

Citizen-Based Sensing of Crisis Events: Sensor Web Enablement for Volunteered Geographic Information

Sven Schade^a, Laura Díaz^b, Frank Ostermann^a, Laura Spinsanti^a, Gianluca Luraschi^a, Simon Cox^c, Manoli Nuñez^b, Bertrand De Longueville^d

^a *European Commission – Joint Research Centre, Institute for Environment and Sustainability, Ispra, Italy*

^b *Universitat Jaume I – Institute of New Imaging Technologies, Castellón, Spain*

^c *CSIRO Earth Science and Resource Engineering, Perth, Australia*

^d *Earth and Life Institute - Environmental Sciences, Université Catholique de Louvain, Louvain-la-Neuve, Belgium*

phone: +39 0332 78 6536

fax: +39 0332 78 6325

email: sven.schade@jrc.ec.europa.eu

URL: <http://inspire-forum.jrc.ec.europa.eu/pg/profile/svenschade>

Note. An initial shorter version of this work was presented as a full paper at WebMGS 2010, 1st International Workshop on Pervasive Web Mapping, Geoprocessing and Services, Como, Italy. The authors were invited to submit an extended version to this special issue. Apart from editorial changes, we included a new section on the sensing of multiple VGI platforms and an extensive discussion section on generalizing the approach towards a nervous system for the next generation of geospatial information infrastructures (Digital Earth). Pointers to related works have been extended, too.

Abstract. There is a long tradition of non specialists contributing to the collection of geo-referenced information. Thanks to recent convergence of greater access to broadband connections, the availability of Global Positioning Systems in small packages at affordable prices, and more participative forms of interaction on the Web (Web 2.0) vast numbers of individuals became able to create and share Volunteered Geographic Information (VGI). The potential of up to 6 billion persons to monitor the state of the environment, validate global models with local knowledge, contribute to crisis situations awareness, and provide information that only humans can capture is vast and has yet to be fully exploited. Integrating VGI into Spatial Data Infrastructures (SDI) is a major challenge, as it is often regarded as insufficiently structured, documented or validated according to scientific standards. Early instances of SDIs used to have limited ability to manage and process geosensor-based data (beyond remotely sensed imagery snapshots), which tend to arrive in continuous streams of real-time information. The current works on standards for Sensor Web Enablement (SWE) aim to fill this gap. This paper shows how such standards can be applied to VGI, thus converting it in a timely, cost-effective and valuable source of information for SDIs. By doing so, we extend previous efforts describing a workflow for VGI integration into SDI and further advance an initial set of VGI Sensing and event detection techniques. Examples of how such VGI Sensing techniques can support crisis information system are provided. The presented approach serves central building blocks for a Digital Earth's nervous system, which is required to develop the next generation of (geospatial) information infrastructures.

Keywords. Sensor Web Enablement, Spatial Data Infrastructure, Volunteered Geographic Information, VGI Sensing, Events

API	Application Programming Interface
O&M	Observations and Measurements (Standard)
OGC	Open Geospatial Consortium
OS	Open Search
SDI	Spatial Data Infrastructure
SensorML	Sensor Modeling Language
SES	Sensor Event Service
SOS	Sensor Observation Service
SPS	Sensor Planning Service
SWE	Sensor Web Enablement
VGI	Volunteered Geographic Information

1 Introduction

Since Web 2.0 provided Internet with colloquial read-and-write functionality, the quantity of digital information accessible online is growing at an overwhelming rate. As a consequence, scientists are faced with a ‘data tsunami’ from which it is increasingly arduous to extract valuable information (Shah et al. 2010). When this online information created by users has a geospatial reference, it is known as Volunteered Geographic Information (VGI – Goodchild 2007).

Our work contributes to the advance of *VGI Sensing*, an emergent research field, which aims at designing a set of standards and techniques to streamline geo-referenced contents published online by citizens as a valuable and timely source of spatio-temporal information (De Longueville et al. 2010). Indeed such techniques are necessary to harness the potential of up to 6 billion humans to monitor the state of the environment, contribute to situation awareness for crisis, validate global models with local knowledge, and provide information that only humans can capture (Goodchild 2007, Jones 2009).

Spatial Data Infrastructures (SDI) are expected to be increasingly able to manage and process geosensor-based data (beyond remotely sensed imagery snapshots), which tend to arrive in continuous streams of real-time information (Annoni et al. 2010). The currently missing timely provision of information and methods for event notification are of particular importance to crisis management scenarios. The current works on standards for Sensor Web Enablement (SWE) are aiming to fill this gap (Botts et al. 2008).

Just as we readily accept the processing of satellite data as an input to many geospatial analyses, we should also aim to better interpret the abundant and freely available signals provided by citizens (De Longueville et al. 2009). A sensor web enablement of VGI would be a major step in that direction. This paper aims at further studying how SWE concepts and standards could be applied to VGI in order to convert it in a timely, cost-effective and valuable information source for SDIs. By doing so, we move towards a next generation of geospatial information infrastructures, or *Digital Earth*. VGI Sensing has already been characterized as a major building block for a Digital Earth’s nervous system (De Longueville et al. 2010b); it is complemented by classical physical sensor networks and environmental simulations (Schade and Craglia 2010).

The remainder of this paper is structured as follows. Related works are outlined in section 2. The process of streamlining VGI into SWE is detailed in section 3, in which we pay special attention to the available sources of VGI. Examples of VGI Sensing for crisis events are described in sections 4 and 5. The former concentrates on VGI Sensing from a single source, while the latter outlines the challenges that arise if sensing VGI gathered from multiple sources. Discussions on the generalization of the approach and of open research issues of VGI Sensing are presented in section 6 before concluding and outlining future work items in section 7.

2 Related Work

The presented work is based on two notions: VGI and SWE. We introduce the concepts and underlying principles below. Pointers to additional readings are included.

VGI: a Great Potential to be Harnessed

The use of a hybrid approach that integrates bottom-up and top-down methodologies has been already demonstrated (Jankowski 2009), with the purpose of integrating user generated information, scientific tools and official information in the same geospatial infrastructure. In this context merging the top-down SDI model with VGI infrastructures have already been described Craglia 2007, Goodchild 2007, and Gould 2007. Out of the multiple available social media, the following are already frequently used for providing user-contributed and location-related content (VGI). Twitter¹ is a social networking and micro-blogging service. Its users can send and read text-based posts of up to 140 characters, so called tweets, which are publicly visible by default. Flickr² is an online application that allows uploading, store and organizing digital photographs. It enables the creation, management and retrieval of the pictures’ metadata, such as title, description, tags, and date, time and location the picture was taken. YouTube allows sharing videos that can be georeferenced. OpenStreetMap and Geonames are both explicit VGI platforms.

¹ <http://twitter.com/>

² <http://www.flickr.com/about/>

There is nowadays a growing consensus to recognize the role of VGI in support to crisis management activities. Numerous case studies stressed the added value of using VGI in various types of crisis events, such as earthquakes (De Rubeis et al. 2009), forest fires (De Longueville et al. 2009), political crises (Bahree 2008), hurricanes (Hughes and Palen 2009), floods (De Longueville et al. 2010), and terrorist attacks (Palen et al. 2009). Emerging internet projects are reducing the gap between data and system, to turn them into valuable resource for better understand our planet. Pachube³ for instance is a platform that helps to build, store, share and discover real-time sensors, energy and environmental data from things (objects and buildings) all around the world.

However, quality concerns may mitigate the enthusiasm VGI raises. Data quality has been recognized as a major concern (Elwood 2008) resulting in a lack of credibility. Flanagin and Metzger (2008) argued that the credibility issue of VGI is mostly due to the apparent lack of control of the data creation process. In addition, the same authors argue that in the data abundance context that characterizes VGI, traditional mechanisms that tend to increase trust in data, such as authoritative sources, well-established data creation methodologies and certified information gatekeepers, are ineffective.

