13], the engagement of public in professional research and activities has had a long history dating back to over two centuries. In essence, the field of citizen science has transitioned from citizens having alternate sources of employment conducting scientific activities out of interest to citizens being employed in institutions for conducting research as professional scientists. This transition to professional scientists observed a major growth since the latter part of the 19th Century [1], with increased institutionalisation of science and scientific activities. With the establishment of organisations, research institutions and universities, research activity itself has undergone a tremendous transformation, underpinned by scientific rig-

---

[5] https://landsat.gsfc.nasa.gov/landsat-8/

[6] https://modis.gsfc.nasa.gov/about/

our, processes and protocols. While this has, in many ways contributed to the alienation of citizen scientists, the role of citizens as contributors has continued throughout this process, albeit in selected areas of study such as archaeology, ecology and natural sciences. Citizen roles, for such areas have mainly focussed on the process of data collection and cataloguing observations. Citizens now have ever increasing means of contributing to citizen science, with the smartphone industry revolutionising how citizens can provide data: actively (explicitly sending information via mobile applications, websites, etc.) and passively (collecting and sending data without an active involvement of citizens). Also, in addition to providing observations and opinions in a standardised and processable manner, citizens can also provide evidence for their observations by submitting media (images, audio, video), along with metadata (e.g., timestamps, geolocation, exchangeable image file format). At the same time, hobbyists and enthusiasts can build their own low-cost physical environmental sensors, which can be easily connected to sensor networks via APIs (Application Programming Interfaces). Similar sensors can also be bought off the shelf and distributed to communities to be deployed in larger areas than covered by highly expensive, professional sensors traditionally provided by local, regional or national authorities or agencies. All of these approaches eventually contribute to a greater awareness of the environmental and physical conditions, with a far wider coverage than previously possible. With the availability of social media and qualitative views of citizens, situations on the ground can be far better understood than before.

A key technical component of citizen science is the process of collecting information from citizens via crowdsourcing. The term crowdsourcing was originally coined by Jeff Howe, contributing editor for Wired Magazine, as a portmanteau of outsourcing and crowd [14]. The Oxford English Dictionary defines crowdsourcing as "*the practice of obtaining information or services by soliciting input from a large number of people, typically via Internet and often without offering compensation*". Although this definition mentions use of the Internet, this does not have to be the case - tasks can be offline as well as online. Indeed, some of the earliest examples of crowdsourcing for citizen science have been offline processes [15] where participants sent information and photographs to researchers via postcards. With the availability of different forms of technologies and increasing possibilities of contributing with information, citizens can now participate in a variety of mechanisms based on their levels of engagement - they can merely provide access to computational resources or be highly active, performing tasks and collecting data. Broadly, citizen science projects have been classified by [16] according to increasing engagement by participant as follows:

## *2.1 Passive Sensing*

In passive sensing, participants allow data generated by equipment owned by themselves (e.g. mobile phones, environmental sensors, GPS units) to be collected and used by researchers. Upon initial setup and configuration, most often the volunteers are not expected to actively participate in the project, and their data is seamlessly collected in the background. For example, the Weather Underground project has a network of over 100,000 personal weather stations across the US, combining with federally-funded 'official' weather stations to provide data for forecasting.

## *2.2 Volunteer Computing*

In volunteer computing, participants provide spare computing resources to enable researchers to create a 'virtual compute grid', enabling larger amounts of data to be processed more economically than would be possible using local compute resources. One of the first examples of volunteer computing is the large SETI@Home project, launched in 1999 as a Internet-based project. In the project, volunteers install a program that downloads and analyses radio telescope data with the purpose of detecting intelligent life outside Earth.

## *2.2 Volunteer Thinking*

Also referred to as Human Intelligence Tasks (HIT), volunteer thinking involves participants donating some of their spare time to perform some tasks such as data analysis, visual observations, annotation, etc. These tasks are usually performed online using computers, tablets or smartphones, mostly requiring classification of images or recognition of patterns. Typically, larger jobs are split into smaller tasks that are distributed to large number of workers who, with little effort contribute toward solving a larger task. Microtasks are generally tasks that are difficult for computers and algorithms to complete accurately such as audio transcription, and handwriting recognition. The solutions are often used as training data and ground truth for machine learning algorithms to help improve the performance of automated audio transcribers etc. At the beginning, volunteers are usually asked to undertake some basic training to understand how to correctly perform the tasks. Overall, the amount of time required is low and individual tasks can be quickly completed. One of the most widely adopted and popular examples of volunteered thinking is GalaxyZoo [17], where volunteers are involved in morphological classification of galaxies from images. Websites such as Amazon Mechanical Turk and CrowdFlower provide platforms for citizen science projects to employ microworkers, define tasks and rewards which can then be made available for crowdworkers to search and select topics of interest to them. Several citizen science projects have also employed crowdworkers to solve tasks such as transcribing historical documents [18], detecting colorectal cancer polyps in image scans [19], text annotation of medical documents [20].

