Journal of Spatial Science

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/tjss20

Improving the reproducibility of geospatial scientific workflows: the use of geosocial media in facilitating disaster response

V. Cerutti, C. Bellman, A. Both, M. Duckham, B. Jenny, R. L. G. Lemmens & F. O. Ostermann

To cite this article: V. Cerutti, C. Bellman, A. Both, M. Duckham, B. Jenny, R. L. G. Lemmens & F. O. Ostermann (2021) Improving the reproducibility of geospatial scientific workflows: the use of geosocial media in facilitating disaster response, Journal of Spatial Science, 66:3, 383-400, DOI: 10.1080/14498596.2019.1654944

To link to this article: https://doi.org/10.1080/14498596.2019.1654944

Published online: 23 Aug 2019.

JOURNAL OF SPATIAL SCIENCE
2021, VOL. 66, NO. 3, 383–400
https://doi.org/10.1080/14498596.2019.1654944

Improving the reproducibility of geospatial scientific workflows: the use of geosocial media in facilitating disaster response

V. Cerutti [a,b], C. Bellman [b], A. Both [b], M. Duckham [b], B. Jenny [c], R. L. G. Lemmens [a] and F. O. Ostermann [a]

[a]Faculty of Geo-Information Science and Earth Observation (ITC), University of Twente, Enschede, The Netherlands; [b]School of Science, RMIT University, Melbourne, Australia; [c]Faculty of Information Technology, Monash University, Melbourne, Australia

ABSTRACT
Reproducibility is widely regarded as crucial for scientific studies, yet there is still a lack of reproducibility in geospatial research. New sources of crowdsourced geoinformation provide new opportunities, but also complicate the reproducibility situation. Consequently, there is untapped potential in the domain of disaster response to reuse scientific methodology. Shared, executable scientific workflows can help in improving reproducibility. In this paper, we created reproducible scientific workflows for disaster response from three published studies using geosocial media sources. They have been adapted to a scientific workflow management system to investigate and evaluate their suitability for the creation of geospatial footprints of wildfire events from Twitter data. We investigated how scientific workflows adapt to various analytical processes and compared their performance using MODIS active fires data as ground truth. A systematic qualitative and quantitative evaluation demonstrated that scientific workflows can help increase the reproducibility of geospatial analytics.

KEYWORDS
Scientific workflow; geospatial footprint; geo-social media; disaster management; reproducibility

Introduction

The past two decades have seen a massive increase in available geoinformation from remote sensing and crowdsourced geosocial media (CGSM), as well as available geocomputation methods (modelling, data mining, artificial intelligence). Because the means of sharing data, methods, and knowledge have not kept up, there is substantial duplication of efforts between research groups, institutions and practitioners. At the same time, we still have insufficient knowledge about the transferability of specific results to other geographic regions. In disaster response scenarios, a fast but reliable response is required, with disaster response experts having to adapt protocols and processes to specific situations. Facilitating reproducibility of methods and workflows would aid disaster response practices, but reproducibility has been identified as a challenge in several domains (Editorials 2016), including geospatial research (Nust *et al*. 2018).

CONTACT V. Cerutti v.cerutti@utwente.nl

One possible approach to increase reproducibility are scientific workflow management systems (SWMS) that help to conceptualise and manage the analysis process, allowing for the creation and reuse of analytical tasks while providing a self-documenting, executable visual language for analytics (Scheider *et al*. 2017). SWMS usually have a graphical user interface, interactively representing an analytic process as a directed graph with nodes (discrete stages or tasks in the analysis), and edges (connections between these nodes representing the flow of data). This process graph supports analysts when composing, executing, assessing, and modifying a workflow. However, SWMS are not commonly used for geospatial analysis of established data sources, and even less so for novel data sources such as CGSM.