Examples showed several possible strategies to overcome VGI's credibility challenge. Firstly, it is possible to reinforce the control on the production chain by establishing a standardized data creation method and by working with a limited number of well-trained volunteers (Lee 1994). Secondly, the quality control itself can be set up as a volunteered process, and the community of users can act as quality filters for VGI as can be found for Wikipedia (Bishr and Mantelas 2008). A third option could be to turn the challenge of data abundance into an opportunity, where reliable information is extracted from vast amounts of VGI with uncertain quality from numerous sources by applying cross-validation mechanisms. In other words, the data quality problem of VGI can be addressed by 'aggregating input from many different people' (Mummidi and Krumm 2008, p. 215), and by processing these VGI clusters to evaluate their relevance in a given context.

This third option is a key concept of VGI Sensing, a set of standards, methods and techniques designed to streamline geo-referenced contents published online by citizens as a valuable and timely source of spatio-temporal information (De Longueville et al. 2010b). This paper aims at contributing to this emerging field of research. The developed notion is inline with the idea of 'crowd sensing' (Siegele 2010), but notably differs from concepts, such as people-centric sensing (PCS – Campbell et al. 2008) and human-centered sensing (HCS – Jiang and McGill 2010), in which persons are considered as kind of sensing devices. In the VGI Sensing approach, on the contrary, peoples' reports on observations serve stimuli, which are topic of sensing. Compared to previous works, such as (Jiang and McGill 2010, Sakaki et al. 2010), the solution suggested in this paper considers not only a single VGI source, such as Flickr, Twitter, OpenStreetMap or Geonames, but includes means for combining information from different platforms.

Sensor Web Enablement

In order to improve interoperability between crisis management systems and sensor networks, the Open Geospatial Consortium (OGC) provides standards for web-based SWE (Botts et al. 2008) run through a series of 'interoperability test-beds' from 2002 to the present. SWE provides a well structured framework, it is based on open standards, and it has a growing user community.

The goal of SWE is to develop an architecture and supporting standards for distributed services related to sensors and observations. The key elements are:

- o Sensor Observation Service (SOS –Bröring et al. 2010), a web interface for requesting observation data;
- o Sensor Planning Service (SPS –Simonis and Echterhoff 2010), a web interface for tasking sensors;
- o Sensor Event Service (SES – Echterhoff and Everding 2008) allows clients to subscribe to events, i.e. enables push-based communication; it generalizes over the Sensor Alert Service (SAS – Simonis 2006);
- o Sensor Model Language (SensorML – Botts and Robin 2007), a model and encoding for describing sensors and sensor systems); and
- o Observations and Measurements (O&M – Cox 2010), a model and encoding for observations and their specific metadata.

SWE is domain- and discipline-neutral and was designed and tested for in-situ, ex-situ and remote observations. The O&M information model is based around the notion of an observation event, and scopes the operation signature of SOS, using the key terms *procedure*, *observed-property*,

³ <http://www.pachube.com/>

feature-of-interest and *result*. The values of any of these may be highly structured internally, and in many contexts be sets or arrays of more primitive elements. The separation of the feature-of-interest, observation (event) and result are the keys to enabling O&M to support the different use cases, and the appearance of an explicit observed property and feature of interest are the keys to observation semantics and cross-domain information discovery and fusion. Furthermore, as the observed-property should be related to the type of the feature-of-interest, a data processing chain is also connected to a sequence of transformations of these.

Expressing VGI as a SWE application is essentially a matter of mapping the elements of a SWE system to the notions listed above. Such work on a sensor web, which is based on human observations, is ongoing. Jürrens and colleagues for example propose SES-based filtering and SOS-based storage of user contributed content that is represented in O&M (Jürrens et al. 2009). We follow a similar approach, but emphasizing the added value for the chain of geospatial information processing.

3 Watching at VGI through SWE Goggles

VGI Sensing provides a novel way of approaching VGI management and processing. In this section, we explain how SWE contributes to the conceptualization and implementation of VGI Sensing. General clarifications are followed by a detailed description of the processing steps that are involved. The scenario considered in this paper is the use of VGI Sensing to support crisis event detection and characterization. The whole processing chain is thus designed to acknowledge the occurrence of a crisis event (a perdurant geospatial entity, such as a fire or flood). The measurement of VGI activity (VGI Sensing) is separated from the detection of such events.

Principle and Overview

As a central principle, we monitor flows of VGI in order to detect events. In contrast to a trivial interpretation of the ‘citizen as sensor’ metaphor, we do not consider the individual citizen as a sensor making measurements on the observed property, but as a foundational element of a (virtual) VGI sensor, where the actual *observed-property* is the flow of VGI harvested under pre-defined conditions. This involves processing vast amounts of VGI, and applying statistical methods in order to derive knowledge. For the moment, this can be considered in the same way as a remote sensing image is processed to translate the spectral signature and patterns of its pixels into geospatial knowledge.

Table 1 presents the central concepts of VGI Sensing and event extraction, as introduced by De Longueville et al. (2010b). We revise the notion of stimulus and sensor compared to the previous version of this table. The analogous process involving remote sensing is provided as a comparison. The various components described in this table have been created in analogy to human sensory system, but can be applied both for VGI and remote sensing.

Each transition between rows in the table represents a processing step in the ‘event detection with VGI Sensing’ workflow. Figure 1 gives an overview. Each step is described in the following subsections. The next two sections each provide an example processing workflow: the first using a single VGI platform (Flickr), and the second using multiple sources.

Step 1: from Stimulus to Signal

We put a virtual VGI sensor in place, that is, software observing the publicly available information on the web to perform measurements related to specific VGI activity. As we illustrate in the following two sections, those sensors may operate on a single VGI platform as well as on multiple platforms. Such VGI sensors may have diverse *features-of-interest*, e.g. the earth surface or a part of the earth surface. The *observed-property* may be the occurrence or the density of VGI. The sensor may be specialized to VGI items including specific tags. VGI sensors are thus designed to detect particular types of VGI items, just as satellite-mounted image sensors are sensitive to particular wavelengths at a particular spatial resolution.

This first step is characterized by web mining processed aiming at gathering VGI related to the feature of interest and observed properties. Such capabilities correspond to the encoding capabilities of the virtual VGI sensor (i.e. the observation procedure, which may be described using SensorML).

Step 2: from Signal to Sensation

Sensors transform signals; traditionally analogue stimuli to digital values. In our case, the transformation is an assignment of VGI items (coming from a particular VGI platform) in respect to a sensor specific 'grid'. This grid divides the geospatial region that is covered by the sensor. The allowed values for each grid cell depend on the selected *observed-property*. A sensation results in a grid of values representing counts or densities, which implements a SWE coverage model, just as n-dimensional satellite image does. However, the grid cells of a virtual VGI sensor may be of any shape and size (e.g. a grid of square 'pixels', but also administrative or natural boundaries such as catchments). The definition of the grid is part of the measurement *procedure*.

For each cell of the grid, specific calibration rules may be applied so that results are comparable even if important factors are expected to influence the amount of expected VGI for each (e.g. population density, technology penetration rate, cultural inclination to report on the Internet). Calibrations may be performed using SPS capabilities (see also section 2). Also the specification of this discretisation method is part of the measurement *procedure*. The restitution of harvested VGI as an organized set of measurements is the result of this second step. These organized observation *results* is represented in O&M. They may be served by a specialized SOS.

Steps 1 and 2 in combination define VGI Sensing. A possible mapping of VGI sensor characteristics to SWE is presented in Table 2; again we use the remote sensing analogy.

Step 3: from Sensation to Perception

Thanks to previous steps, vast amounts of heterogeneous VGI have been turned into an organized set of measurements. The next step is analogous to any signal processing. The grid of values provided by the virtual VGI sensor is analyzed in order to identify specific patterns (e.g. a significant raise of 'flood pictures' in the primary Donahue catchment or cluster of 'fire pictures' in adjacent grid cells). This can be compared to entity extraction in remote sensing. Just as the spectral signature that characterizes a satellite image pixel can be a rich source of information about the corresponding geographic area, analysis of contemporaneous VGI for a given grid cell (and of its neighbors) informs about the phenomenon of interest for this portion of the earth.

Ultimately, the goal of this step is to detect and characterize patterns from sensor's signal, spatiotemporal events (perdurant geospatial entity), for instance. The result of this step can be provided as part of (or close to) a SES.

