A New Generation of Tools for Crowdsourcing and Citizen Science to Support Earth Observation Calibration and Validation

Linda See, Steffen Fritz, Eduardo Dias, Elise Hendriks, Bas Mijling, Frans Snik, Piet Stammes, Fabio Vescovi, Gunter Zeug, Pierre-Philippe Mathieu, Yves-Louis Desnos and Michael Rast

***Abstract*—Citizens are providing vast amounts of georeferenced data in the form of in-situ data collection as well as interpretations and digitization of Earth Observation (EO) datasets. These new data streams have considerable potential for supporting the calibration and validation of current and future products derived from EO. Referred to as crowdsourcing and citizen science among many other terms, we provide a general introduction to this growing area of interest and review existing crowdsourcing and citizen science initiatives of relevance to EO. We then draw upon our own experiences to provide case studies that highlight different types of data collection and citizen engagement, and discuss various barriers to adoption. Finally, we highlight opportunities for how citizens can become part of an integrated EO monitoring system in the framework of the EU Space program including Copernicus and other monitoring initiatives.**

***Index Terms*—Crowdsourcing, citizen science, Earth Observation, in-situ data, ground truthing, Sentinel, Copernicus**

I. Introduction

Citizens have a long history of being involved in scientific research or what has more recently been referred to as ‘citizen science’ [1]. One of the main drivers behind the recent proliferation of citizen science projects has been technological, i.e. interactivity made possible via Web 2.0 resulting in a thriving culture of social media, movement towards the Internet of Things through smart sensors, and GPS-enabled mobile devices. These components have made it possible for citizens to become environmental sensors, collecting and analyzing information on a massive scale that also has real scientific value. There are many successful examples of citizen science that have led to new scientific discoveries such as new knowledge about protein structures [2] and discovering new galaxies [3], as well as websites for public reporting of illegal logging / deforestation [4] and illegal waste dumping [5]; these have demonstrated how citizens can have a visible impact upon the environment and local governance.

Another significant development has been the opening up of satellite imagery for viewing purposes through providers such as Google Earth and Bing. This has given citizens access to vast volumes of spatial data about the entire world. This trend continues as we move into the ‘Sentinel Era’ of big data where access to the data is truly open. Complementing the vast amounts of information already being collected via Sentinel 1, which is part of the European Space Agency’s (ESA) Copernicus program, there are several planned Sentinel missions in the near future. Other initiatives include the Biomass mission planned for 2020, and the 131 satellites that will be launched in 2015 by Planet Labs [6], where there will be an open access data model. All of these missions will require greater volumes of calibration and validation data.

The collection of ground truth for remotely-sensed products has traditionally been undertaken by experts. However, with tightening budgets and an explosion in the volume and frequency of data acquisition, new sources need to be considered, particularly those from citizens. Citizen science and crowdsourcing represent considerable opportunities to support data collection for Earth Observation (EO). At the same time, citizen involvement can promote EO more widely through awareness raising and education, which is often a secondary but fundamental goal of many citizen science projects.

In addition to vast quantities of data from space, citizen science and crowdsourcing, i.e. the involvement in citizens in tasks such as data collection, are also generating big data (especially from social media such as geotagged photographs), particularly in terms of the frequency and the variety of data sources. For this reason, crowdsourced data present numerous challenges, e.g. managing large volumes of data from diverse

This paper was submitted on 20 Jun 2015. This work was supported by the EducEO project, which is funded by the European Space Agency (ESA). Additional funding by the Zurich Flood Resilience Alliance is gratefully acknowledged.

L. See and S. Fritz are with the International Institute for Applied Systems Analysis, Laxenburg, Austria (e-mail: see@iiasa.ac.at; fritz@iiasa.ac.at).

E. Dias is with Geodan, Amsterdam, Netherlands. (e-mail: eduardo.dias@geodan.nl).

E. Hendriks, formerly of the Royal Netherlands Meteorological Institute (KNMI), and F. Snik are with Leiden University, Netherlands (email: hendriks@strwleidenuniv.nl; snik@strw.leidenuniv.nl)

Bas Mijling and Piet Stammes are with KNMI (email: bas.mijling@knmi.nl; stammes@knmi.nl)

F.D. Vescovi is with Airbus Defense and Space, Farnborough, UK (email: Fabio.Vescovi@astrium.eads.net)

G. Zeug is with Terranea UG, Germany (email: gunter.zeug@terranea.de).

P.-P. Mathieu, Y.-L. Desnos and M. Rast are with the European Space Agency (ESA), Frascati, Italy (email: pierre.philippe.mathieu@esa.int; yves-louis.desnos@esa.int; michael.rast@esa.int).

inputs, how to ensure data quality, how to build up communities and motivate participants and how to ensure sustainability of crowdsourcing activities. There is already a growing body of literature on many of these issues from both ecology/conservation [7]–[9] and the geographic literature (e.g. [10]–[13]). Overcoming these challenges will be critical if the data collected by citizens are to become a serious and rigorous input to support EO in the future.

The aim of this paper is to highlight the potential of citizen science and crowdsourcing for calibration and validation of EO in the context of current and future big data streams from space. In the first section we provide a general introduction to crowdsourcing and citizen science and review existing initiatives of relevance to EO. We then provide a few case studies of best practice from our own experiences and discuss barriers to adoption including issues such as data quality, bias and the digital divide. Finally, we highlight opportunities for how citizens can become part of an integrated EO monitoring system in the framework of Copernicus and other monitoring initiatives.

II. CROWDSOURCING, CITIZEN SCIENCE AND EO

A. Definitions

Citizen science (CS) can be defined as the involvement of the wider public in scientific research, from data collection to research design [14], [15]. The term CS first appeared in a book of the same name by Irwin in 1995, where Irwin expressed it as the idea of local knowledge to complement knowledge from more scientific sources. Around the same time, Rick Bonney of the Cornell Lab of Ornithology used the term as a synonym for public participation in scientific research (PPSR) [16].

PPSR and CS are only two terms of many that have appeared in the literature to describe the same basic phenomenon in which citizens have been involved in carrying out some type of task. Another commonly used term is crowdsourcing, which was coined by Howe [17]. Combining the words ‘crowd’ and ‘outsourcing’, it literally means to outsource tasks to the crowd. Within Howe’s definition is also the idea that this model has value for businesses and is therefore often used in a more commercial sense. Crowdsourcing platforms such as Amazon Mechanical Turk represent one mechanism in which businesses can find low cost labor to carry out a range of micro-tasks but it may also be an inexpensive source of data [18].