This paper aims to contribute to our understanding on how to create reproducible workflows for disaster response, using CGSM. Three scientific workflows from the literature are adapted to SWMS to investigate and evaluate their performance for creating geospatial footprints of wildfires. Determining affected or vulnerable areas of wildfires from CGSM is a crucial and challenging step in successful disaster response for which geosocial media as of now has still untapped potential (Granell and Ostermann 2016). We evaluate how scientific workflows perform in modelling these processes, identify potentials and limitations and demonstrate how a scientific workflow can be a self-documenting and interactive tool to support reproducibility of geospatial analysis.

The key contributions of this work are: (1) a demonstration of the implementation of three recent disaster response analytics processes in a SWMS; (2) a qualitative and quantitative evaluation of the outcomes of the three workflows; (3) an exploration of the key common operations; (4) integration of the outcomes into the knowledge on improving reproducibility of geosocial media analytics for disaster response.

Key concepts and related work

Geosocial media

Technologies such as smartphones with receivers for global navigation satellite systems, cloud computing, and Web 2.0 allow users with no professional background to produce geographic data, transforming the way it is collected, used, and disseminated. This has facilitated new possibilities to discover geographic knowledge and to analyse human behaviour from social media data (Miller and Goodchild 2015, Capineri *et al*. 2016). In addition to offering previously inaccessible local and tacit knowledge, social media are accessible in real time and at a fraction of the cost with respect to traditional methods. However, many challenges remain, in particular relating to large volumes of noise in the data (Ostermann and Spinsanti 2012), bias in contributions associated with demographics, internet use and technology access (Haworth *et al*. 2012), local variation in usage (Zahra *et al*. 2017), integration with authoritative data sources (Schade *et al*. 2012), trust and credibility (Flanagin and Metzger 2008), legal issues around privacy and liability (Scassa 2013), and reproducibility due to volatility of content and terms of service (Ostermann and Granell 2017). Despite these issues, studies have shown the potential of CGSM to detect and track information about events, with application in several fields such as disaster management, urban planning or health (Alexander 2013, Grajales *et al*. 2014, Shelton *et al*. 2015).

Geospatial footprints

Geospatial events can be represented by their geospatial footprint, i.e. the area that is or has been directly affected. Typically, there are three common steps in generating geospatial footprints from diverse data sources: filtering, clustering, and shape reconstruction.

Filtering techniques are applied to clean the data, reduce noise, and retain only event-related information. Filtering can be geospatial, restricted to a certain region or area; temporal, restricted to a certain timeframe; or based on semantics, utilising the textual content. Geospatial filtering of CGSM is challenging because of the heterogeneity and variety of locational information: a single post can contain coordinates provided by the device's GNNS, a place name suggested by the social media platform, a home location in the user profile provided by the contributor, and any place mentioned in the actual content. While the first two usually indicate where a post is from, the user home location can indicate local knowledge, and the content can indicate which place(s) the post is about. Furthermore, any place names must be geocoded before being useful for further analysis, with the frequent problems of place name recognition, disambiguation, and grounding (areas represented as points, etc.). In contrast, temporal filtering is more straightforward, since most CGSM have a well-defined temporal signature, allowing temporal filtering to detect events in CGSM streams by looking for a peak occurrences of a phenomenon within a short time window. Lastly, semantic filtering focuses on the content of the messages, aiming to remove unrelated messages and enrich related. Semantic filtering often relies on natural language processing (NLP) techniques to extract keywords or content from unstructured text, and match or encode this extracted content with knowledge representation methods such as ontologies. Established NLP tools need adaptation to deal with the short text of CGSM annotations/tags (Cheng *et al*. 2010). Most studies use keyword-based queries (Kongthon *et al*. 2012, Cerutti *et al*. 2016) or keyword co-occurrence (Spinsanti and Ostermann 2013).