Step 4: from Perception to Attention

This step aims at assigning levels of relevance for the events that have been identified. It allows to inform decision makers with the events that require most immediate attention and to filter irrelevant events. Depending on the constraint model, different decision makers may specify diverse conditions for notification. The raw VGI data that contributed to the detection of an event can be further analyzed at this stage. The analysis may include data from additional sources (e.g. land use, soil moisture, or weather forecast data for assessing potential severity of floods). For the events that fulfill any pre-defined condition, an alert can be triggered by a SES.

Step 5: from Attention to Action

At this step, VGI Sensing information is integrated in a decision support system, thus helping crisis managers to plan the appropriate actions. Acquisition of additional data, such as satellite imagery, can be part of such actions, thus emphasizing the complementarities between information sources and sensor types. Notably, the results/impact(s) of actions may again be observed by citizens, who create VGI. Loops can be performed in the context of situation awareness, early response, and damage assessment.

4 VGI Sensing of a Single Platform: A Flood Example

This section presents an illustrative example of event extraction based on VGI Sensing, with a focus on data transformation that occurs at each step the process. The aim is to detect and locate floods on a geographic zone, United Kingdom in this case. In the example, only a single VGI platform: Flickr is used as the source of VGI. For reasons of unpredictability, we present the example based on historic data instead of describing a real-time monitoring case. All illustrations can be directly projected. The example is further described in (De Longueville et al., 2010a).

Virtual VGI Sensor for Flood Pictures in the UK

Flickr offers numerous features that make it an interesting VGI platform. The first of them is the multiplicity of uploading options, which includes direct upload from camera-enabled mobile phones. Such devices are becoming common in the mass market and many of them also include built-in GPS sensor, so it is expected that Flickr will contain in the future a growing number of geo-referenced contents that will be available within seconds after a photo has been taken. The possibility for users to assign a location to pictures – ‘to geotag’ - is another important feature. Indeed, the wide majority of cameras do not include a GPS device that automatically inserts picture location in the image file metadata. Flickr users can thus manually add this information using an online map interface. Flickr allows users to associate keywords – called ‘tags’ - to their pictures.

In step 1, we collect information about pictures related to floods, such as the time and place where they have been taken. During the retrieval phase, queries are submitted through the Flickr Application Programming Interface (API), and their results are saved locally for further processing. The Flickr API offers numerous options to submit queries using the *flickr.photos.search* method. Research parameters can include the date the picture has been taken, the date it has been uploaded, portions of text to be searched in its title and description, the presence of one or several tags, the *id* of the group it belongs to, the *id* of the user that uploaded it, the place where it has been taken (bounding box or distance radius around a given location). Figure 2 gives an overview of the spatio-temporal distribution for geotagged pictures taken between 01/01/2007 and 31/03/2009, and related to floods retrieved from Flickr.

In step 2, the collected VGI items are captured using a 'grid'. The grid (including geospatial resolution), the temporal resolution, the *observed-property*, and the discretization *procedure* are designed to fit a given purpose just as sensor specifications are set up to address a pre-define use for this sensor.

A VGI item, to be captured, needs to be validated. The validation is a formal step to control if the minimal information required to process the data is available in the proper format. We defined a set of validation criteria: geographic extent (some of the pictures uploaded have clearly invalid latitude and longitude, e.g. equal to 0), valid temporal extend (e.g. date of publication or creation has to be valid and it has to be provided), and a picture has to be tagged.

Validated VGI items are then allocated to grid cells (i.e. a spatio-temporal segmentation of the retrieved VGI lot is performed). In this work our spatio-temporal segmentation method is based on three hypothesis: (i) *Geospatial pattern*: A flood is an event that occurs in a defined area, i.e. pictures that record a single flood event have specific spatial relationship; (ii) *Temporal pattern*: A flood is an event recorded in a discrete time period, i.e. flood pictures in the same area but at different times refer to different events; and (iii) *Event pattern*: A flood is an event that should be documented by significant images, i.e. the more there are people affected by the flood, and the more pictures will be uploaded on Flickr.

On this basis, we formulated three criteria we used to build relevant events:

- *Geospatial Criteria*: The geospatial segmentation can be based on a regular grid of cells (e.g. square pixels) or using appropriate territorial units. For detecting floods, polygons representing catchments are a logical choice.
- *Temporal Criteria*: The temporal segmentation is performed by creating time intervals, in accordance with the expected characteristics of the event of interest (e.g., a flood in Europe typically lasts several days – not seconds or months) and the expected sensitivity to temporal change of the system. Time slices can be created arbitrarily or with statistical methods. In this case, we used the Jenk's Natural Breaks (Jenks and Coulson 1963).
- *Event criteria*: a flood with bigger impact is documented by more citizens.

These three criteria allow are used as specifications for our VGI virtual sensor. Figure 3 is a graphical representation of the results obtained for our VGI virtual sensor for floods on Flickr, i.e. it represents the grid. The vertical dimension depicts the geographic segmentation (each position on the y-axis corresponding to a different catchment, ordered along a North-South axis) and the horizontal dimension depicts the temporal segmentation (each position on the x-axis corresponding to a time period defined in Step 2). The size of each bubble represents the amount of VGI retrieved for a given spatio-temporal cell (corresponding to the x and y values of its centre).

Detecting Flood Events as Patterns in the Grid and Assessing Relevance of Identified Flood Events

In step 3, the grid of data provided by the virtual VGI sensor can be analyzed in a way analogous to classifying an array of measurements. In the floods case, a particular care will be given to the

temporal distribution of VGI for each catchment separately (a peak in VGI flow corresponding most likely to a flood event), while analysis of geospatial distribution should take into account the hydrological relations (i.e. how catchments are connected to each other).

When a peak of VGI appears for a given spatio-temporal cell, further analysis can be performed in order to assess the likelihood this peak corresponds to an event of interest. For example, the semantic similarity between the tags associated with the photo and concepts associated to floods can be measured in order to establish a ranking of flood likelihood. In other words, the ranking reflects the likelihood the collected VGI pictures represents a *floods* (and not any other type of accumulation of water).

In order to implement step 4, a pre-defined alert threshold is set, which can be subject to calibration based on socio-economic factors related to the likelihood of presence of citizens with appropriate devices and willingness to report the event. When a rank value exceeds this threshold, an alert is triggered.

In our flood example, we look for VGI amounts exceeding a threshold pre-calibrated for each catchment, on Flickr. The ranking value is calculated by processing the picture tags and it can be used to reduce noise (i.e. by eliminating pictures that are most likely not corresponding to flood event or evaluate the probability that an event can be confused with another type of flood. Figure 4 shows the time series of VGI Sensing values for the Severn catchment, together with a possible alert threshold. In this case, alerts that would have been triggered correspond to two major flood events that took place between the 21st and the 30th of July 2007, and between the 15th and the 26th of January 2008 (source: the Dartmouth Floods Observatory⁴). The figure provides an example of the value added information we are seeking for. Similar timelines could be provided by real-time monitoring.

Creating Flood Alerts

Corresponding to step 5, the flood alert can be propagated from the SES to relevant authorities. In addition to the alert itself, that may have been triggered earlier by other means (such as flow measurements of rivers and forecast models) the collection of spatio-temporally indexed VGI (text, picture, videos) that is associated to the event can contribute to situation awareness and support the early response phase of the event.

5 VGI Sensing of Multiple Platforms: A Forest Fire Example

The case study presented here is part of larger research project on social media, VGI and crisis management. Details about the background and framework can be found in the literature (Ostermann et al. 2010, Spinsanti and Ostermann 2010). This case study aims to combine several sources of social media, and to investigate the potential utility of VGI on a broader scale, i.e. looking at an entire season of forest fires in one European country; France in this case. In the example, two VGI platforms, Twitter and Flickr, are used as the source of VGI. Again, the example is based on historic data.

VGI Sensing of Wild Fires in France

The harvesting of VGI (step 1) started with the forest fire season at 16th of July, and continued until 30th of September. It used a continuous stream of Tweets, that was collected by querying the public Twitter streaming API with a filter of wildfire-related keywords (such as 'fire', 'helicopter' and 'evacuation') in eight different languages. In total, 24.5 GB of data, equalling around 8 million Tweets, was collected. The second source of data was the online image sharing service Flickr. Using the same set of keywords for searching Flickr, meta-data for around 700 thousand images was retrieved. Around 1% of that data is already geo-coded.