Volunteered geographic information (VGI) has emerged from the geographical literature [19] with a main focus on citizens as sensors, gathering spatially referenced data and providing this voluntarily. In VGI, end-users contribute geographic information to augment and replace existing sources of information such as printed maps, remotely-sensed images, and other web content. We also make the distinction between active data collection where participants go out and take measurements versus more passive data collection from social media or where sensors are connected to the internet and data are automatically collected, e.g. data from amateur weather stations. A primary benefit of VGI is that users are often more familiar with local geographic conditions and might contribute local geographic information more often and faster than governmental mapping organizations. End-users may therefore be better at detecting changes in their local environments. One of the most famous examples of VGI is OpenStreetMap (OSM), a community mapping initiative to provide open and free access to basic mapped features. Goodchild [20] notes that geo-registration errors between authoritative sources and non-authoritative sources are often similar while a number of papers have shown reasonable positional accuracy between OSM and authoritative data [21], [22]. Moreover, the Royal Institute of Chartered Surveyors [23] proposed the use of mobile phones for crowdsourcing of land tenure information in developing nations as a way to help build regional or national Land Administration Systems. Finally, VGI offers the potential of much improved spatial resolution compared to satellite data.

Other related terms can be found in the literature, e.g. user-generated content, neogeography, collaborative mapping, the GeoWeb 2.0, etc., all of which have commonalities with CS, crowdsourcing and VGI. However, in this paper we will simply refer to both CS and crowdsourcing in recognition of the fact that not all projects involving geospatial data collection by the crowd are driven purely from a scientific perspective.

B. EO-related CS and Crowdsourcing Projects

There are a number of CS and crowdsourcing projects that involve the use of EO data; a selection of these have been summarized in Table 1, organized by the type of data collected that could be of relevance to EO. The type of data are not based on any predefined categories but are an attempt to describe sites that collect similar data. In most cases the primary purpose is not for calibration or validation of EO although there are exceptions, e.g. Geo-Wiki and View-IT were developed specifically for validation of land cover, the Precipitation ID Near the Ground (PING) project collects rainfall for radar calibration while the ForestWatchers project allows citizens to mark areas of deforestation and correct results from automatic classification of imagery. However, data collected by the other projects might potentially be used for calibration and validation depending upon the distribution of the data collected. For example, a number of sites are concerned with the collection of weather and cloud data, which could be used to verify data from geostationary or weather satellites. Other sites involve the collection of data on land cover or involve image interpretation, which could be used in calibration and validation of land cover products. The same is true of sites that involve the collection of geotagged photographs, both at a given point in time or over time to monitor landscape change. Photographs from the Degree Confluence project have already been used for validation of land cover in the past [24]. Note that generic sites for collecting georeferenced photos such as Flickr, Panoramio and Google Earth have not been specifically listed in Table 1 but these photographs could also be used for calibration and

TABLE I
EXAMPLES OF RELEVANT CROWDSOURCING AND CITIZEN SCIENCE INITIATIVES TO EO

Data Collected	Crowdsourcing and CS Projects and Initiatives
Aerial imagery	Public Laboratory Balloon and Kite Mapping (http://publiclaboratory.org/tool/balloon-mapping)
Air pollution / quality, aerosols, noise pollution	CITI-SENSE (http://www.citi-sense.eu/), Omniscientis (http://www.omniscientis.eu/), iSpex (http://ispex.nl/en/), Noise Tube (http://noisetube.net/)
Clouds, sunspots, solar flares	The NOVA Labs: Cloud Lab and Sun Lab (http://www.pbs.org/wgbh/nova/labs) Students' Cloud Observations Online (S'COOL) (http://scool.larc.nasa.gov) SatCam (http://satcam.ssec.wisc.edu/)
Environmental data	Cobweb (Citizen Observatory Web) (http://cobwebproject.eu), Eye on Earth (http://www.eea.europa.eu/mobile) Global Learning and Observations to Benefit the Environment (GLOBE) (http://globe.gov) FieldScope (http://education.nationalgeographic.com/education/programs/fieldscope)
Extreme events, hurricanes, earthquakes, landslides	iSeeChange: The Almanac (http://www.thealmanac.org/), SkyWarn (http://www.skywarn.org/) Did You Feel It? (http://earthquake.usgs.gov/earthquakes/dyfi/), Did You See It? (http://landslides.usgs.gov/dysi/form.php)
Flood levels and extent	WeSenseIt (http://www.wesenseit.eu/)
Forest fires	Mount Diablo Fire Monitoring http://nerdsfornature.org/monitor-change/diablo.html
Fracking	Frack Finder (http://crowd.skytruth.org/)
Geotagged photographs of landscapes	Field Photo Library (http://www.eomf.ou.edu/photos/), Degree Confluence Project (http://www.confluence.org/)
Humanitarian and crisis response, disaster mapping	Humanitarian OpenStreetMap (http://hot.openstreetmap.org/), MicroMappers (http://www.micromappers.com/ International Network of Crisis Mappers (http://crisismappers.net/), Ushahidi (http://www.ushahidi.com/) TOMNOD (http://www.tomnod.com/nod/challenge/) The Digital Humanitarian Network (http://digitalhumanitarians.com/)
Land cover validation	Geo-Wiki (http://www.geo-wiki.org), VIEW-IT (Clark and Aide (2011))
Light pollution	Dark Sky Meter (http://www.darkskymeter.com/) Dark Skies ISS (http://crowdcrafting.org/app/darkskies/, http://www.citiesatnight.org/)
Mapped features (buildings, roads, POIs, land cover, land use, etc.)	OpenStreetMap (http://www.openstreetmap.org), Google Map Maker (http://www.google.com/mapmaker) The National Map Corps (http://nationalmap.gov/TheNationalMapCorps/), Wikimapia (http://wikimapia.org/)
Northern lights	Aurorasaurus (http://aurorasaurus.org/)
Ocean water colour	Citclops (http://www.citclops.eu/)
Phenology	USA National Phenology Network (https://www.usanpn.org/)
Photosynthesis	Public Laboratory Infrared Camera (http://publiclaboratory.org/tool/near-infrared-camera)
Precipitation / snow depth	CoCoRaHS: Rain, Hail, Snow Network) (http://cocorahs.org/), Rainlog.org (http://rainlog.org) Tracking Climate in Your Backyard (http://www.priweb.org/outreach.php?page=citizenscienceed/TCYIB) Precipitation ID Near the Ground (PING) (http://www.nssl.noaa.gov/projects/ping/) Snow Tweets http://www.snowtweets.org/,
Radioactivity of the oceans	Our Radioactive Ocean (http://ourradioactiveocean.org/helpus.html)
Tree/forest cover, deforestation, biomass	Deforestation Mapping in Canada (https://cfsnet.nfis.org/deforestation/), Forest Watchers (http://forestwatchers.net/), Treezilla (http://treezilla.org/), Urban Forest Map (http://urbanforestmap.org), Urban Tree Survey (http://www.nhm.ac.uk/urban-tree-survey), EarthWatchers (http://dfa.tigweb.org/)
Water quality and biodiversity	FieldScope (http://education.nationalgeographic.com/education/programs/fieldscope)
Weather	Weather Underground (http://www.wunderground.com/), WOW (http://wow.metoffice.gov.uk/), Citizen Weather Observer Program (http://www.wxqa.com/)

validation of land cover or the extraction of land use, which is an emerging area of research [25].