## 2.2 Environmental and ecological observation

Environmental and ecological observations involve participants in monitoring and observing the environment for different purposes. Many of the most established examples of citizen science and crowdsourcing fall into this category. For example, the National Audubon Society has been conducting the Christmas Bird Count annually since 1900, involving collecting observations from thousands of participants globally. In the UK, the British Trust for Ornithology conducts similar surveys.

## 2.2 Participatory sensing

Engaged with a higher degree of control and influence over the data collected and analyses, participatory sensing enables volunteers with a greater amount of participation. Activities are typically initiated by external research organisations with close cooperation with volunteers. Such activities typically exploit advanced technologies available in mobile phones. Some examples of participatory sensing include air quality sensing [21], noise level sensing via mobile's microphone readings and geographical locations [22].

## 2.2 Civic / Community Science

Initiating from the citizens and communities themselves, civic science are initiatives where many (if not all) stages of the scientific processes are conducted by non-professional scientists. For example, residents on the Pepys Estate in London, concerned about air quality and pollutants released into the atmosphere by a local scrapyard initiated and guided data collection from citizens. External research organisations were also contracted to perform more specific analyses.

Passive sensing and volunteer computing, while being simpler ways of engaging citizens in science and collect data which are relatively trustworthy, their engagement and often, cognitive capability, is wasted. At the same time, a larger participation from citizens can be possible particularly for volunteers who either do not have enough time to commit to physical or cognitive effort, or those with lower educational attainment. Volunteer thinking and environmental and ecological observation are usually the most commonly deployed and historically established forms of crowdsourcing. Compared to environmental and ecological observation, volunteer thinking however expects a much lower commitment from participants in terms of physical engagement: volunteers are not expected to take readings by visiting locations, carry equipment etc. Instead, volunteer thinking tends to involve rapid or widespread data collection/analysis - e.g. Galaxy Zoo [17] involves the rapid assessment and categorisation of images of galaxies. Participatory sensing and civic science heavily rely on significant engagement from the volunteers, not just in the collection and interpretation of collected data but also in organisation, management, and curation of the project. Particularly for civic science, citizens and professional scientists collaborate as peers in all stages of the scientific process.

## 3 Stakeholders in the EO domain

The success of citizen science and crowdsourcing often relies on understanding the various actors who are involved in the process - not only in data collection, but also analysis, project management and end users. This is a highly important aspect as it helps understanding various expectations from different types of stakeholders. Stakeholders are, in this case the groups potentially interested in crowdsourcing activities and the results originating from them. In the context of the Earth Observation domain, the Crowd4Sat project identified several groups of different nature: decision-makers, scientific teams, industry and citizens. As data from OS does not always meet the expectation of stakeholders in terms of spatial and temporal resolution or information need, such groups have a significant interest in understanding how crowdsourcing can help in improving, validating and extending such datasets. Decision-makers are in general highly interested in citizen engagement and crowdsourcing as a source of data as well as gathering awareness on key societal issues. In addition to understanding local issues and concerns, authorities also rely on highly accurate information such as OS maps to support planning activities and investigations. Through a large scale citizen participation, crowdsourced data can provide them with in-situ measurements that can validate OS and fill-in various gaps identified in such datasets. At the same time, engaging with citizens can help authorities understand highly evolving situations on the ground. For example, during emergencies, citizens tend to be the first observers of unexpected events and hence can provide highly accurate real-time information on events. Crowdsourcing and citizen science enables citizens to become highly involved and engaged actors in decision-making processes, eventually helping authorities make decisions better understood by citizens and better aligned with community's interests and concerns. In spite of the improved situational awareness, encouraging citizen participation by authorities is a potentially high risk activity. It is important to consider several factors in such scenarios for example, ensuring citizens are not put in harm's way as a result of their interest in helping authorities during emergencies. Furthermore, authorities need to also consider the responsibilities and implications of citizen participation such as the need to act upon all information arising out of citizens during large scale events.