In order to start the first iteration of the workflow with the best possible data, the data is narrowed down to smaller sample. First, a geographical region is focused by the case study. We decide for France (and therefore Tweets expressed in French), because an exploratory examination showed a better geographical focus than for other languages: English is used almost on a global scale, Italian and Spanish share some keywords and include South America, in Germany there were no major fires during the case study period, and Greek raises additional issues related to character encoding. The subset for the French keywords is further reduced by considering only those with the keyword

⁴ <http://www.dartmouth.edu/~floods/Archives/index.html>

'incendie' (wild fire) in their text, description, or tags. Finally, data without any explicit geocoding (in the form of coordinates) or implicit geo-coding (in the form of place names) is discarded. After checks for syntactical correctness, we enrich the data. At this stage of our work, this enrichment focuses on the geographical location in order to make more data available for subsequent analysis. For a very early simple geocoding, we submit the data without coordinates but with toponyms to Yahoo! Placemaker⁵. All these steps are summarized in Table 3.

The next step (step 2) is to prepare the spatio-temporal clustering in the VGI data, with the aim to compare any finding to the fires as reported by the European Forest Fire Information System (EFFIS)⁶. Depending on the parameter settings, we expect to find all EFFIS reported fires, but also to detect noise in the form of False Positives. The preparation consists of a qualitative and a quantitative phase. First, the data is examined qualitatively using a common off-the-shelf Geographic Information System by visualizing all VGI located in France (n=680), varying the scale and the symbology to distinguish various attributes, such as source, location, time, and type. Then, the data is converted into the format required by a clustering software⁷ and clustering parameters are defined. As in the flood example, there are geospatial, temporal and event criteria to limit cluster size (spatial, temporal, population) and location (geographic overlap between clusters). For this case study, we use two different settings for the clustering: (i) the default values and (ii) a set of modified parameters. The former include no restrictions on size, but prohibit spatial overlap. The latter includes a limit to the spatial extent of the clusters to 50 km, because it is sensible to assume that any relevant VGI is close to the fire event. This also allows a possible geographic overlap between clusters, because some fire events are geographically close. Ruling out geographic overlap would result in not detecting clusters that are close geographically, but distant in time.

Strictly speaking, this approach, which is required to address VGI coming from multiple social media (platforms), breaks the analogy between VGI Sensing and remote sensing. Instead of creating a grid of which each cell represents a measure of VGI flow (VGI occurrence, density of VGI etc.) we generate a point set, i.e. a vector file, of VGI items. We store this in a format that suites later clustering and prepare sensible (event specific) parameters for calibrating the clustering algorithm. We have to consider each VGI item separately, because we need to understand the interplay between single contributions across VGI platforms (Flickr and Twitter in this case). Creating a grid out of heterogeneous sources would not account for potential correlations. Still, we can apply the mapping to core SWE concepts:

- The *feature-of-interest* remains the Earth surface and its periphery, in the example this is restricted to France.
- The *observed-property* is the individual occurrence of a VGI item.
- The *procedure* includes the search criteria for each included social media/VGI platform, the post processing in order to prepare the format required by the clustering algorithms, and the set of particularly identified parameter values.
- The *result* of this VGI sensing is a complex measurement including separate sets of VGI items with point locations for each harvested platform. For this case this corresponds to two different point sets, one harvested from Twitter, the other from Flickr.

Detecting Wild Fire Events as Patterns of VGI and Assuring their Quality

Out of the different options for searching spatio-temporal clusters (step 3), two scan statistics can be applied to the type of point data at hand: a spatio-temporal permutation model (Kulldorff et al. 2005), or a Bernoulli model (Bertsekas and Tsitsiklis 2002). The latter requires a population distribution data which in our case would represent the number of possible providers of forest fire related VGI. The methodological difficulties associated with creating this data, like spatial granularity and choice of attributes, lead to a focus on the space-time permutation model, which only needs case data, i.e. the VGI data itself. There are several options for running the space-time permutation model, first and foremost whether to base the spatial location of the scanning windows on the cases or on other locations. From a conceptual point of view, this resembles the choice to look whether to detect clusters in the data without prior knowledge about possible events, or whether the known events are represented in the data. We use both methods in order to compare their results. As describes above, we use two different parameter sets for each. Thus, in total there are four sets of results. Each set consists of a number of likely clusters ($p \leq 0.0001$ estimated from 9999 Monte Carlo simulation runs), and the type of data that belongs to them.

⁵ <http://developer.yahoo.com/geo/placemaker/>

⁶ <http://effis.jrc.ec.europa.eu/>

⁷ <http://www.satscan.org/>

Considering step 4, the assignment of relevance to a recognised pattern, we examine the quality of our results by comparing the detected potential wild fires with forest fires that have been registered in EFFIS. Table 4 shows a comparison of the clusters detected with the 6 official fires registered in the EFFIS system and the clusters predicted by the 4 models. It reveals that the a-priori assumption was correct: depending on the parameters settings, all fires were detected but also a significant number of false positives were returned. As *True Positives* we count the number of recorded fires that are represented in the reported clusters. As *False Negatives* we count the number of recorded fires who have not been reported. Notably, the false negatives are probably due to sparse data, and that the large number of yet not geo-coded VGI may contain information on these. As *False Positives* we count the number of clusters that do not coincide with a recorded fire as reported in the EFFIS data. Here, it is highly possible that there has been an actual fire that was not included in the official data because it was too small (below 20 hectares), or for other reasons. It would be interesting to compare the results once a larger part of the data has been geo-coded with a finer granularity.

Table 5 shows the characteristics of the 6 registered fires: the name, the size in hectares of burnt area and the date. The fires occurring in the same province and the same date are treated as one for the purposes of determining the accuracy of the clustering, as for the second cell of the table. For each fire the correlation with a related cluster of each model is shown. A closer inspection of Table 5 reveals several aspects that deserve attention. First, there are three clusters that are identical in all four cases. Second, the clusters containing a majority of images from Flickr have a higher average number of VGI items per user. This means that usually people post more than one photo of an event. In the assessment of the cluster this aspect must be taken into account, for example by grouping assessment value per user. Additionally, there are clusters that are formed by both types of media (Tweets and photos), but many clusters have a distinct preponderance of one form of media. The correlation of clusters event with the forest fire information, such as burnt area, and demographic information, such as internet penetration, must be carefully analyzed to find interesting correlations. Figure 5 shows the VGI events in southern France colour-coded according to the clusters as detected in case 1. The z-Axis shows the temporal dimension of the data.

Apart of diverging from the remote sensing analogy, the analysis also highlights that noise reduction is a key issue when using VGI. Noise becomes naturally introduced through various degrees of freedom in tag assignment. Wild or forest fires events are particularly affected by this phenomenon, because there will always be a large number of VGI about other types of fires or burning things. The results show the necessity to assess the quality both at the level of single pieces of information and at cluster level. In order to not discard any information, the quality should be assessed by introducing attributes like credibility and relevance for each piece of VGI and each cluster. In this way, the threshold for alerts via the SES can be easily adjusted in field trials without having to change the entire processing work flow.

Creating Fire Alerts

These results show that every fire event can be detected and they confirm the social network activity around disaster reports. Moreover, the time analysis highlight that messages are correlated in real-time to the fire event. This can lead to a future development of a almost-real-time system for VGI monitoring, in which the spatio-temporal clustering is performed in short intervals. Again, in order to implement step 5, the fire alert can be propagated from the SES to relevant authorities. It is worth notice that the analysis of false positive clusters leads to the hypothesis of detecting 'non-official' fires. This highlights a potential improvement for alert systems that today is only linked to hot spots analysis of satellite imagery.

6 Discussion

The two examples provide first insides to VGI Sensing. Whereas the flood example illustrates the analogy to remote sensing and establishes some core principles, the wild fire case indicates the complexity of the topic and provides directions for further research. Practically speaking, we have to develop means to harvest VGI from multiple platforms. Beyond this, we also require theoretical work addressing meaningful analysis techniques and approaches for integrating different senses. We address these issues in the remainder of this section and outline the requirements for applying the presented approach to crisis events in general.