Clustering consists of grouping a set of objects so that the within-group similarity is higher than the between-group similarity, allowing the identification of events or features of an event. Many clustering algorithms exist. We limit this introduction to the two common algorithms that have been used in the three published studies that have been adapted in this paper: k-means and DBSCAN. K-means algorithm (Macqueen 1967) is a distance-based method that produces elliptical clusters around a set of centroids and is very sensitive to noise. It requires the number of clusters (k) as input, determines k centroids in the data and clusters points by assigning them to the nearest centroid. DBSCAN (Ester *et al*. 1996) is a density-based method that uses density thresholds around each object to distinguish relevant data from noise. It infers the number of clusters based on the data and can discover clusters of arbitrary shape. DBSCAN requires the estimation of two parameters: how close points should be to each other to be considered a part of a cluster (eps), and how many neighbours a point should have to be included into a cluster (MinPoints). Using these parameters, it classifies points into core points, border points (which define the clusters) and noise. A variation of DBSCAN is ST-DBSCAN (Birant and Kut 2007), which explores the spatiotemporal behaviour of events using a spatiotemporal distance function with space and time thresholds.

Shape reconstruction algorithms are not commonly used in CGSM studies. Galton and Duckham (2006) compare various methods for producing footprints, such as gift wrapping, swinging arm, close pairs, Delaunay triangulation (Arampatzis *et al*. 2006), Voronoi-based region approximation (Alani *et al*. 2010), alpha-shape (Edelsbrunner *et al*. 1983), and power crust (Amenta *et al*. 2001). Convex hull algorithms are the simplest and most frequently used, computing the smallest convex polygon that contains all the points in the given geometry, without having any angle that exceeds 180 degrees between two neighbouring edges. However, the convex hull does not represent well the boundaries of a given set of points with a pronounced non-convex distribution. Concave hull algorithms can be more efficient in these cases. A concave hull of a set of points can be defined as the shape which minimises the area of the containing shape, allowing any angle between the edges. The characteristic-shape algorithm used by Zhong *et al*. (2016) is an example of a concave hull algorithm.

Reproducible scientific workflows

One important benefit of executable scientific workflows is their potential to increase reproducibility. Reproducibility of a scientific study means that independent researchers can derive the same results of the study by using the same data and methods. A close concept to reproducibility is replicability, which means deriving similar results and conclusions without necessarily using the same input data and methods (Ostermann and Granell 2017). Ideally, scientific workflows are self-contained, executable, and self-documenting and can be easily shared, facilitating reproducibility of scientific research (Scheider *et al*. 2017).

Scientific workflows have different levels of granularity. A node (task) is atomic if it cannot be decomposed into smaller nodes (i.e. it provides a single task/processing step), it does not contain sub-task(s), and it has input ports and outputs ports that connect it to other (atomic or composite) nodes. Atomic nodes can be grouped into composite or meta-nodes.

Scientific workflows are data-oriented, using mostly sequential or iterative dataflows to communicate between nodes (next task uses the output of previous tasks as input, previous tasks need to be completed to start the execution of the next task) (McPhillips *et al*. 2009).

Multiple granularity levels and consistency in the dataflows allow modularity: complex processes can be decomposed in modular, reusable parts composed of simpler operations. Interchangeable modules can be connected and subsequently executed. This modularity makes scientific workflows adaptable to different users' skills and needs, and if combined with a user-friendly graphical user interface, allows non-expert users to use and combine existing modules and expert users to modify or adapt modules or nodes according to their specific requirements. Every change is also documented and saved in the workflow.

Methods

Selecting studies to reproduce

The studies to be reproduced should have four characteristics:

(1) Datasets: Use of CGSM either as a standalone data source or in combination with authoritative data (e.g. population data, meteorological data, etc.), for testing the adaptability of scientific workflows to various datasets.
(2) Methods: Use of different methods to filter, cluster and summarise their point-based data, for testing the modularity of the scientific workflows and the reuse of tasks.
(3) Case study: Focus on disasters from natural hazards, for testing the adaptation of scientific workflows to different applications.
(4) Year: Recent works, from 2012 onward.