Various economic factors also drive the need for improving OS via crowdsourcing for the industry. For example, various companies often look for extended coverage and accurate measurements for resource estimation. With massive crowdsourcing activities, such data can provide in-situ measurements to validate OS products at costs far lower than traditional surveys or data collection activities. For instance, insurance providers need to evaluate risks of natural disasters and limit potential losses that may arise as a result. Such providers can be assisted by improved OS quality by using crowdsourcing using more accurate and up-to-date measurements. The scientific community now face significant challenges due to funding constraints and can hence benefit from large pools of volunteers who could contribute by either providing in-situ measurements, validating observations or even conducting analyses or performing tasks. Many researchers rely on crowdsourcing to provide essential data for their research. Citizens can also benefit from crowdsourcing either personally or via citizen associations (e.g. hikers associations, bird watching associations, local action groups). Citizen data, in large amounts, can bring interesting benefits to citizens themselves through better understanding of their environment, activism around local issues, assisting in existing activities and hobbies and so on. Furthermore, several solutions exist that exploit citizen generated data and are made available to consumers as products. For example, Google Traffic uses passively crowdsourced mobile phone traces to estimate traffic conditions which is subsequently used for journey planning.

## 4 Demonstration Projects

Crowd4Sat addressed key scientific and societal problems through four demonstration projects (DPs) by combining OS with crowdsourced observations. Each DP had a set of stakeholders who were approached to gather initial sets of requirements, which helped set a clear focus on the user needs to ensure a profitable uptake of the products and services offered by the project. The process of requirements gathering was conducted via user meetings during the first three months of the project. DP1 was aimed at validating snow coverage maps produced from MODIS and Sentinel products with crowdsourced information collected through a dedicated mobile application, distributed within the hiking community in Catalonia, Spain. This activity involved interviewing stakeholders from multiple organisations: Federació d'Entitats Excursionistes de Catalunya (Hikers Association), Agéncia Catalana de l'aigua (Water Agency) and Asociación turística de estaciones de esquí y montaña (Ski Resorts Association). The monitoring of snow melt is a key parameter for management of water resources and runoff modelling. In this context, OS is very useful and has reached operational maturity. Furthermore, new satellites such as ESA's Sentinel-1A is expected to improve monitoring of snow cover areas with greater accuracy and improved revisit times. However, measuring snow cover areas with SAR satellite image processing comes with intrinsic slant-range distortion problems such as foreshortening, layover, shadowing. This DP proposed to exploit crowdsourced observations from hikers to validate processed snow coverage maps. Hikers were approached via hiking associations and were provided with Android and iOS mobile applications, which would allow sending reports of snow presence/absence information (along with geo-localised images) to a server collecting all information. The mobile apps would also provide hikers with hiking route tracks to provide further means of engagement.

DP2 was aimed at understanding how opportunistically crowdsourced vehicle telematics data can be combined with satellite remote sensing and in-situ data to improve pollution mapping and modelling for local authorities of large metropolitan areas. The interviewees for this activity were primarily Sheffield City Council (Traffic and Planning, Air Quality, Strategy, Sustainability), South Yorkshire Transport Planning Executive body, and South Yorkshire Intelligent Transport Systems. This DP aimed at enabling usage of mass road crowdsourced mobility data to help better understand road pollution. Existing methodologies to observe road pollution involve few ground based calibrated monitoring stations, but such sensors fall short of providing fine grained emissions across road networks. Such sensors are only few, owing to the high cost of procurement and installation as well as located in locations not ideal due to the need for installation in safe places. The project seeked to address this data gap by using a combination of crowdsourced mass mobility GPS trace data, Corine OS Land Usage data and digital elevation models.