Technical Feasibility

To harvest VGI from multiple platforms, we require means to increase the number of VGI sources to be integrated. (Fonts et al, 2010) exposed a mechanism to use a simple query interface to integrate a bigger set of Web 2.0 Services and improve data accessibility. Based on this previous research we propose a web service, the *Web 2.0 Broker*. The functionality of this service is to access Web 2.0 Services functionality (search interface, geographic content data type) through a unique entry point implementing a common simple query interface: OpenSearch (Gonçalves 2010; Clinton, 2010).

OpenSearch (OS) defines a minimal interface to query a search engine that is extensible by adding extra parameters to define other filtering criteria. Such extensions include the *time-extension*, the *semantic-extension*, and the *geo-extension*. The latter allows for the use of multiple location filters: bounding box, circle, polygon and place name.

OS and its geo extension (Turner et al. 2010) is proposed as the query interface to access geographic content, both for Web 2.0 Services and also for SDI services. The Web 2.0 Broker is able to receive OS queries propagate them to a set of Web 2.0 Services and return the results encoded in standard data formats as GeoRSS, GeoJSON, KML, or ATOM.

As shown in Figure 6, the Web 2.0 Broker aggregates Web 2.0 adapters that translate an original OS query to the concrete syntax of each Web 2.0 Service API. This approach implies the development of OS adapters for each Web 2.0 Service instead of using the proprietary search tools of each Web 2.0 resources. It has the advantage that potential calibration mechanisms can be encapsulated in well defined components, which direct connect and use the specific Web 2.0 Service APIs. However, the search criteria based on the OS need to be mapped into the specific Web 2.0 Service API and this means that we could lose accuracy in certain parts of a query. This may have an impact on the numbers of VGI items retrieved.

For the first prototype, the Web 2.0 Broker encompasses the adapters for a selection of services:: Twitter, Flickr, YouTube, OpenStreetMap and Geonames (see Figure 6). To test this component within our second scenario to detect wild fires in France we need an OS client able to send an OS (Geo extension) query, in our study case this query would be:

$$q= \textit{incendie} \ \& \ l=\textit{France}$$

The Web 2.0 Broker receives this query and broadcasts it to the different Web 2.0 Services by delegating in the aggregated Web 2.0 adapters (Figure 6). Depending on the request, the collected results are translated from JSON to KML or ATOM in order to return a well formed client side response. The first results show that for this specific query, and using the Web 2.0 Broker we are able to offer a simple way to make the same query to Twitter, Flickr and YouTube at once. Consequently the amount of data collected in the previous study is increased with 589 YouTube results which could add extra information after their evaluation. The Web 2.0 Broker increases the number of available Web 2.0 Services by providing a unique and standard entry point to search for VGI data. Using the previous use case the Web 2.0 Broker increases the number of obtained results, having new data from different web services, such as YouTube, Wikipedia, OpenStreetMap or Geonames, providing more detailed information for step 1 of the VGI sensing workflow (Figure 1).

The Complexity behind VGI

The flood and wild fire examples illustrate the complexity of VGI Sensing. As long as considering only one VGI platform, event extraction can be performed similarly to feature recognition in remote sensing. We presented how this may be achieved for flood detection using Flickr. Nevertheless, although the presented principles are independent of a particular type of crises event (such as flood, fire, hurricane, political crisis etc.), calibrations might be required. Whereas a flooding occurs in a relatively short time frame and might affect people immediately, other events (for example oil spills) have different geospatial characteristics and different temporal impact.

Accordingly, the patterns in VGI differ depending on the considered types of phenomena and the associated geospatial, temporal and event criteria.

In addition, it remains to be examined how different phenomena impact the content of different VGI platforms and the attention of crises by (social) media and by VGI platforms needs to be analyzed in order to calibrate VGI Sensors correctly. As oil spills commonly occur off-shore, there will be fewer pictures available on Flickr, than comments and concerns posted on Twitter. Only the consequences of an oil spill (dirty beaches, dead birds etc.) will be reported in pictures. Earthquakes and outbreaks of diseases cannot directly be photographed at all, but again their impacts may be reported in pictures, whereas floods, fires and landslides are a more visual features that can be directly pictured. Twitter posts can be generated in any case.

Complementary to this, social networks also act as an information repeater, i.e. information about (crisis) events becomes repeated/propagated through a series of posts. The big 2010 fires in Russia provide just one example. They were discussed throughout the globe and only few posts directly addressed a particular wild fire location. This echo effect may be addressed using language processing and semantic technologies. Assuming that echoed information is less accurate, posts (such as Tweets) which are spatially or temporally 'far' from the mentioned event, or of 'weak' level of detail may be filtered out.

We can expect that, depending on the type of crisis event VGI can be valuable at various stages of the emergency management cycle. Flickr images for example may report on an ongoing wild fire, on the burned area shortly after the fire, as well as on the different phases of the re-creation process. We expect that if we want to trace a crisis event throughout its lifecycle, i.e. from onset to damage assessment, to reconstruction, we cannot rely on a single platform. Developing means to access multiple platforms in a harmonized manner is a necessary building block that has to be provided as an initial step, but it will be sufficient for implementing VGI Sensing.

Considering more than one VGI platform increases complexity, and adds another level of required calibrations. As the second example (the detection of forest fires) illustrated, VGI items may interrelate between platforms and contributors to different platforms behave differently. Apart for the above mentioned calibration for pattern recognition on a single-platform (first level calibration), some patterns may only be recognized when considering multiple VGI sources. The combination of platforms and the *observed-property* of the 'flow of VGI', i.e. this second level of calibration, will again depend on the type of event under consideration. Technically, it has to be evaluated if the required calibrations have to be realized at deploy time, i.e. when putting a VGI Sensor into place, or at run time. The latter option would result in a general VGI Sensor, which could be tasked for multiple types of events. Still, in depth investigations on the impact of crisis events on the content of multiple VGI platforms and the interplay requires in-depth investigations.

From the Peripheral to the Central Nervous System

Combining different sources of social media might increase the potential utility of VGI since it would provide more search results to be used in the study to improving the accuracy. Just as the brain acts as a system to integrate stimuli belonging from different sense organs and aggregate them in one single interpretation, we should make a step forwards and characterize the different social media stimuli. In this manner, Flickr can be associated to the eyesight while Twitter to the sense of hearing. A central system that is able to process the two channels is mandatory at some development point. We move from the peripheral nervous system of single VGI Sensors to the central nervous system, which integrates information from multiple heterogeneous sources and makes sense out of it.

Following this line of thought and considering the fact that casual events that trigger crises cannot be always be detected (and/or followed) by humans, VGI should complement other sources for Earth Observation. This again holds for oil spills, where numerous occasions are detected by remote sensing. Many systems using traditional sensor networks are already in place and they become more and more connected to the Internet. Advances in sensor web and model web (Geller and Melton 2008) technologies should be used in combination with VGI Sensing. First conceptual solutions on an integrated approach have been presented (Schade and Craglia 2010), but implementations still have to consider the complex relationships between all the different channels. Only if we find a way of integrated monitoring and event detection, we will be able to establish a central nervous system (a brain) for the Digital Earth. The development of such a system should be the goal for the coming years. This would particularly serve crisis management, but would also contribute to more general challenges of understanding the interplay between the society and their environment (ICSU 2010).

7 Conclusions and Future Work

The potential of up to 6 billion humans to monitor the state of the environment, to validate global models with local knowledge, and to provide information that only humans can capture is vast and has yet to be fully exploited. In this paper, we shifted the focus of attention from ‘citizen as sensors’ to VGI sensing, that is the sensing of VGI flows. SWE proved useful for clarifying this approach. We presented a mapping between central SWE concepts (such as *feature-of-interest* and *observed-property*) and VGI Sensing, as well as a possible application of SWE technologies (such as O&M and SOS) to VGI. A workflow for event detection based on VGI Sensing was refined and improved based on previous work. The refinement included the new role of (virtual) VGI sensors. An example walkthrough was provided for the case of flood detection in the UK based on Flickr images. We provided a second example, in which we included multiple VGI platforms (Twitter and Flickr) to detect wild fires in France. This illustrated the limits of the analogy between VGI and remote sensing and provided further insights to VGI Sensing as a research field in its own rights. For example, VGI Sensing has to include two levels of calibration (i) calibrating the sensor for processing content from specific platform, such as Flickr; and (ii) calibrating the sensor in order to integrate VGI coming from multiple Web 2.0 sources. All this advances a novel way for processing VGI and will help to establish the Digital Earth’s nervous system.