Since our objective was not a comprehensive or representative meta-study but a first investigation and exploration, we combined convenience sampling from our professional network and a systematic literature search to choose three studies. The systematic literature search used the search terms 'social media', 'spatial analysis', and 'disaster response' on Scopus database. We used a two-stage reviewing process for the returned results: first scanning of abstracts, then in-depth evaluation of promising candidates. The following paragraphs, Table 1 and Figure 1 all provide an overview of the three chosen studies that fulfil all criteria.

Saravanou *et al*. (2015) use visual analytics on Twitter data to identify the areas affected by a flood event. They collected Tweets through the Twitter Streaming API and used lexicon keywords to extract flood-related Tweets. K-means clustering then aggregated the point data into areas, experimenting and comparing the results using different *k* values. Using a prioritisation approach, areas are sorted according to the signal-to-noise ratio derived from comparing flood-related Tweets with the total number of Tweets. To evaluate the results, the authors used ground truth information.

Zhong *et al*. (2016) describe a technique for real-time tracking of wildfire perimeters based on curated crowdsourced data in the form of emergency calls, and integrate it with authoritative data (population density and dynamic wind fields) to increase estimation accuracy. Using ST-DBSCAN and shape-reconstruction (characteristic shape) techniques, their analysis filtered and clustered emergency calls based on topic, spatial location, and timing, constructing an evolving footprint of the wildfire perimeter.

Spinsanti and Ostermann (2013) designed a prototype to process information on wildfires from Twitter and Flickr. The main prototype modules are: (1) the sensor that retrieves CGSM from Twitter Streaming API and Flickr Search API, (2) the analyser that classifies the

Table 1. Overview of selected studies' main characteristics.

Paper Reference	Data sources	Analytical methods	Case study
(Saravanou *et al*. 2015)	Twitter, ground truth data	Keyword filtering, k-means algorithm, visual analytics	Flood
(Zhong *et al*. 2016)	Emergency calls, population density, wind speed and direction	ST-DBSCAN + χ- shape algorithm	Wildfires
(Spinsanti and Ostermann 2013)	Twitter, Flickr, LAU2 municipality data, CORINE land cover, MODIS hotspot data	Filtering, geocoding, enrichment, spatiotemporal clustering, quality assessment	Forest fires

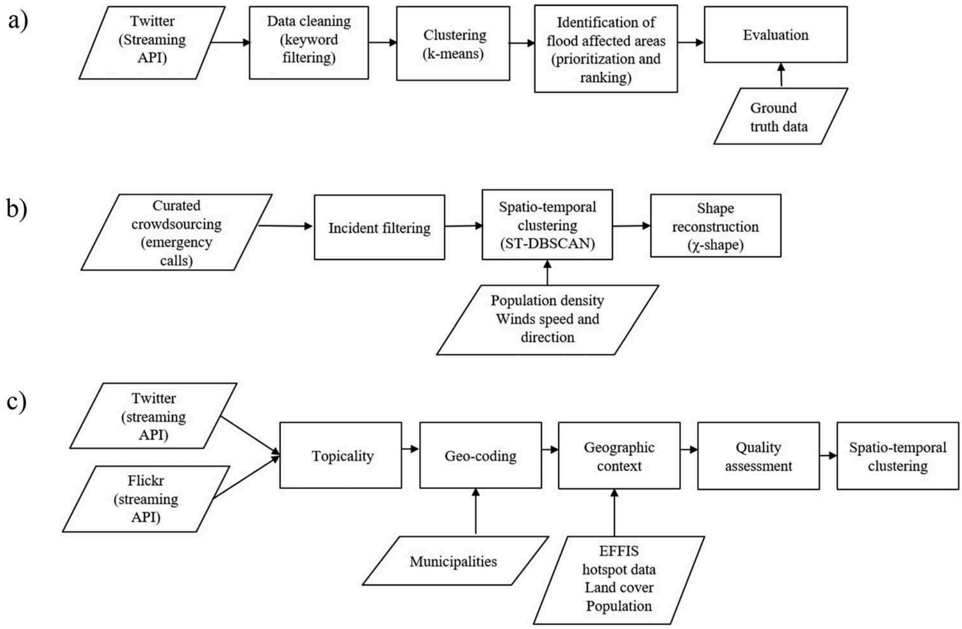

Figure 1. Overview of selected studies' main steps: (a) Saravanou *et al.* (2015), (b) Zhong *et al.* (2016) and (c) Spinsanti and Ostermann (2013).