DP3 evaluated how opportunistically crowdsourced Social data (eg. Twitter, Youtube, Facebook etc.) can complement crisis mapping from remote sensed data, improving standard workflow of emergency mapping services as Copernicus Emergency Management Service, that provides crisis maps, actually mainly based on satellite images. Several organisations were interviewed as a part of this activity: Civil Protection Department (Headquarter and Sardegna Regional Office), National Authority for Civil Protection (Portugal), Civil Contingencies Secretariat (UK), Administration of the Republic of Slovenia for Civil Protection and Disaster Relief (Slovenia) and Doncaster Metropolitan Borough Council (United Kingdom). One of the main user requirements of the Copernicus EMS is to receive first crisis information within the first 24 hours after the disaster, while today it is not unusual to experience delays up to 72 hours, mainly due to the availability of the first usable post event satellite image, caused by satellite tasking and orbital constraints, bad weather conditions preventing the collection of optical images, late activation. Moreover, the crisis maps, purely based on satellite information, have known quality limitations due to the physical constraints of satellite acquisitions (e.g. resolution, analysis technique) that affect the thematic accuracy of the analysis. The aim of this DP was to investigate the possibility of using crowdsourced social data to improve the quality and timeliness of emergency flood mapping services. During the project, crowdsourced social data Social data (e.g. Twitter, YouTube, Pinterest), related to the historic flood event that occurred on February 2014 in the Bridgwater area, in United Kingdom, was collected. The data was analysed to generate, before the availability of a post-event satellite image, a “warning map” providing a preliminary information about the areas more affected, and to refine the quality of flood extent delineation of the final Crisis map produced once the satellite image is available.

DP4 centered on land use in the Bacchiglione river catchment area used data from citizens, social media and the last version of CORINE land cover data (CLC 2012) [23] to validate the land cover information via crowdsourced data. This activity involved stakeholders from Alto Adriatico Water Authority (AAWA), Agenzia Regionale per la Prevenzione e Protezione Ambientale del Veneto (ARPAV), Project Unit of Civil Engineers (Veneto Region), Regional Department for Soil Protection (Veneto Region), Planning Strategic Section and Cartography (Veneto Region), Urban planning department of Vicenza and Regional Forest Service of Padua. This DP involved validation of land use map CORINE land cover (CLC 2012 dataset) through crowdsourcing observations by using participatory crowdsourcing mechanisms through the involvement of professional groups. The Water authority of Alto Adriatico was directly involved in the project and organised a dedicated campaign. Similar to the DP1, a professional group represented by AAWA was provided with an Android and iOS mobile phone application. The campaign was in the city of Vicenza and surroundings, directly involving volunteers while opportunistic crowdsourcing observations were also collected manually from Panoramio. Images provided by users were initially checked via a tagging API, which automatically classifies a text tag along with a confidence value. Users selected from a list of ten of the most relevant tags, which were further compared with the CLC2012 dataset values.

The range of different demonstration projects served to help understand how different types of crowdsourcing can be used to collect data to improve OS products. DP2 employed opportunistic sensing to collect vehicle mobility traces from telematics data. DP1 and DP4 employed participatory sensing in two different settings. DP3, on the other hand employed opportunistic sensing via social media.

## 5 Results

This section discusses the results of the demonstration projects and, based on the experience of the project highlights various aspects that need to be considered while engaging with volunteers and participants. DP1, although highly advertised via social media and association channels and shared among the hiking community and hiking associations did not provide any information via crowdsourcing. The DP followed several recommendations typically applicable for crowdsourcing such as recruiting participants through groups of special interests, incorporating the CS mechanism within their own framework as well as close communication with crowdsourcers and hiking groups to co-design the crowdsourcing apps. Several reasons could be attributed to the lack of data in the setting of the DP1 - it could be possible that the hikers had a higher expectation from the information received on mobile phones to serve information beyond their practice of activities. While it is important to provide information typically unavailable to users, it is important to note that users should not be overloaded with too much information to process. Furthermore, hikers rely on well-existing practices and sources of information - there may be strong barriers for new actors that need to demonstrate long-term relevance and validity. Hikers could also be engaged and preoccupied during their hiking activities, therefore unwilling to focus on using technologies.

DP2, using passive opportunistic sensing collected large volumes of data in the region of interest - this is primarily due to the data being collected passively without the need for users to engage in the project regularly. DP2 developed a new algorithm for detecting elevation data by combining ground based survey data, LIDAR data and the crowdsourced mobility traces to better understand road surface elevation and subsequently improve models for estimating road pollution by accounting for the slope of each road segment. The stakeholders in the DP identified clear advantages from better understanding of EO data, which is largely unused in the sector. Another strong potential domain was identified as traffic management, where there is a need for understanding macro-regions. The approach of using crowdsourced data was seen to be strongly positive among stakeholders, however there is a need to evaluate similar approaches on a wider region to better evaluate CORINE data.