Our work indicates that VGI Sensing can be complementary to remote sensing, and ‘traditional’ in-situ and ex-situ sensors. It can provide high-scale value-added information at low cost. Furthermore this approach could be used as to enrich crisis management models inputs or to refine its output results. As a next step, we will investigate this relation, especially in respect to shared *features-of-interest*, *observed-properties* and measurement *procedures*. VGI Sensing relies on human reporting changes in their environment and it is the human input to Web 2.0 that is sensed. ‘Traditional’ sensing, on the contrary focuses on the environmental changes directly. So does environmental simulation/modeling. It remains to be explored how both sensing principles can benefit from each other (Schade and Craglia 2010).

As argued previously (De Longueville et al. 2010b), sensor technology can be used at various abstraction levels. Especially the potential of cascading SOS remains to be investigated. Depending on the purpose, even events may be provided as a sensor, implementing notions such as ‘I observed a flood’ or ‘I observed a tiger’. Categorizing events as features of interest and the definition of according observable properties are topic to ongoing discussions. One key issue is to balance between reusability and efficiency when deciding on the level of abstraction. When features of interest remain close to the raw data being sensed the reusability of the exposed measurements is higher than that of measurements indicating detected events; although this would be more efficient in certain scenarios. Developments of the Web 2.0 broker are ongoing in parallel; it will provide an important means to harness VGI and to use its full potential.

Acknowledgements

This work has been partially supported by the European FP7 Project nr. 226487 called EuroGEOSS and the exploratory research project ‘Engaging citizen in the next generation Digital Earth – Case Study on Forest Fire’ of Joint Research Centre, Institute for Environment and Sustainability (JRC-IES).

References

- Annoni, A., Craglia, M., de Roo, A., and San-Miguel, J. (2010). Earth Observations and Dynamic Mapping: Key assets for risk management. In M. Konecny, S. Zlatanova and T. Bandrova, eds., *Geographic Information and Cartography for Risk and Crisis Management*. Springer-Verlag Berlin Heidelberg, 3 - 22.
- Bahree, M. Citizen Voices (2008). *Forbes Magazine*.
- Bishr, M. and Mantelas, L. (2008). A trust and reputation model for filtering and classifying knowledge about urban growth. *GeoJournal* 72, 3-4, 229-237.
- Botts, M., Percivall, G., Reed, C., and Davidson, J. (2008). OGC Sensor Web Enablement: Overview and High Level Architecture. *Lecture Notes in Computer Science* 4540, December, 175–190.
- Botts, M. and Robin, A. (Eds.) (2007). *Sensor Model Language*. Open Geospatial Consortium Implementation Specification.

- Bröring, A., Stasch, C. and Echterhoff, J. (Eds.) (2010). Sensor Observation Service v2.0. Open Geospatial Consortium Implementation Specification.
- Campbell, A.T., Eisenman, S.B., Lane, N.D., Miluzzo, E., Peterson, R.A., Lu, H., Zheng, X., Musolesi, M., Fodor, K. and Ahn, G.S. (2008) IEEE Internet Computing, 12, 12-21.
- Clinton, D. (2010) OpenSearch 1.1 Draft 4 Specification.
- Cox, S.J.D. (Ed.) (2010). Observations and Measurements v2.0. Open Geospatial Consortium Abstract Specification, Topic 20 (also published as ISO DIS 19156:2010).
- Craglia, M. (2007). Volunteered Geographic Information and Spatial Data Infrastructures: when do parallel lines converge? Specialist Meeting on Volunteered Geographic Information. 13-14 December 2007. Santa Barbara
- De Longueville, B., Smith, R.S., and Luraschi, G (2009). OMG, from here, I can see the flames!: a use case of mining location based social networks to acquire spatio-temporal data on forest fires. Proceedings of the 2009 International Workshop on Location Based Social Networks, ACM (2009), 73-80.
- De Longueville, B., Luraschi, G., Smits, P., Peedell, S., and De Groeve, T. (2010a). Citizens as sensors for natural hazards: a VGI integration workflow. *Geomatica* 64, 1 (in press).
- De Longueville, B., Annoni, A., Schade, S., Ostländer, N., and Whitmore, C. (2010b). Digital Earth's Nervous System for crisis events: real-time Sensor Web Enablement of Volunteered Geographic Information. *International Journal of Digital Earth* 3, 3 (in press).
- De Rubeis, V., Sbarra, P., and Tosi, P. (2008). Web based macroseismic survey: fast information exchange and elaboration of seismic intensity effects in Italy. In J. Landgren and S. Jul, eds., *Proceeding of the 6th International ISCRAM Conference Gothenburg, Sweden*. ISBN - 978-91-633-4715-3. 2009.
- Díaz, L., Granell, C., Fonts, O. and Gil, J. (2010). D.2.6.1. Report on Web 2.0 and alternative search mechanism. Deliverable EuroGEOSS. November 2010.
- Bertsekas, D.P. and Tsitsiklis, J.N. (2002). *Introduction to Probability*. Athena Scientific, Massachusetts ISBN 1-886529-40-X.
- Echterhoff, J. and Everding, T. (Ed.) (2008). *OpenGIS® Sensor Event Service Interface Specification v0.3.0*. Open Geospatial Consortium Discussion Paper.
- Elwood, S. (2008). Volunteered geographic information: Key questions, concepts and methods to guide emerging research and practice. *GeoJournal* 72, 3-4, 133-135.
- Flanagin, A. J. and M.J. Metzger (2008). The credibility of volunteered geographic information. *GeoJournal*, 72, 137-148.
- Fonts, O., Huerta, J., Díaz, L. and Granell, C. (2010). OpenSearch-geo: The simple standard for geographic web search engines. *Actas de las IV Jornadas de SIG Libre 2010 (SIG Libre 2010)*. Girona, Spain, Mar 2010.
- Geller, G. and Melton, F. (2008). Looking forward: Applying an ecological model web to assess impacts of climate change, *Biodiversity* 9, no. 3&4.
- Goodchild, M.F. (2007). Citizens as voluntary sensors: spatial data infrastructure in the world of Web 2.0. *International Journal of Spatial Data Infrastructures Research* 2 24-32.
- Gould, M. (2007). Vertically interoperable geo-infrastructure and scalability. Specialist Meeting on Volunteered Geographic Information. 13-14 December 2007. Santa Barbara.
- Hughes, A.L. and Palen, L. (2009). Twitter Adoption and Use in Mass Convergence and Emergency Events. *Proceeding of the 6th International ISCRAM Conference*.
- ICSU (2010). *Grand Challenges in Global Sustainability Research: A Systems Approach to Research Priorities for the Decade*. International Council for Science, pre-publication version, June 2010.
- Jankowski, P. (2009). Towards participatory geographic information systems for community-based environmental decision making. *Journal of Environmental Management* 90. 1966-1971.
- Jenks, G.F. and Coulson, M. (1963). Class intervals for statistical maps. *International Yearbook of Cartography* 4 3, 119-134.
- Jiang, M. and McGill, W.L. (2010). Human-Centered Sensing for Crisis Response and Management Analysis Campaigns. *Proceedings of the 7th International ISCRAM Conference - Seattle, USA, May 2010*.
- Jones, M.T. (2009). Unexpected Change - Transforming the GeoWeb for a New World. Keynote speech.
- Jürrens, E.H., A. Broering and S. Jirka (2009). A Human Sensor Web for Water Availability Monitoring. *Proceedings of: OneSpace 2009 - 2nd International Workshop on Blending Physical and Digital Spaces on the Internet*. September 1. 2009. Berlin, Germany.
- Kulldorff, M., Heffernan, R., Hartman, J., Assunção, R., and Mostashari, F. (2005). A Space-Time Permutation Scan Statistic for Disease Outbreak Detection. *PLoS Med*, 2(3).
- Lee, V. (1994). Volunteer monitoring: a brief history. *The Volunteer Monitor* 6, 1.