topicality of each item using a decision tree, then geocodes relevant items based on extracted place names, and scores their credibility based on contextual information such as distance to nearest reported MODIS (Moderate Resolution Imaging Spectroradiometer) hotspots; (3) the clusterer that searches for spatio-temporal patterns (events) in the data; and (4) an interactive web map to visualise and communicate the results.

Creating scientific workflows in KNIME analytics platform

This paper uses the KNIME analytics platform (Berthold *et al.* 2009) to create scientific workflows that replicate the selected published studies. KNIME is an open source platform based on Eclipse, designed for data analysis, predictive analytics and modelling. It has a graphical user interface, making it intuitive and user-friendly. Additionally, it provides a broad range of functions including geospatial ones, which is uncommon in other SWMS (Scheider *et al.* 2017), and a large community of users that continue to develop extensions. For example, with the recently developed Open Spatial Analytics (OSA) plugin (Bellman *et al.* 2018), KNIME supports a typical CGSM processing pipeline, i.e. it allows connection to the Twitter Streaming/Search API, supports text processing, location extraction/geocoding using a gazetteer and has DBSCAN, k-means, and self-organising map algorithms already implemented in it. In addition, non-standard components can be added using R or Python scripting.

From here on, we refer to the scientific workflows created from the three studies of Saravanou *et al.* (2015), Zhong *et al.* (2016) and Spinsanti and Ostermann (2013) as SW1, SW2, and SW3, respectively. We decided to use a single new data set for all three

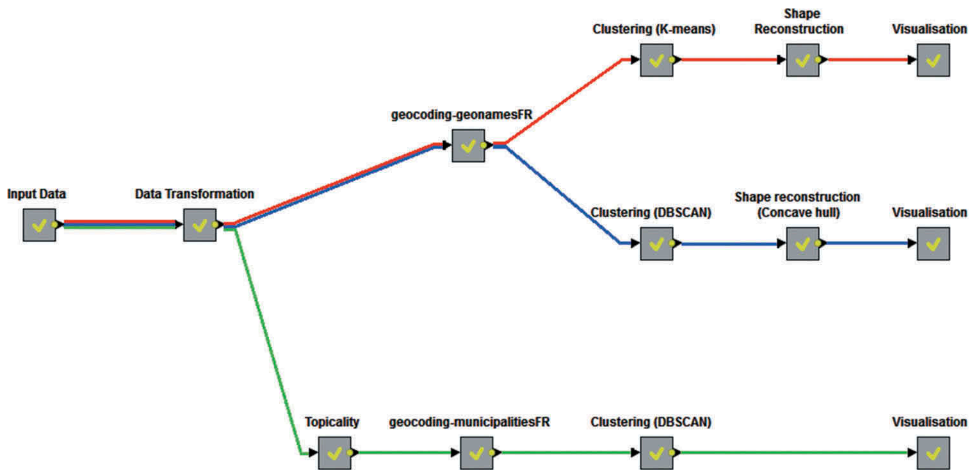

Figure 2. Combined view of SW1 (red – topmost), SW2 (blue – middle), SW3 (green – bottom) in the KNIME interface. Reuse of tasks allows a reduction in execution time.

workflows for two reasons: first, most of the original data was inaccessible because of limitations in terms of service and API capabilities. Second, using different data set for the three workflows would complicate a comparative evaluation.