Some of the initiatives listed in Table 1 are quite generic and are focused on map making and cartography, e.g. OpenStreetMap, Wikimapia, and Google Map Maker, but they use EO data as part of online digitization of features. There are sites collecting different types of environmental data, e.g. water quality, air quality, light pollution, etc., which could be used to validate satellite-derived products that measure different environmental parameters.

The worldwide Global Learning and Observations to Benefit the Environment (GLOBE) program is an example of a CS initiative for environmental science which began in 1995. The program aims to increase environmental awareness by actively involving students in science. An essential part of the program is that the students perform measurements that are of research quality and report their observations to archives designed for the study of the Earth. Another goal is to generate public outreach for EO satellite missions.

One of the environmental parameters measured in the framework of the GLOBE Program is air pollution in terms of aerosols, by measuring the Aerosol Optical Thickness (AOT), an indicator for the amount of aerosol in the atmospheric column. Since 2001, students have measured AOT with the aid of a small and easy to handle sun photometer especially designed for GLOBE. Apart from creating awareness of aerosols and their role in climate and air quality, GLOBE sun photometer AOT measurements can be of significant value for the validation of AOT by satellites [26], [27]. GLOBE also has a partnership with NASA, where GLOBE campaigns in 2014 were aimed at validating data from the Soil Moisture Active Passive (SMAP) mission, measuring precipitation for comparison with data from the Global Precipitation Measurement (GPM) satellite mission, collecting data about clouds for comparison with data from the CloudSat mission and in-situ data collection for use by the CALIPSO mission.

Another key theme for crowdsourcing and EO initiatives is in disaster response. In the direct aftermath of a disaster, the collected data can be used to provide a better operational picture and situation awareness. This was e.g. achieved in the aftermath of the Haiti earthquake, when volunteers analysed and mapped incoming text messages. Through the aggregation of individual reports, the crisis mappers were able to identify clusters of incidents and urgent needs, helping responders target their response efforts. Another large group of volunteers built the most complete map possible for the affected areas using satellite imagery donated by DigitalGlobe, GeoEye, the World Bank and the US Government. Since the events in Haiti in 2010, several digital volunteering networks have been set up, e.g. the Digital Humanitarian Network, the International Network of Crisis Mappers and the Humanitarian OpenStreetMap Team (HOT). These volunteer networks collect, analyse, map and disseminate data in the form of text, blogs, video, pictures, and maps [28]. Besides harvesting the web for meaningful information, they are using dedicated tools for analysing tweets, photos, and satellite imagery. Software tools such as Ushahidi provide a mechanism for pulling in information from multiple sources and for mapping out the situation while TOMNOD uses the crowd to find different features on Digital Globe imagery that can aid in post-disaster management.

EO is also widely used for post-disaster damage assessment, e.g. using geotagged field photographs for verification. In the aftermath of typhoon Haiyan, a strong tropical cyclone that hit and devastated parts of the Philippines in November 2013, a community of volunteers used the MicroMappers tool to evaluate damages visible on photos posted to social media. All photos classified as having “severe damage” were geotagged and published in a live crisis map. Damages exist which are difficult or not identifiable from satellite imagery and related maps cannot provide the level of information needed for the quantification of damage intensity [29]. Crowdsourcing can therefore provide the relevant ground truth.

Overall, the initiatives and projects outlined in Table 1 highlight that many EO-relevant initiatives exist but that most are not currently being utilized for calibration and validation of EO datasets. In the next section, three case studies are provided that illustrate different ways in which CS and crowdsourcing have been used together with EO, which are based on our experiences. The main challenge is scaling up these efforts to collect larger quantities of data and to expand the spatial outreach.

III. Case Studies

A. *The iSPEX Project*

Aerosol measurements from professional ground-based stations suffer from lack of spatial coverage. Satellite measurements provide spatial coverage but an insufficient temporal resolution, often providing only one measurement per day. The iSPEX project is an example of a large-scale CS project that was developed as a way of collecting aerosol data at a much higher spatiotemporal resolution. An inexpensive add-on to a smartphone that uses a multitude of built-in phone functionalities and a simple protocol for taking the measurements has enabled wide-scale citizen participation in providing crucial information about atmospheric particulates, which affect air quality and influence our climate (Fig. 1).

Fig. 1. The iSPEX device (left) and a demonstration of its use (right).

In 2013 experiments were run on three separate days in the Netherlands in which more than 10K measurements were collected. The quality of the measurements when averaged over many devices at the same location was extremely high, proving that measurements from an army of iSPEX observers can complement professional measurements from the ground [30]. The iSPEX measurements were also collected across the Netherlands and compared with data from MODIS, showing good correspondence in the spatial patterns. Thus large amounts of measurements with sufficient resolution in time and space can provide data that are complementary to products derived from EO. Moreover, the upward-looking iSPEX measurements can be used to calibrate and validate satellite products that need complex corrections for ground scenes.

Future plans include European wide experiments in aerosol data collection via the iSPEX add-on and further citizen involvement. The enthusiasm of the Dutch volunteers to take the measurements has been one of the main factors in the success of this CS project. Motivating larger crowds at a European wide scale will be a future challenge.

B. *The Geo-Wiki Project*

Geo-Wiki was established in 2009 as a tool for involving citizens in determining whether global land cover maps accurately characterize the Earth’s surface based on Google Earth imagery [31]. Since then there have been a number of crowdsourcing competitions, which have collected around

250K pixels of land cover classifications; these have been used to both train and validate land cover maps (e.g. [32], [33]). More recently, a Geo-Wiki game called Cropland Capture (Fig. 2) resulted in the collection of around 4.5 million pixels from 3,000 players over a 6 month period [34]. These data are currently being used to develop a hybrid global cropland map for 2010.

Fig. 2. The Cropland Capture game in which players were asked whether they could see evidence of cropland in the images and photographs.

As part of the ESA-funded EducEO project, a new game called Picture Pile has been developed, which uses many of the same game mechanics as Cropland Capture. This game broadens image classification to any type of land cover organized into different piles of pictures for interpretation (Fig. 3). Picture Pile has the potential to become a new generic tool in the ESA toolbox, which will be devoted to validation of Sentinel and other imagery. The idea will be to develop a large community of citizen 'Sentinel Truthers', similar to the communities that have been built up by the Zooniverse project for the classification of galaxies.