Mummidi, L. and Krumm, J. (2008). Discovering points of interest from users' map annotations. *GeoJournal* 72, 3-4, 215-227.

Ostermann, F., Spinsanti, L., De Longueville, B. and Craglia, M. (2010). User Generated Geographic Information and Natural Hazards - How Volunteered Geographic Information Can Contribute to the Assessment and Communication of Forest Fire Risks. *GIScience 2010 - Sixth international conference on Geographic Information Science Zurich*, 14-17th September, 2010.

Palen, L., Vieweg, S., Liu, S., and Hughes, A. (2009). Crisis in a networked world: Features of computer-mediated communication in the April 16, 2007, Virginia Tech event. *Social Science Computer Review* 27, 4, 467-480.

Sakaki, T., Okazaki, M. and Matsuo, Y. (2010). Earthquake Shakes Twitter Users: Real-time Event Detection by Social Sensors. *WWW2010, April 26-30, 2010, Raleigh, North Carolina*.

Schade, S. and Craglia, M. (2010). A Future Sensor Web for the Environment in Europe. *EnviroInfo2010, Cologne/Bonn, Germany*.

Shah, A.R., Adkins, J.N., Baxter, D.J., et al. (2010). Applications in Data-Intensive Computing. In *Advances in Computers*, 79, 1-70, Elsevier.

Siegele, L. (2010). A Sea of Sensors. *The Economist – A special report on smart systems*, 6th of November 2010.

Simonis, I. (Ed.) (2006). *OGC Sensor Alert Service Candidate Implementation Specification v0.9*. Open Geospatial Consortium Best Practise.

Simonis, I. and Echterhoff, J. (Eds.) (2010) *Sensor Observation Service v2.0*. Open Geospatial Consortium Implementation Specification.

Spinsanti, L. and Ostermann, F. (2010). Validation and Relevance Assessment of Volunteered Geographic Information in the case of Forest Fires, *ValGeo Workshop, Ispra, Italy*.

Turner, A (2010) *The OpenSearch Geo extension (draft 2)*. http://www.opensearch.org/Specifications/OpenSearch/Extensions/Geo/1.0/Draft_2.

	Event detection based on VGI Sensing	Event detection based on remote sensing
Stimulus	A new item of VGI is entered into a Web 2.0 platform.	Waves are reflected or emitted by a surface.
Sensor	VGI items entered by citizen are detected and 'discretized' according to a VGI sensor specific array/grid.	Waves are detected and digitized thanks to a satellite-mounted sensor, i.e. camera.
Sensation	Heterogeneous information is centralized and organized in a grid of measurement results, according to the VGI virtual sensor's specifications.	Series of remote sensing images are created according to the image sensor's specifications.
Perception	Patterns are found in results, and events and situations are identified thanks to prior knowledge.	Image series are processed to detect signals with specific characteristics leading to the identification of events of a specific type(s).
Attention	Alerting mechanisms are triggered according to context.	Alerts are triggered according to context.
Reaction	Sensor network information is integrated in information systems, where appropriate tasks are prioritized, related to: <ul style="list-style-type: none"> o early response to crises; o situation awareness, request for additional information; o monitoring of parameters; etc. 	

Table 1: Event detection based on VGI Sensing compared to remote sensing (adapted from De Longueville et al. 2010b).

VGI at Web 2.0 Platforms

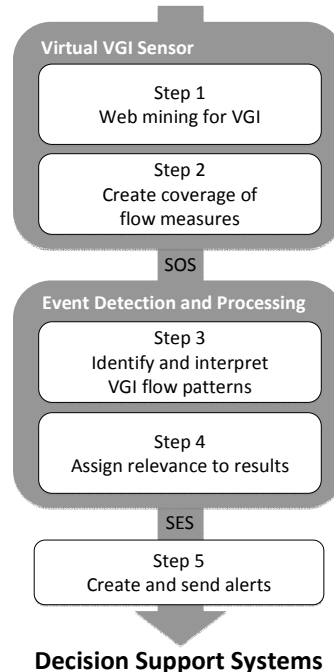


Figure 1: Overview of the VGI Sensing based event detection workflow.

SWE (O&M) element	VGI Sensing	remote sensing
<i>feature-of-interest</i>	Europe (earth surface and its periphery)	Europe (earth surface)
<i>observed-property</i>	Density of VGI tagged with 'flood'	Power in a certain wavelength
<i>Procedure</i>	Calculating densities for each cell of the grid	Assigning digital values to each pixel of the image sensor
<i>Result</i>	Cells are filled with data such as: flood; 2010-06-16; (-33, 135)	Satellite image product

Table 2: Mapping the sensing elements to SWE.

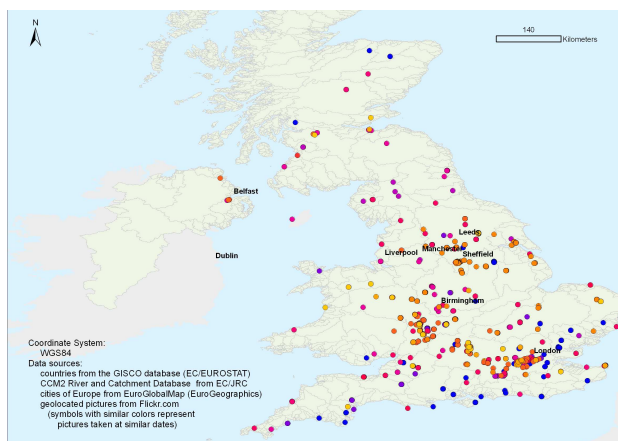


Figure 2: Spatio-temporal distribution of retrieved Flickr flood images (similar colors means similar acquisition time).

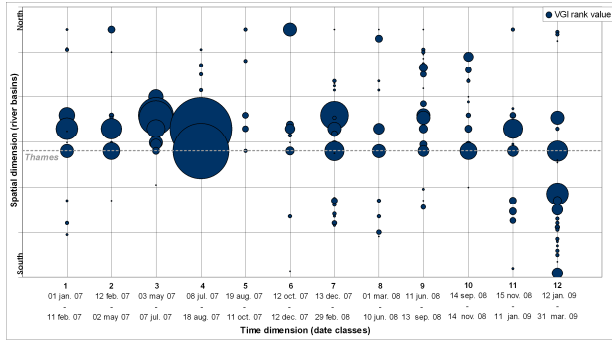


Figure 3; VGI virtual sensor results: spatio-temporal distribution of VGI aggregates.

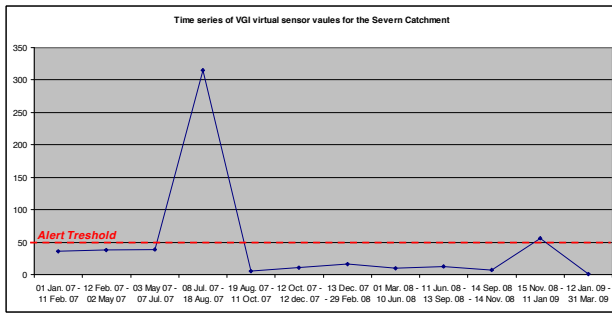


Figure 4: Example VGI Sensing time series with alert threshold.

	Entries in Summer 2010	At least one keyword in FR	Keyword 'incendie'	Coordinates	Indirect Location	Located in FR
Twitter	> 8 millions	611,274	6,754	31	1,123	437
Flickr	> 700 thousand	61,697	458	133	293	243
Total	> 8.7 million	672,971	7,212	164	1,416	680

Table 3. Dataset for French Fires Case Study.

Case	Total number of Clusters	True Positives	False Positives	False Negatives
1	8	6	2	0
2	74	7	67	0
3	4	3	1	3
4	6	3	3	3

Table 4: Accuracy of clustering methods.