DP3 sought to demonstrate the value of information such as images, video and text data related to crisis events, shared among social channels such as Twitter, Facebook, Flickr etc. Such posts are sometimes geotagged via GPS positioning sensors embedded in devices such as smartphones and tablets or can be geolocated through the toponyms of point of interests contained in the textual information. In order to understand how social media can help understand events better, a historical dataset of the UK Bridgwater floods (occurring during 6th and 10th February in 2014) was collected. Stakeholders in the project were provided access to the data along with a WebGIS deployment and invited to provide their feedback and suggestions, raise concerns as well as comment on their observations.The stakeholders identified that information from Social Media can provide a significant advantage and potentially save a lot of time sifting through large number of articles and websites. It is important to note that such an experiment was on a historical dataset and hence, lacked the real-time urgency Emergency Responders experience during disaster events. Indeed, taking in account the quality of Social data gathered in an opportunistic way in terms of lack of geotagging, redundancy of the information, etc, it is important to develop automatic mechanisms to retrieve, filter, geolocate and ranking the Social data to reduce the time spent in their analysis.

DP4 involved participatory sensing in the same spirit of DP1, but applied in a different use case and setting to understand how crowdsourced observations can help validating land use maps. The engagement of participants for crowdsourcing activities was a highly successful event. Around 1200 observations were received. Observations were post-validated based on images sent together with the land use observations. Third-party tools for image classification were used to help validate observations. A major drawback in the DP was an accessibility issue - many areas were private property, and observations about such areas were sent from locations close to the areas and hence would introduce potentially unwanted noise. Furthermore, analysis conducted in the DP highlighted that often, the resolutions of the EO datasets did not often reflect what would be in-situ data - for e.g., the minimum size in the CLC2012 dataset for a land cover area is 25 hectares and 100m in linear scale. Smaller areas are not reflected in the dataset - hence, areas such as a small vineyard surrounded by fields would be classified as ‘Land principally occupied by agriculture’ - discrepancies such as these can be easily filled-in by in-situ sensing.

While participatory sensing in DP1 generated no crowdsourced observations, DP4 was highly successful. However, part of the success could be attributed to the crowdsourcing activity being restricted to a day event. The Alto Adriatico region benefits from a very strong volunteer base that are professionally trained to respond to disasters. Finally, opportunistic social sensing provided significant insights into understanding disaster areas and providing initial warning map that can be potentially exploited to indicate evolving scenarios on the ground.

## 6 Experience from Demonstration Projects

The four demonstration projects provided an excellent opportunity to understand various facets of crowdsourcing when applied in practice to solve societal and technical challenges. The DPs provided a lot of insight into how crowdsourcing tools can be developed and how citizens can be engaged. One of the most interesting findings from the demonstration projects was to understand the limitations involved in engaging citizens to provide large volumes of information. While the process of developing technologies for crowdsourcing and collecting data from citizens and communities is a straightforward process, the project clearly observed there are significant challenges in the process. The primary goal of the project was to understand these challenges as feasibility studies and assess how crowdsourcing and citizen science can practically add value to space products, datasets and services. The roadmapping stage conducted in the initial stages of the project highlighted several recommendations - a primary one being the need to engage with communities by feeding back information from developed technologies. Although this was a primary consideration in DP1, where hiking maps and submitted snow reports were provided to users via the mobile application, the participation in terms of submission of reports was minimal. Several possible reasons could be attributed to this observation - users may have been unwilling to rely on sources other than their traditional medium of information. Another reason could be decreased snowfall during the year may have reduced public interest and hence participation in the crowdsourcing task. A further reason could be the practicalities involved in providing observations - taking time out during hiking to take photographs and submitting text reports on an expensive smartphone while navigating difficult terrain can be a cumbersome process and could be an unfeasible task.

Further work is necessary in order to understand what could be the potential reason for reduced participation, as this was out of scope for the project. At cases such as these, it could be possible to also investigate other forms of crowdsourcing such as passive opportunistic data collection like GPS traces, wearables and smartwatches to seamlessly collect data which could be used to infer snow coverage. However, it serves to demonstrate that engaging with citizens is not a trivial task and continued interest may not necessarily guarantee crowdsourced observations.

Passive opportunistic sensing, on the other hand provided large volumes of data, which could be successfully used by the project to improve models for pollution estimation. While a significant amount of data was collected in the project, it is necessary to be aware of practical consideration when EO data is complemented by CS data. Observations from Satellites are on a very high scale, and comparing with high granularity of the data provided by CS may introduce challenges. For example, in the DP2, estimating road elevation levels and hence the slopes is critical to improving pollution models. Smaller sections such as roundabouts are designed by city planners to have minimal elevation gradients. As a result, augmenting satellite data with ground based sensors at a larger scale are more promising as the elevation gradients are much more pronounced.