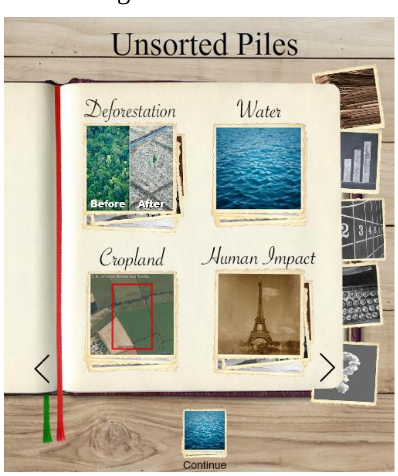

Fig. 3. Examples of different piles in the Picture Pile game corresponding to different themes

Quality control will be implemented through checking citizen performance against expert controls and providing the same image to more than one person to interpret. The first version of the game will be launched in October 2015, with a focus on deforestation.

C. *The Earthwatchers Project*

Geodan set up the Earthwatchers project as a way of engaging students in the near-real time detection of deforestation in Borneo while simultaneously teaching them about remote sensing. Embedded within a high school level course, students are charged with keeping watch on an area, and as near-real time Landsat imagery is processed and made available, they alert the system when changes are detected. These changes are then investigated on the ground by NGOs (non-governmental organizations) working closely with the local communities. Such a joined up approach has resulted in the successful detection of illegal deforestation and real action on the ground while raising awareness among school populations living in areas far removed from these illicit activities. The information collected through the Earthwatchers system also provides validation data for change detection algorithms.

As part of the EducEO project, the Earthwatchers application is being extended to time series of Sentinel 1 SAR imagery (Fig. 4) to determine whether this data stream can be used to identify deforestation, particularly in areas that are prone to persistent cloud cover. Part of the exercise will test the effectiveness of different renderings of SAR data, i.e. true colour, false colour, vegetation indices or radar, in comparison with Landsat 8 imagery. We expect that the students will find patterns and issues that might never be uncovered using automated methods alone and will provide valuable validation data. At the same time, the students will learn to analyze SAR imagery and understand its potential as part of their mainstream education.

Fig. 4. The Earthwatchers interface showing the areas adopted by the students and the Sentinel 1 imagery.

IV. Challenges and Barriers to Adoption

There are a number of key challenges associated with crowdsourcing and CS projects, which have relevance to the use of the data for supporting EO. These are discussed below.

A. Data Quality

Crowdsourced geospatial data production is typically an open, lightly controlled process with few constraints, specifications or quality assurance processes. This is in contrast to the authoritative geospatial data production practices of national mapping and cartography agencies (NMCAs) and businesses, which are typically less flexible and more controlled [35]. The potential massive use of these production methods has been a concern, especially to government organizations and academic institutions, due to differences in the methods of production. However, EO-based mapping projects are nearly always lab-based and lack field checking. Therefore, in some cases, crowdsourced data may be the only option for ground-based calibration and validation. Moreover, experiences from the crowd could help to inform EO data producers about product quality and fitness for purpose, which could help to market EO products or services in the future.

ISO 19113, which is an international standard for reporting on the quality of geographic data, proposes a set of data quality descriptors including: completeness; logical consistency; positional accuracy; temporal accuracy; and thematic accuracy. This also should form a structured approach for considering quality assurance protocols in a crowdsourcing context. There are a number of different ways that quality can be controlled for. Where possible, automated quality assurance methods should be used, e.g. when individuals contribute data that fall outside an allowable range. When this is not possible, then one of the most common approaches is to use majority agreement in which more than one citizen is given the same task; this is essentially drawing upon the wisdom of the crowd [13], [36].

Another approach is to use control data or 'gold standard' data, which are results from experts against which the crowd can be compared. Controls are only possible for a small sample but this can provide information about the performance of individuals and can be used to put different weights on their individual contributions when used for subsequent purposes. Peer review is another good mechanism that can be applied to improve data quality. Answers from individuals can be rated by others or discussion forums can be used to discuss individual answers, which additionally provide a learning process for contributors. Volunteers often self-organize into hierarchies of expertise, e.g. in OSM and Wikipedia, where more experienced individuals provide advice or mentoring to less experienced contributors.

A number of studies have examined the thematic quality of image classifications from Geo-Wiki [10], [37], [38], which showed varying levels of performance across contributors and across land cover types. However, one of the issues with this dataset was not having sufficient classifications at the same location by multiple contributors in order to develop statistically robust relationships between contributor performance, land cover type and other factors such as image resolution, location, etc. In the Cropland Capture game, this was rectified and each image classification has multiple answers to help understand the quality of the contributions by the crowd.

For low-cost sensors, the data from a single unit is often not accurate enough, but through averaging over many sensors, accurate data can still be obtained as shown in the iSPEX project. To reach a desired accuracy, the cost of a device times the number of devices (N) required often takes the form of sqrt (N). The collection of large quantities of data also permits outlier analysis, and intrinsic statistical analysis.

When considering social media as a potential source for crowdsourcing, rumors and intentional misinformation must be considered. Especially during a crisis, unconfirmed reports tend to spread rather quickly on social media. Very often manipulated photographs are circulating through the internet which makes it difficult finding and verifying accurate information. However, to make full use of social media they must be verified. One attempt to overcome that is the Verily platform (https://veri.ly/) for the verification of social media. Volunteers are asked to answer a question, e.g. related to a photo or a statement. The answer must be justified by providing evidence either in the form of an image, video or as text. It is again the use of majority agreement and the wisdom of crowds which is used for verification.

B. Representativeness of the Data

With CS projects and in particular crowdsourcing and VGI, there is a bias in the spatial distribution of the data collected. For example, in OSM, there is higher completeness in urban compared to rural areas [39]. The same is true of photos contributed through sites such as Flickr and Panoramio. There are exceptions to this, e.g. the Degree Confluence project collects stories and photos in four directions at each intersection of latitude and longitude, while Geo-Wiki campaigns have created samples for image interpretation that are relevant to each research question posed. However, the representativeness of the data has implications for how it can be used in other types of research since a particular sampling strategy has not been imposed *a priori*. The representativeness of the data may have less effect on calibration activities but are more critical for validation, where a stratified sample based on land cover type might be employed for validating land cover maps. This may become less of an issue for desk-based image interpretation tasks when the geographical collection of data is sufficiently dense that sub-samples of the data could be extracted and bootstrapping could be used to examine the uncertainty across samples. This is only possible in an age of crowdsourcing and CS, as traditional validation exercises would not have had the luxury of having sufficient in-situ data available. The development of an open online land cover validation tool and validation data repository through the LACO-Wiki project is one step towards realizing this goal [40].