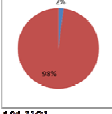
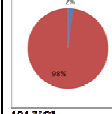
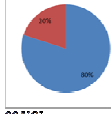
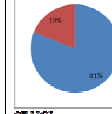
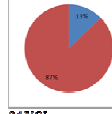
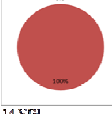
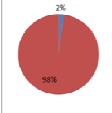
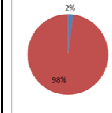
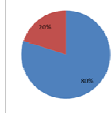
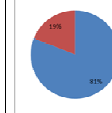
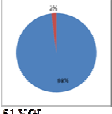
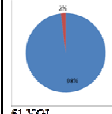
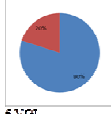
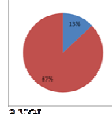
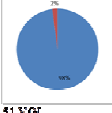
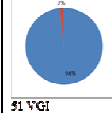
Fire: Sausset-les-Pins (Bouches-du-Rhône)		Fire: Saint-Bauzille-de-Montmel and Meze (Hérault)		Fire: Cassis (Bouches-du-Rhône)	
07-24-2010 840 hectares		06-30-2010 2747 + 452 hec (two close fires)		08-30-2010 152 hectares	
■ Flickr ■ Twitter		■ Flickr ■ Twitter		■ Flickr ■ Twitter	
CASE 1  181 VGI 1.4 avg VGI user	CASE 2  181 VGI 1.4 avg VGI user	CASE 1  99 VGI 3.3 avg VGI user	CASE 2  97 VGI 3.3 avg VGI user	CASE 1  24 VGI 1.1 avg VGI user	CASE 2  14 VGI 1.1 avg VGI user
CASE 3  181 VGI 1.4 avg VGI user	CASE 4  181 VGI 1.4 avg VGI user	CASE 3  99 VGI 3.3 avg VGI user	CASE 4  97 VGI 3.3 avg VGI user	CASE 3 NOT DETECTED	CASE 4 NOT DETECTED
Fire: Arzonnan (Aude)		Fire: Saugninet (Landes)		Fire: Boisse-Penchot (Aveyron)	
07-30-2010 66 hectares		09-03-2010 127 hectares		08-10-2010 65 hectares	
■ Flickr ■ Twitter		■ Flickr ■ Twitter		■ Flickr ■ Twitter	
CASE 1  51 VGI 8.5 avg VGI user	CASE 2  51 VGI 8.5 avg VGI user	CASE 1  5 VGI 2.5 avg VGI user	CASE 2 NOT DETECTED	CASE 1  9 VGI 1.5 avg VGI user	CASE 2 NOT DETECTED
CASE 3  51 VGI 8.5 avg VGI user	CASE 4  51 VGI 8.5 avg VGI user	CASE 3 NOT DETECTED	CASE 4 NOT DETECTED	CASE 3 NOT DETECTED	CASE 4 NOT DETECTED

Table 5: Accuracy of clustering methods.

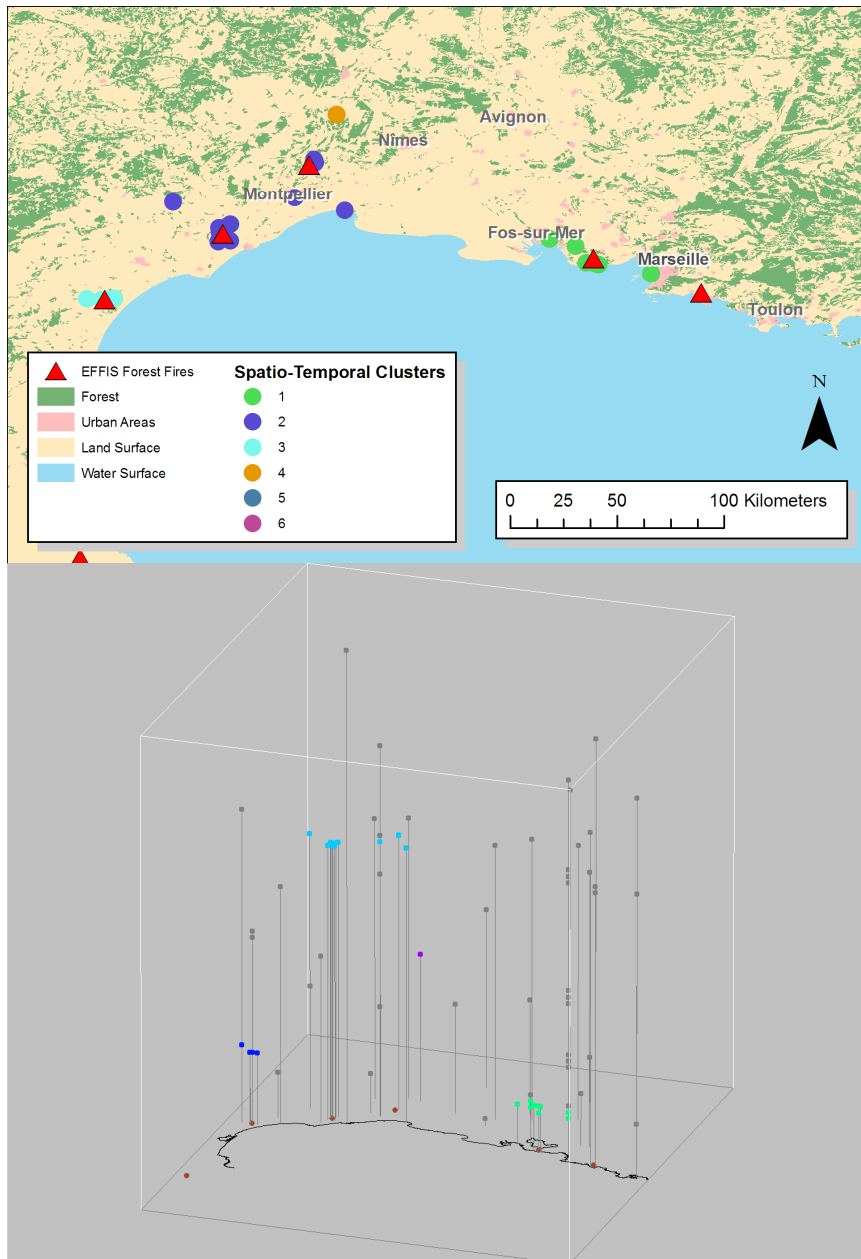


Figure 5: Space-time cube of forest fire VGI clusters.

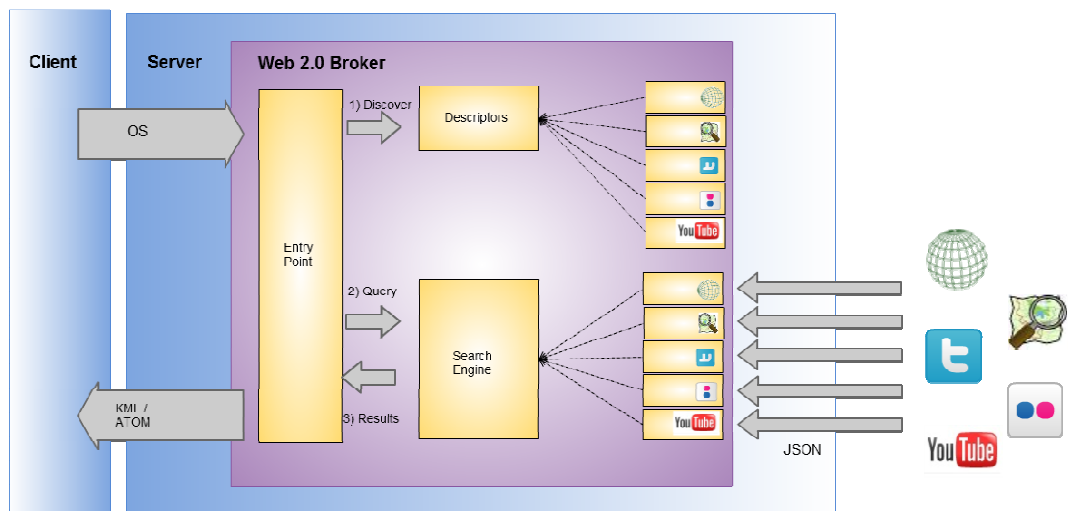


Figure 6: A Web 2.0 Broker for VGI (based on the open search API).