Participatory sensing data, involving trained volunteers providing categories of land use data based on visual observations also needs to be handled with care. In the DP4, data was provided by trained volunteers comprising of geotagged images along with manually classified categories. A large number (1200) of observations were recorded, which was a significant success, particularly from the context of crowdsourcing and citizen science projects. However, it is important to note that the task itself may introduce noise and errors at times - for example, some areas may be inaccessible either due to difficult terrain or restricted access. In such cases, automatically geotagged images could identify the observation as relevant to a different area than the one in question. The scale of EO and CS data is different - EO provides imagery from a very high level, while in-situ observations can be highly detailed. As a result, EO data may generalise information to a higher level and inconsistencies may arise - e.g. a vineyard in a farming area would be identified by volunteers as a vineyard while EO could generalise the entire area as farmland. Furthermore, discrepancies can exist in the ability of humans to observe and classify data. Some CLC2012 dataset categories such as ‘urban fabric’, ‘roads’ can be fairly easily identified by users, while categories such as ‘discontinuous urban fabric’ are more difficult to identify.

Crowdsourced data collected from Social Media can be a significant source of information - however, dealing with such data requires a great deal of consideration. Social data is high in volume and constantly increasing; often duplicated, incomplete, imprecise and potentially incorrect; informal (short, unedited and conversational) and less grammatically bounded text; generally concerning the short-term zeitgeist and covering every conceivable domain. These characteristics make automating intelligence gathering task difficult and the DP3 aimed at understanding the actionable information that can be collected from social media. The applicability of the data varies widely on the use case and as a result, different sources provide more contextually relevant information than others. For example, in flood emergency scenarios, Twitter has resulted the most relevant channel to achieve information, not only as primary source provided by the users, but also as indirect way to access to other social data contents, through the sharing of other information channels (eg. news, institutions and public bodies providing usually more relevant information as compared to citizens sharing information). Especially videos and images shared on YouTube and Twitter provide immense help in understanding scenarios on the ground. Anyway, in order to exploit such datasets, there needs to be strategies in place to deal with missing information, as lack of geotagging. For instance, a very small fraction of Twitter data (< 1% of analysed Tweets) actually contained geolocation information and in most cases, the Tweets were positioned outside the area of interest. Moreover, it has to be considered a mechanism to manage the redundancy of the information and to limit the effort due to the analysis of social data. 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.

## 7 Discussion and Recommendations

Rapid changes in modern communications and mobile devices have made citizen participation in science much easier and widespread (e.g., [25]). Large numbers of volunteers can be recruited over wide geographical areas to collect, submit, and interpret data at low cost [26]. Such widespread data collection (potentially over extended temporal periods) would be simply infeasible without citizen volunteered information. Indeed, such geographical spread is essential to understanding the processes behind many of the most important global challenges of today: vegetation loss, climate change, natural resource management, migration patterns, etc. In addition to geographical coverage, the volume of observation data (satellite-, airborne- and land-based) - some of which can only be interpreted by humans - is constantly growing. The ability to crowdsource detailed, high-resolution annotations of such data facilitates timely scientific analysis and decision making. Citizen science has evolved from a hobby through to serious science and is rapidly becoming a preferred approach to conducting large scale research [27].

The advantages of crowdsourcing extend beyond traditional scientific endeavour; the ability to mobilise a large group of volunteer data analysts (for example, performing satellite imagery annotation) can have significant humanitarian and emergency response benefits. For example, Tomnod crowdsources object and place annotation from satellite imagery to assist in major emergencies ranging from mapping a drought-stricken area of Somalia for the UNHCR to searching for the missing Malaysia Airlines plane MH370. The FP7 SPACE Project GEO-PICTURES crowdsourced geo-referenced in-situ images for the purpose of improving flood assessment from Radar EO Images [28].

Each type of citizen science and crowdsourcing initiative relies on a different degree of engagement from the citizen (and hence a different degree of engagement on behalf of the organiser - usually an NGO or research organisation). Similarly, the geographic extent of the initiative has strong implications for the suitability of particular forms of citizen science that can be employed - none is excluded at any one particular scale but organisational and governance overheads can become problematic.