C. Recruitment, Motivation and Incentives for Participation and Sustainability

Recruitment of participants is needed to raise awareness of a project's existence, which begins with identification of the target audience, e.g. school children vs amateur astronomers. The promotion and recruitment process should then be customized for the audience using a variety of media channels, e.g. TV, radio, print, various types of social media and advertisement in specialist publications. Launch events can be effective ways for scientists to meet the public while word-of mouth recruitment from existing participants is a powerful means of recruiting new volunteers in an ongoing project.

Understanding motivation is also a critical prerequisite to developing successful CS and crowdsourcing projects and is an area of active research [8], [9], [41]–[44]. Based on these studies (which are situated mostly in the fields of ecology, biology and nature conservation), a summary of different volunteer motivations includes: the desire to learn more about the science behind the project; helping the environment; getting to know other people with similar interests and as a way to make new friends; feeling like an active participant and co-owner of the project; relevance to the community; ability to see the impact of their work, e.g. visualization of their data collection efforts or further use within a scientific or policy application; and gaining recognition for their input, e.g. through feedback and interaction with scientists and peers, and through gaining achievements, e.g. progression to expert status or from simple to more complex tasks requiring additional responsibility. Where possible, tasks should also be fun and participation should be made as easy as possible, minimizing technical, logistical, legal and intellectual barriers.

Motivation is also clearly linked to maintaining participation in the longer term to ensure sustainability of the project should this be desired, e.g. some tasks may be completed after a finite period of time. Sustainability refers to retaining a community as well as the services or apps that have been developed as part of a project. In terms of retaining a community, methods that have been shown to work include giving rapid feedback to participants and providing regular communication about their contributions. Volunteers like the idea of knowing that their work is important and that their contributions can help scientists make better and more comprehensive analyses. They also like to see their contributions, e.g. if you upload a new track to OSM, you will see the change reflected very quickly. Rewarding citizen scientists is another effective way to encourage and support participation, e.g. by providing participants with certificates of recognition, providing access to the data, providing different levels of progression or reputation ranking (e.g. used by eBay and TripAdvisor), gamefication to introduce an element of competition between participants, different kinds of prizes, and inclusion on scientific publications.

In the Geo-Wiki project, different prizes such as Amazon vouchers and electronics have been used as incentives as well as co-authorship on papers [32], [45]. In the Earthwatchers and guardians applications, there were no extrinsic motives, but the users could alert local authorities when detecting illegal deforestation, giving them the feeling of empowerment and effectiveness of their contributions. In the iSPEX project, it was noted that people who are intrinsically motivated are the most valuable. They commit to the project by buying the iSPEX add-on for a couple of Euros. These participants keep on measuring because they like contributing to science and in the end the data may help to improve air quality. They also contribute because it is fun. For this reason, iSPEX units will not be free anymore. There is also instant qualitative feedback on the measurements, and the measurement appears on a live map with all the others, which creates a sense of community.

D. The Digital Divide and Other Inequalities

Individuals are motivated by different drivers and these differ across communities and across different demographic groups. Some communities are excluded and identifying the barriers to participation is important for finding solutions to widening participation. For example, certain age groups have lower rates of household internet penetration and mobile phone usage so online CS and crowdsourcing activities will not reach this age group as effectively as others. Technology is also a potential barrier due to the cost as well as the ability to use it effectively, which will also be partly determined by socio-economics. Certain ethnic groups have been underrepresented within CS projects in the USA and there has been an overrepresentation by more affluent groups [8]. Barriers may also be a result of language because many CS projects and websites are in English. Such a language barrier can make it difficult to manage a large-scale project that is meant to be geographically distributed. In the future, as technology and the Internet of Things become even more prevalent, barriers to participation may be significantly lowered; however, other inequalities may still persist and should be considered.

E. Data Issues: Protocols, Interoperability/Standards, Legal Issues and Data Privacy

Many CS projects have established data collection protocols, particularly those in the fields of ecology, biology and nature conservation. However, this is less true of VGI and projects such as OSM, where contributors typically operate without central coordination or strict data collection frameworks. Moreover, the GPS in mobile phones may not be accurate enough to meet the minimum data specifications of some NMCAs. Despite some of these known limitations, if crowdsourced data are to be used alongside authoritative sources of data, then they must meet certain quality standards. Data collection protocols or templates are one way of helping to ensure that these standards are met. Project design will inevitably involve trade-offs between achieving scientific goals, e.g. gathering comprehensive, high quality data according to rigorous scientific protocols, and the ease of data collection. If the data collection is too complex or too time consuming, volunteers often lose their desire to participate and thus understanding and adapting the program to the skills, expectations and interests of the volunteers is critical. Training is another way of helping to ensure more accurate data collection, e.g. [46] found that the accuracy of data points collected in OSM improved when a collaboration was set up between OSM and the US Geological Survey, which provided feedback to volunteers on quality.

There are issues related to the use of data from social media platforms as the data, which may often need to be scraped, cannot be archived, curated or made available for re-use. This represents a barrier to reproduction of scientific results, which makes it difficult to evaluate research that uses data from these sources.

Data interoperability and the use of standards are other issues that become relevant when crowdsourced data are shared between distributed systems. Unlike authoritative data, crowdsourced geospatial data do not normally have metadata that conform to specific standards. Kalantari et al. [47] have recently proposed a new metadata standard for VGI, called Geospatial Metadata 2.0. The INSPIRE Directive may provide some useful guidance, particularly for CS and crowdsourcing initiatives in the EU. Long term preservation and curation of the data are also issues related to sustainability and require a data management plan.

Finally, addressing legal issues and data privacy are key challenges for CS and crowdsourcing, which are complicated by laws that differ from country to country, particularly when the data are collected in one country and then stored in the cloud within another country. Cho [48] outlines a series of legal concerns with the use of geospatial information from crowdsourcing. Ownership of the data can be difficult to trace when there are many contributors to the database or what is referred to as the 'wiki' effect [49]. Contributions from the crowd that infringe copyright, e.g. from third party providers, must be dealt with through removal of the material and should be part of an agreement between the user and the provider to shield providers from potential damages. The risk of liability is another legal issue that needs to be considered, especially in those situations where the crowdsourced data could lead to negligence. To protect against liability, crowdsourcing sites should insist that users accept the terms of use of their data. At the same time, they must be responsible for ensuring data quality. Finally, when crowdsourced data are collected and hosted on a cloud computing platform, there are a number of additional issues that must be addressed such as security, unauthorized use of the service and the protection of personal and confidential data.