Although most of the processing steps in the three studies can be replicated in KNIME using existing nodes, some steps required nodes to execute custom R or Python code. To increase consistency and modularity, we reused nodes within the three workflows whenever possible, combining individual nodes into five KNIME metanodes (see also Figure 2):

(1) Input data metanode: Some nodes require a specific data format, which led to the implementation of additional data transformation operations, increasing the number of steps and nodes. The interchangeable modules for the different types of input data support stream of Tweets, spreadsheets, JSON files, and multiple files readers.
(2) Data transformation metanode: input data are filtered, transformed, and geocoded if necessary. While KNIME has a Google Address Geocoder node, the rate limit of Google Maps API of 2500 free requests per day is too low for our data set, leading to the implementation of different geocoding nodes (see next section).
(3) Clustering metanode: we use k-means and DBSCAN clustering algorithms which are readily available in KNIME.
(4) Shape reconstruction metanode: converts cluster points into areas, using convex hull or concave hull nodes.
(5) Visualisation metanode: visualisation of the spatial data and the results can be performed within the platform or using any external GIS and exported as files.

Details and specifications of the metanodes for each SW are explained below (see the section on 'Reproducing workflows in KNIME').

Case study results and evaluation

This section describes the case study that applies and evaluates each of the workflows to data from a wildfire event. Twitter data were collected during the summer of 2017, using a similar query to the public Streaming API as Spinsanti and Ostermann (2013), as it has shown to produce usable data. For our case study, we limited the input data to Tweets generated on 25 July 2017, when major wildfires spread around France. The data consist of 88,281 Tweets. The dataset of tweet IDs as well as the SWs are available as additional material (DOI: 10.5281/zenodo.3271291) to this paper. Using the same input data for all SWs allows us to compare systematically the SW outputs qualitatively and quantitatively.

Reproducing workflows in KNIME

For the qualitative evaluation, we focused on the structure of the SW: Which steps are reproduced without change, and which steps required an approximation or slightly different replication? Since all SWs share the same input data, data transformation, and visualisation metanodes, their evaluation is next, followed by the SW-specific metanodes.

Input data: we adapted the input data metanode to read Tweets saved as JSON objects in text files from a folder in a local repository. This change was required because, as mentioned, the original data used in the selected works were not available.

Data transformation: this step was necessary to properly read the JSON data as a table into KNIME and to remove unnecessary columns and symbols in the text of the Tweets. This metanode approximates the data cleaning/filtering step adopted in the published studies, which were not sufficiently described to reproduce exactly.

Visualisation: visualisation of results was not considered as a step in the original workflows (expect for SW3 that used a Web Mapping Service to visualise and disseminate results). The MapViewer node allows to visualise the results within KNIME, but the results were also exported as shapefiles.

SW1 Geocoding: The published study used a geographic bounding box to obtain Tweets from the Streaming API, and then applied a keyword filtering on the geolocated Tweets. This case study inverts the process by first retrieving Tweets based on keywords, then using a lexical matching of content unigrams with the French subset of the GeoNames gazetteer (http://www.geonames.org/). This allows to increase the amount of usable data (usually only a small fraction (1–2%) of Tweets are geocoded) and to identify what a Tweet is about, as opposed to where is it from.

SW1 Clustering: the published study is replicated using k-means algorithm with the same k-values.

SW1 Footprint reconstruction: this metanode replicates the steps of prioritisation and ranking by counting the number of Tweets in each cluster and ranking the clusters in a descending order. The clusters containing the highest numbers of Tweets have been used to identify the footprint.

SW2 Geocoding: similar to SW1. Also, for SW2 we had to approximate the geocoding technique since the published study used emergency phone calls data that have been filtered and structured into spatio-temporal data by an external office.

SW2 Clustering: the published study used ST-DBSCAN, but for our case study the DBSCAN node is sufficient since a fine temporal resolution of 1 day is used and the event is already known and detected.

SW2 Footprint reconstruction: the published study is replicated by using the concave hull algorithm from the OSA extension of KNIME.

SW3 Topicality: this metanode replicates the algorithm described in the published study using a Python script to execute a decision tree created through supervised learning (Spinsanti and Ostermann 2013). This is the only metanode in the SWs using a scripting node to perform an operation, due to the unavailability of existing KNIME nodes to execute that specific task.