Citizen science and the wider active participation of citizens in science and governance will continue to grow. The benefits are mutual. Citizens can enjoy making a positive impact at local, national or international levels in a variety of domains (e.g., environment, ecology, medical research, etc.). Professional researchers and decision makers can tap into an unprecedented wealth of knowledge and expertise or simply leverage the size of the crowd to collect large amounts of data across large geographical areas - all much more quickly and at must lower costs compared to traditional research approaches.

At the core of opportunities to EO 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.

Based on the findings from the demonstration projects and stakeholder analysis [24] of their understanding of how the field will evolve in the next few years, a set of recommendations are proposed:

### *Funding Support and Embedment*

Exploitation of citizen science and crowdsourcing should be a priority in future funding calls on EO in order to support existing initiatives and foster uptake in a larger scale. This will also help drive research into various facets of crowdsourcing such as engagement, security, privacy, user experience etc. Attention should also be focused on supporting wider set of participants including users not proficient in technology. Endeavours by all funders are essential to support citizen engagement and science learning at all ages. Initiatives that can help citizens relate to their own interests can have a greater acceptance and engagement, thereby fostering continued collaboration with citizens. This is a highly challenging problem and much work is needed in understanding how different forms of crowdsourcing can be used in different scenarios and settings.

### *Understanding*

A wider understanding of the benefits of crowdsourcing could inspire new strategies and initiatives. Organisations currently exploiting crowdsourcing could provide guidelines and standards on how best to collect, analyse and re-use citizen science data (including metadata, data protection, ethics and privacy). Providing easy access and means of visualisation of the data generated by contributors can provide them with an immediate observable feedback as well as provide a sense of accomplishment. Furthermore, it is important to co-design iterative solutions with user communities to ensure a continued interest and shared ownership among user communities.

### *Outreach and Communication*

Sharing the benefits, potential and opportunities of citizen science with different communities can be an excellent avenue for connecting with the public. One of the potential avenues could be to explore introducing citizen science within school curriculum with differentiating content and communication based on target age, starting at the very first classes in school. Team building activities based on citizen science could be used as means for developing interest in citizen-driven collaborations. Hackathons and fab labs could be another means of increasing general interest, while inspiring younger people and enthusiasts to explore new avenues to bring disruptive ideas to market.

### *Widening Participation*

Citizen science, in general is still not representative of all in society and this unequal participation, in many cases could have inherent biases based on gender, ethnicity, age and socioeconomic backgrounds. While efforts are being made to increase participant diversity, there is still much work involved in engaging large numbers of participant communities. A strong societal bias in STEM engagement and a strong bias in addressing societal and community issues could potentially result in having an incomplete and inaccurate understanding of all communities in the public. While focusing on individual communities and target groups can be helpful to provide a set of highly engaged users, there is a significant risk of alienating other pool of users and communities.

The recommendations are indeed provided as ways to start designing tools and technologies to engage citizens. The field of crowdsourcing and citizen science is highly complex that requires an in-depth understanding of users and their expectations. It is important to carefully consider all such aspects in order to build innovative business models that can leverage different approaches to engage with user communities.

**Acknowledgments** This research was supported by the European Space Agency funded Crowd4Sat initiative. The authors would also like to sincerely thank all the stakeholders and organizations for their inputs and interesting ideas. We would also like to thank the EU FP7 funded

WeSenseIt project (Grant agreement No. 308429) and EU H2020 funded Seta project (Grant agreement No. 688082).

## References

[1] Haklay, M., 2013, Citizen Science and Volunteered Geographic Information – overview and typology of participation in Sui, D.Z., Elwood, S. and M.F. Goodchild (eds.), 2013. Crowdsourcing Geographic Knowledge: Volunteered Geographic Information (VGI) in Theory and Practice . Berlin: Springer. Pp 105-122 DOI: 10.1007/978-94-007-4587-2_7

[2] H.E. Roy et al. Understanding Citizen Science and Environmental Monitoring. Technical report. 2012.

[3] Goodchild, M.F. Citizens as sensors: The world of volunteered geography. GeoJournal 2007, 69, 211–221.

[4] Elwood, S. Volunteered geographic information: Future research directions motivated by critical, participatory, and feminist GIS. GeoJournal 2008, 72, 173–183.

[5] Longueville, B.D.; Luraschi, G.; Smits, P.; Peedell, S.; Groeve, T.D. Citizens as sensors for natural hazards: A VGI integration workflow. Geomatica 2010, 64, 41–59.

[6] Gill, A.; Bunker, D. Crowd Sourcing Challenges Assessment Index for Disaster Management. Available online: https://pdfs.semanticscholar.org/7725/5b503bdab5b82627fa3e801042ef81bbd669.pdf (accessed on 15 May 2017).