When data are collected on the ground e.g. through a mobile application, users leave a spatial footprint. This can mean a significant impact on one's privacy. For example, the risk of providing location information through Twitter and websites like http://geosocialfootprint.com/ is described in [50], which could be used to determine a person's place of work and home and therefore abused not only for criminal activities.

Another data-related issue is that smartphone cameras are different across models and platforms, and they have different mechanical interfaces. This makes it difficult to create a generic solution, e.g. for the iSPEX project. However, this situation is improving for electronic interfaces to smartphones, which are being regulated by the EU.

V. Toward an Integrated EO Monitoring System

As stated in the introduction, one of the clearest ways in which CS and crowdsourcing can support EO is through the calibration and validation of satellite missions. Some projects are already actively supporting calibration and validation activities. For example, Geo-Wiki has collected a considerable amount of calibration and validation data using high resolution imagery. The data have been used to create new land cover products [45], [33], [51]) and to validate others [52], [32]. To highlight this potential in a scientific context, the use of crowdsourced data in [45], [33] and [51] resulted in improvements in overall accuracy of around 5% compared to existing products. More importantly, they improved upon the spatial distribution of land cover. In terms of validation, [52] and [32] provide examples of how good quality validation data can be collected in short periods of time and provide policy-relevant inputs, e.g. in [32], the validation exercise resulted in a downgrading of estimates of global land availability for biomass by as much as a magnitude over previous estimates.

The ongoing LACO-Wiki project will develop a new online validation tool for land cover that is designed for use by NMCAs, regional mapping agencies, researchers and students, where the validation data will be gathered into a single repository that could be used for other calibration and validation activities. This 'expert'-sourced database could complement ones being collected from the crowd. Other examples are SatCam and S'COOL, which are projects that gather information on clouds and are used to validate satellite products from MODIS and other NASA missions, and iSPEX, which gathers data on aerosols that can be used to validate MODIS products.

There is a great deal of potential in using mobile phones for in-situ data collection. There are existing apps such as Geo-Wiki Pictures, which allow users to collect georeferenced photos of the landscape and indicate the land cover type. There are also a number of generic data collection apps appearing, e.g. EpiCollect and GeoODK, where the latter is currently being used to validate remotely-sensed drought indicators on the ground. There are many current biodiversity-related CS projects that collect information on habitats, and there is a considerable amount of land cover and land use now being collected by OSM, which could also be used for calibration and validation purposes. Not only would this benefit current EO datasets but it has considerable potential for Sentinel 2. A GMES Masters prize for 2012 was awarded to a Norwegian company, AnsuR, who have developed the ASIGN app for the ground-based verification of SAR flood mapping products.

Soil moisture from EO needs much more validation, particularly in areas like Africa where there are few ground-based measurement stations. Low cost soil moisture sensors are now available and have the potential to validate soil moisture products in data sparse areas. This is relevant not only for current EO data from ASCAT and SMOS but also new data coming from Sentinel 1.

Opportunities exist in the collection of crowdsourced biomass measurements using customized apps such as the Relasphone. Recent developments in radar devices for mobile phones could mean that by the time the Biomass satellite is launched in 2020, 3D measurements by smart phones will be commonplace. Using the crowd for large-scale biomass measurements could then become a real possibility. These developments also have great potential for urban mapping.

On-going developments of small and cheap instruments for environmental trace gases further increase the potential of CS for air pollution data collection in support of EO. Satellite measurements are essential to monitor the day-to-day variation, and geographical distribution of air pollution. Crowd sourcing of air pollution on the ground, in addition to a very limited number of other professional ground-based observations, could be of enormous value for the validation of the satellite measurements and add detail to the global satellite scale.

NO2 is the main short-lived air pollutant that can be observed by satellite. The possibility of a citizen science-based NO2 trace gas detection can therefore be of large benefit to satellite missions for air pollution, such as for the Tropheric Monitoring Instrument (TROPOMI) aboard Sentinel 5P (to be launched early 2016) or its predecessor Ozone Monitoring Instrument (OMI) aboard NASA's AURA satellite. These satellites measure trace gases including NO2 and aerosols. CS-oriented sensors such as the Air Quality Egg (http://airqualityegg.com) are available but have not been investigated yet in comparison with scientific instrumentation. Other sensors for testing air quality are currently being trialed in the CITI-SENSE project (see Table 1), which may result in low-cost sensors that could be rolled out to citizens in the future.

All of these aforementioned opportunities require significant buy-in from the producers of EO derived-products. The data contributed by citizens must be integrated into the workflows of product development. Instead of using only professional sources of calibration and validation data, mindsets regarding the value of CS and crowdsourcing need to change. The onus lies partly with individual CS and crowdsourcing projects that must demonstrate that scaling up is feasible and can generate good quality data for a wide geographical area and at a temporal resolution that is fit-for-purpose. EO map producers need to be willing to experiment with data from CS and crowdsourcing projects, but this requires recognition of the opportunities and investment. The EducEO project is one positive step towards bringing producers and CS together, but it represents only the start of a new, collaborative process toward an integrated EO monitoring system that includes citizens as a valuable component.

References

[1] A. Miller-Rushing, R. Primack, and R. Bonney, "The history of public participation in ecological research," *Frontiers in Ecology and the Environment*, vol. 10, no. 6, pp. 285–290, Jul. 2012.

[2] F. Khatib, F. DiMaio, F. C. Group, F. V. C. Group, S. Cooper, M. Kazmierczyk, M. Gilski, S. Krzywda, H. Zabranska, I. Pichova, J. Thompson, Z. Popović, M. Jaskolski, and D. Baker, "Crystal structure of a monomeric retroviral protease solved by protein folding game players," *Nature Structural & Molecular Biology*, vol. 18, no. 10, pp. 1175–1177, 2011.

[3] D. Clery, "Galaxy Zoo volunteers share pain and glory of research," *Science*, vol. 333, no. 6039, pp. 173–175, Jul. 2011.

[4] A. Nayar, "Model predicts future deforestation," *Nature News*, Nov. 2009.

[5] G. Milčinski, "The rise of crowd-sourcing: how valuable data can we get out of VGI," presented at the CAPIGI, Amsterdam, Netherlands, 2011.

[6] R. Taylor, "The mini-satellite revolution gives new views of Earth," *BBC News*, 2014. [Online]. Available: http://www.bbc.com/news/technology-27422596. [Accessed: 18-Jun-2015].

[7] K. Crowston and A. Wiggins, "Supporting citizen involvement in scientific research," presented at the HICSS 2011 SCI Workshop, 2011.

[8] R. E. Pandya, "A framework for engaging diverse communities in citizen science in the US," *Frontiers in Ecology and the Environment*, vol. 10, no. 6, pp. 314–317, Aug. 2012.