SW3 Geocoding: this metanode replicates the geocoding of the published study by looking at the Tweets' text and compares unigrams with simple lexical matching to place names of the LAU2 classification for French municipalities.

SW3 Clustering: the original spatio-temporal clustering of SatScan is replaced with the DBSCAN algorithm, making the same assumption on the data as for SW2.

To summarise, SW1 replicates the original workflow the most accurately, SW2 is the workflow with more approximations compared to the published study, and SW3 falls in between. For a summary of replications and approximations used see Table 2.

For a quantitative evaluation of the SW outputs, we looked into details of the number of records and nodes as summarised in Table 3. After filtering, transformation and geocoding, in SW1 and SW2 about half of the initial Tweets remained, while SW3 reduced the data set to fewer than 7000 Tweets, if filtering for only the top two of four topicality classes. This reduces notably the execution time for the geocoding and clustering metanodes in SW3. The metanodes that require more time to execute are the geocoding in SW1 and SW2 and the clustering using DBSCAN algorithm in SW2 (SW2 uses different parameters than SW3).

Table 2. Summary of metanodes used in the SWs.

Task/SW	SW1	SW2	SW3
Data Input	A	A	A
Data Transformation	A	A	A
Geocoding	A	A	R
Clustering	R	A	A
Footprint reconstruction	R	R	A
Topicality	–	–	R
Visualization	A	A	A

Notes: R = Replicated, A = Approximated, – = Not applicable.

Table 3. Characteristics of SWs in numbers. *topicality + geocoding.

Number of ...	SW1	SW2	SW3
Records in input	88,281	88,281	88,281
Records after geocoding	43,280	43,280	6847*
Clusters	100,500,1000,5000	2	6
Atomic nodes	39	33	30
Metanodes	7	8	7
Common metanodes	4	5	5
Scripting nodes	0	0	1

Geospatial footprints results

For the qualitative evaluation of geospatial footprints resulting from the SWs, we examine how similar or different the footprints are and how they compare to the ground truth.

We tested SW1 using the same values as in the published study (k = {10, 100, 500, 1000}) but added k = 5000 to test with larger number of clusters. Point clusters are then transformed into areas using the convex hull algorithm and ranked according to the number of Tweets in each cluster. For our dataset and case study, k = 1000 produces the most meaningful results. Lower k values return very large areas with many Tweets, while k = 5000 results in small areas and likely errors of omission, with too few Tweets to derive meaningful or actionable information on the event. Figure 3 shows the resulting footprints of SW1. Although several parts of France seem to have been affected by wildfires, the south clearly is affected the most.

SW2 has the same input, data transformation and geocoding meta-nodes as SW1, but uses DBSCAN and concave hull. By iterative adjusting the two DBSCAN parameters of eps and MinPoints, the best values for this case study are: eps = 0.1 and MinPoints = 10. The resulting footprint of SW2 (Figure 4) shows fewer and smaller potentially affected areas compared to the result of SW1, which are mostly located in the south coast of France and in Corsica.