[7] Fritz, S., Fonte, C. C., & See, L. (2017). The Role of Citizen Science in Earth Observation. Remote Sens. 9(4), 357.

[8] Irwin, A. (1995). Citizen science: A study of people, expertise and sustainable development. Psychology Press.

[9] Bonney, R., Cooper, C. B., Dickinson, J., Kelling, S., Phillips, T., Rosenberg, K. V., & Shirk, J. (2009). Citizen science: a developing tool for expanding science knowledge and scientific literacy. BioScience, 59(11), 977-984.

[10] Cohn, J. P. (2008). Citizen science: Can volunteers do real research?. BioScience, 58(3), 192-197.

[11] Dickinson, J. L., Zuckerberg, B., & Bonter, D. N. (2010). Citizen science as an ecological research tool: challenges and benefits. Annual review of ecology, evolution and systematics, 41, 149-72.

[12] Paulos, E., Honicky, R., & Hooker, B. (2008). Citizen science: Enabling participatory urbanism. Urban Informatics: Community Integration and Implementation.

[13] Science Communication Unit, University of the West of England, Bristol (2013). Science for Environment Policy In depth Report: Environmental Citizen Science. Report produced for the European Commission DG Environment, December 2013. Available at: http://ec.europa.eu/science-environment-policy

[14] Howe, J. (2006). The rise of crowdsourcing. Wired magazine, 14(6), 1-4.

[15] Audubon Society (2017). Christmas Bird Count. Retrieved 13 April, 2017, http://www.audubon.org/conservation/science/christmas-bird-count.

[16] Haklay, M. (2015). Citizen science and policy: a European perspective. The Wodrow Wilson Center, Commons Lab.

[17] Lintott, C. J., Schawinski, K., Slosar, A., Land, K., Bamford, S., Thomas, D., ... & Murray, P. (2008). Galaxy Zoo: morphologies derived from visual inspection of galaxies from the Sloan Digital Sky Survey. Monthly Notices of the Royal Astronomical Society, 389(3), 1179-1189.

[18] Lang, A. S. I. D., & Rio-Ross, J. (2011). Using Amazon Mechanical Turk to transcribe historical handwritten documents. The Code4Lib Journal, 15.

[19] Nguyen, T. B., Wang, S., Anugu, V., Rose, N., McKenna, M., Petrick, N., ... & Summers, R. M. (2012). Distributed human intelligence for colonic polyp classification in computer-aided detection for CT colonography. Radiology, 262(3), 824-833.

[20] Yetisgen-Yildiz, M., Solti, I., Xia, F., & Halgrim, S. R. (2010, June). Preliminary experience with Amazon's Mechanical Turk for annotating medical named entities. In Proceedings of the NAACL HLT 2010 Workshop on Creating Speech and Language Data with Amazon's Mechanical Turk (pp. 180-183). Association for Computational Linguistics.

[21] Cuff, D., Hansen, M., & Kang, J. (2008). Urban sensing: out of the woods. Communications of the ACM, 51(3), 24-33.

[22] Maisonneuve, N., Stevens, M., & Ochab, B. (2010). Participatory noise pollution monitoring using mobile phones. Information Polity, 15(1, 2), 51-71.

[23] Copernicus Programme Land Monitoring Service. Retrieved 15 May, 2017, http://land.copernicus.eu/pan-european/corine-land-cover/clc-2012

[24] S. Mazumdar, S.N. Wrigley and F. Ciravegna (2017). Citizen science and crowdsourcing for earth observations: an analysis of stakeholder opinions on the present and future. Remote Sensing 9(1) 87-107.

[25] E. Paulos, R. J. Honicky, and B. Hooker. “Citizen Science: Enabling Participatory Urbanism”. In: Handbook of Research on Urban Informatics: The Practice and Promise of the Real-Time City (2009), pages 414–436.

[26] Jeff Howe. “The rise of crowdsourcing”. In: Wired magazine 14.6 (2006), pages 1–4.

[27] Kathleen Toerpe. “The Rise of Citizen Science”. In: The Futurist 47.4 (2013). URL: http://www.wfs.org/futurist/2013-issues-futurist/julyaugust-2013-vol-47-no-4/rise-citizen-science

[28] Validating Space Observations for Flooding with Crowdsourcing In-Situ Observations by ANSUR. Technical report 30. 2012.