[9] J. C. Tweddle, L. D. Robinson, M. J. O. Pocock, and H. . Roy, "Guide to citizen science: developing, implementing and evaluating citizen science to study biodiversity and the environment in the UK," Natural History Museum and NERC Centre for Ecology & Hydrology, 2012.

[10] A. Comber, L. See, S. Fritz, M. Van der Velde, C. Perger, and G. Foody, "Using control data to determine the reliability of volunteered geographic information about land cover," *International Journal of Applied Earth Observation and Geoinformation*, vol. 23, pp. 37–48, Aug. 2013.

[11] G. M. Foody and D. S. Boyd, "Using volunteered data in land cover map validation: Mapping tropical forests across West Africa," in *Geoscience and Remote Sensing Symposium (IGARSS), 2012 IEEE International*, 2012, pp. 6207–6208.

[12] M. F. Goodchild and L. Li, "Assuring the quality of volunteered geographic information," *Spatial Statistics*, vol. 1, pp. 110–120, May 2012.

[13] M. Haklay, S. Basiouka, V. Antoniou, and A. Ather, "How many volunteers does it take to map an area well? The validity of Linus' Law to volunteered geographic information," *The Cartographic Journal*, vol. 47, no. 4, pp. 315–322, 2010.

[14] R. Bonney, "Citizen science: A lab tradition," *Living Bird*, vol. 15, no. 4, pp. 7–15, 1996.

[15] R. Bonney, C. B. Cooper, J. Dickinson, S. Kelling, T. Phillips, K. V. Rosenberg, and J. Shirk, "Citizen science: A developing tool for expanding science knowledge and scientific literacy," *BioScience*, vol. 59, no. 11, pp. 977–984, Dec. 2009.

[16] R. Bonney, H. Ballard, R. Jordan, E. McCallie, T. Phillips, J. Shirk, and C. C. Wilderman, "Public Participation in Scientific Research: Defining the Field and Assessing its Potential for Informal Science Education," Center for Advancement of Informal Science

Education (CAISE), Washington DC, A CAISE Inquiry Group Report, 2009.
[17] J. Howe, “The rise of crowdsourcing,” *Wired Magazine*, vol. 14, no. 6, pp. 1–4, 2006.
[18] M. Buhrmester, T. Kwang, and S. D. Gosling, “Amazon’s Mechanical Turk A New Source of Inexpensive, Yet High-Quality, Data?,” *Perspectives on Psychological Science*, vol. 6, no. 1, pp. 3–5, Jan. 2011.
[19] M. F. Goodchild, “Citizens as sensors: the world of volunteered geography,” *GeoJournal*, vol. 69, no. 4, pp. 211–221, Nov. 2007.
[20] M. Goodchild, “NeoGeography and the nature of geographic expertise,” *Journal of Location Based Services*, vol. 3, no. 2, pp. 82–96, Jun. 2009.
[21] M. Haklay, “How good is volunteered geographical information? A comparative study of OpenStreetMap and Ordnance Survey datasets,” *Environment and Planning B: Planning and Design*, vol. 37, pp. 682–703, 2010.
[22] S. Zheng and J. Zheng, “Assessing the completeness and positional accuracy of OpenStreetMap in China,” in *Thematic Cartography for the Society*, T. Bandrova, M. Konecny, and S. Zlatanova, Eds. Springer International Publishing, 2014, pp. 171–189.
[23] Royal Institute of Surveyors, “Crowdsourcing Report,” Royal Institute of Surveyors, London, Research Report, 2011.
[24] K. Iwao, K. Nishida, T. Kinoshita, and Y. Yamagata, “Validating land cover maps with Degree Confluence Project information,” *Geophysical Research Letters*, vol. 33, no. 23, p. n/a–n/a, 2006.
[25] J. Estima and M. Painho, “Flickr geotagged and publicly available photos: Preliminary study of its adequacy for helping quality control of Corine land cover,” in *Computational Science and Its Applications – ICCSA 2013*, B. Murgante, S. Misra, M. Carlini, C. M. Torre, H.-Q. Nguyen, D. Taniar, B. O. Apduhan, and O. Gervasi, Eds. Springer Berlin Heidelberg, 2013, pp. 205–220.
[26] D. R. Brooks and F. M. Mims, “Development of an inexpensive handheld LED-based Sun photometer for the GLOBE program,” *J. Geophys. Res.*, vol. 106, no. D5, pp. 4733–4740, Mar. 2001.
[27] K. F. Boersma and J. P. de Vroom, “Validation of MODIS aerosol observations over the Netherlands with GLOBE student measurements,” *Journal of Geophysical Research*, vol. 111, no. D20, Oct. 2006.
[28] M. Laituri and K. Kodrich, “On line disaster response community: People as sensors of high magnitude disasters using Internet GIS,” *Sensors*, vol. 8, no. 5, pp. 3037–3055, May 2008.
[29] G. Lemoine, C. Corbane, C. Louvrier, and M. Kauffmann, “Intercomparison and validation of building damage assessments based on post-Haiti 2010 earthquake imagery using multi-source reference data,” *Natural Hazards and Earth System Sciences Discussions*, vol. 1, no. 2, pp. 1445–1486, Apr. 2013.
[30] F. Snik, J. H. H. Rietjens, A. Apituley, H. Volten, B. Mijling, A. Di Noia, S. Heikamp, R. C. Heinsbroek, O. P. Hasekamp, J. M. Smit, J. Vonk, D. M. Stam, G. van Harten, J. de Boer, C. U. Keller, and 3187 iSPEX citizen scientists, “Mapping atmospheric aerosols with a citizen science network of smartphone spectropolarimeters,” *Geophys. Res. Lett.*, vol. 41, no. 20, p. 2014GL061462, 2014.
[31] S. Fritz, I. McCallum, C. Schill, C. Perger, L. See, D. Schepaschenko, M. van der Velde, F. Kraxner, and M. Obersteiner, “Geo-Wiki: An online platform for improving global land cover,” *Environmental Modelling & Software*, vol. 31, pp. 110–123, May 2012.
[32] S. Fritz, L. See, M. van der Velde, R. A. Nalepa, C. Perger, C. Schill, I. McCallum, D. Schepaschenko, F. Kraxner, X. Cai, X. Zhang, S. Ortner, R. Hazarika, A. Cipriani, C. Di Bella, A. H. Rabia, A. Garcia, M. Vakolyuk, K. Singha, M. E. Beget, S. Erasmi, F. Albrecht, B. Shaw, and M. Obersteiner, “Downgrading recent estimates of land available for biofuel production,” *Environ. Sci. Technol.*, vol. 47, no. 3, pp. 1688–1694, Feb. 2013.
[33] L. See, D. Schepaschenko, M. Lesiv, I. McCallum, S. Fritz, A. Comber, C. Perger, C. Schill, Y. Zhao, V. Maus, M. A. Siraj, F. Albrecht, A. Cipriani, M. Vakolyuk, A. Garcia, A. H. Rabia, K. Singha, A. A. Marcarini, T. Kattenborn, R. Hazarika, M. Schepaschenko, M. van der Velde, F. Kraxner, and M. Obersteiner, “Building a hybrid land cover map with crowdsourcing and geographically weighted regression,” *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 103, pp. 48–56, 2015.
[34] L. See, T. Sturn, C. Perger, S. Fritz, I. McCallum, and C. Salk, “Cropland Capture: A gaming approach to improve global land cover,” in *Proceedings of the AGILE’2014 International Conference on Geographic Information Science*, Castellon, Spain, 2014.
[35] M. T. Rice, F. I. Paez, A. P. Mulhollen, B. M. Shore, and D. Caldwell, “Crowdsourced Geospatial Data: A Report on the Emerging Phenomena of Crowdsourcing and User-generated Geospatial Data,” US Army Engineer Research and Development Center, Alexandria, VA, Annual Report BAA #AA10-4733, Contract #W9132V-11-P-0011, 2012.
[36] J. Surowiecki, *The Wisdom of Crowds*. New York: Anchor Books, 2005.
[37] G. M. Foody, L. See, S. Fritz, M. Van der Velde, C. Perger, C. Schill, and D. S. Boyd, “Assessing the accuracy of volunteered geographic information arising from multiple contributors to an internet based collaborative project,” *Transactions in GIS*, vol. 17, no. 6, pp. 847–860, 2013.
[38] L. See, A. Comber, C. Salk, S. Fritz, M. van der Velde, C. Perger, C. Schill, I. McCallum, F. Kraxner, and M. Obersteiner, “Comparing the quality of crowdsourced data contributed by expert and non-experts,” *PLoS ONE*, vol. 8, no. 7, p. e69958, Jul. 2013.
[39] P. Neis, D. Zielstra, and A. Zipf, “Comparison of volunteered geographic information data contributions and community development for selected world regions,” *Future Internet*, vol. 5, no. 2, pp. 282–300, Jun. 2013.
[40] L. See, J. Weichselbaum, S. Fritz, C. Perger, M. Hofer, C. Hoffman, R. Grillmayer, and C. Dresel, “Tools for Land Cover Validation – Integrating LAnd COver VALidation

(LACOVAL) with Geo-Wiki," presented at the Sentinel-2 for Science Workshop, ESRIN, Frascati, Italy, 2014.

[41] B. Bruyere and S. Rappe, "Identifying the motivations of environmental volunteers," *Journal of Environmental Planning and Management*, vol. 50, no. 4, pp. 503–516, 2007.

[42] J. L. Dickinson, J. Shirk, D. Bonter, R. Bonney, R. L. Crain, J. Martin, T. Phillips, and K. Purcell, "The current state of citizen science as a tool for ecological research and public engagement," *Frontiers in Ecology and the Environment*, vol. 10, no. 6, pp. 291–297, Aug. 2012.

[43] H. . Roy, M. J. O. Pocock, C. D. Preston, D. B. Roy, J. Savage, J. C. Tweddle, and L. D. Robinson, "Understanding Citizen Science & Environmental Monitoring," NERC Centre for Ecology & Hydrology and Natural History Museum, Final Report on behalf of UK-EOF, 2012.

[44] H. A. Van Den Berg, S. L. Dann, and J. M. Dirkx, "Motivations of adults for non-formal conservation education and volunteerism: Implications for programming," *Applied Environmental Education & Communication*, vol. 8, no. 1, pp. 6–17, Jun. 2009.

[45] L. See, I. McCallum, S. Fritz, C. Perger, F. Kraxner, M. Obersteiner, U. Deka Baruah, N. Mili, and N. Ram Kalita, "Mapping cropland in Ethiopia using crowdsourcing," *International Journal of Geosciences*, vol. 4, no. 6A1, pp. 6–13, 2013.

[46] S. Jackson, W. Mullen, P. Agouris, A. Crooks, A. Croitoru, and A. Stefanidis, "Assessing completeness and spatial error of features in volunteered geographic information," *ISPRS International Journal of Geo-Information*, vol. 2, no. 2, pp. 507–530, Jun. 2013.

[47] M. Kalantari, A. Rajabifard, H. Olfat, and I. Williamson, "Geospatial Metadata 2.0 – An approach for Volunteered Geographic Information," *Computers, Environment and Urban Systems*, vol. 48, pp. 35–48, Nov. 2014.

[48] G. Cho, "Some legal concerns with the use of crowd-sourced Geospatial Information," *IOP Conference Series: Earth and Environmental Science*, vol. 20, p. 012040, Jun. 2014.

[49] T. Scassa, "Legal issues with volunteered geographic information," *The Canadian Geographer / Le Géographe canadien*, vol. 57, no. 1, pp. 1–10, Mar. 2013.

[50] G. Friedland and R. Sommer, "Cybercasing the joint: On the privacy implications of geotagging," in *Proceedings of the Fifth USENIX Workshop on Hot Topics in Security (HotSec 10)*, Washington, D.C., 2010.

[51] S. Fritz, L. See, I. McCallum, L. You, A. Bun, E. Moltchanova, M. Duerauer, F. Albrecht, C. Schill, C. Perger, P. Havlik, A. Mosnier, P. Thornton, U. Wood-Sichra, M. Herrero, I. Becker-Reshef, C. Justice, M. Hansen, P. Gong, S. Abdel Aziz, A. Cipriani, R. Cumani, G. Cecchi, G. Conchedda, S. Ferreira, A. Gomez, M. Haffani, F. Kayitakire, J. Malanding, R. Mueller, T. Newby, A. Nonguierma, A. Olusegun, S. Ortner, D. R. Rajak, J. Rocha, D. Schepaschenko, M. Schepaschenko, A. Terekhov, A. Tiangwa, C. Vancutsem, E. Vintrou, W. Wenbin, M. van der Velde, A. Dunwoody, F. Kraxner, and M. Obersteiner, "Mapping global cropland and field size," *Glob Change Biol*, vol. 21, no. 5, pp. 1980–1992, May 2015.

[52] S. Fritz, L. You, A. Bun, L. See, I. McCallum, C. Schill, C. Perger, J. Liu, M. Hansen, and M. Obersteiner, "Cropland for sub-Saharan Africa: A synergistic approach using five land cover data sets," *Geophys. Res. Lett.*, vol. 38, no. 4, p. L04404, Feb. 2011.