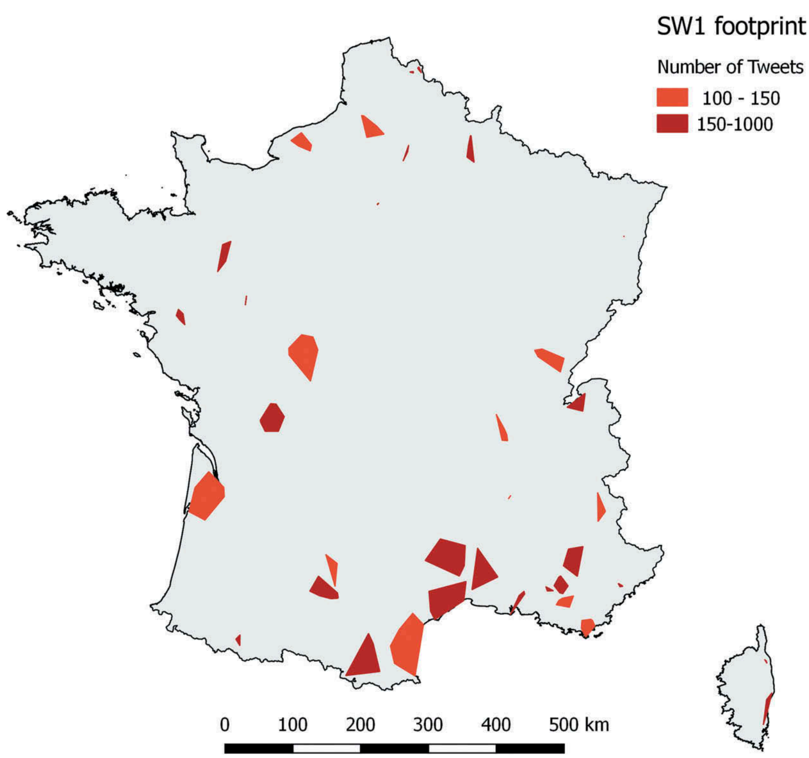

Figure 3. Resulting footprint of SW1.

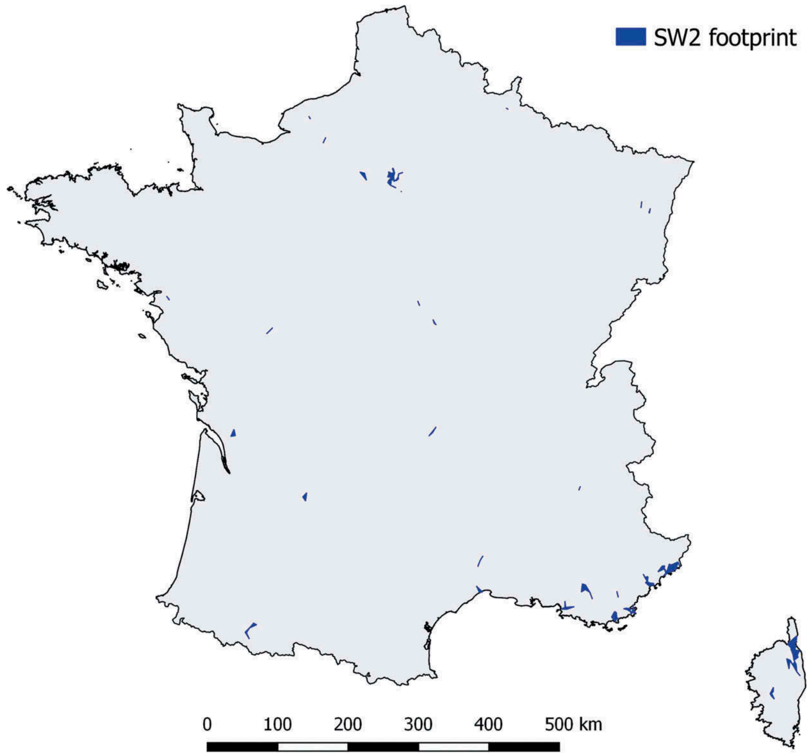

Figure 4. Resulting footprint of SW2.

Since SW3 has fewer input data for the clustering, we used different values for DBSCAN parameters to obtain meaningful results: eps = 1, MinPoints = 50. Figure 5 shows the results of SW3: potentially affected areas are grouped in 5 clusters; non-affected areas are classified as noise. Clusters 0, Cluster 1 and Cluster 4 show a density of municipalities that is substantially higher compared to the other ones, where municipalities are dispersed. Since many wildfires cross municipal boundaries and affect larger areas, it is reasonable to suggest that Cluster 0, Cluster 1 and Cluster 4 identify areas affected by the forest fire event of 25 July 2017.

To understand each SWs performance in identifying affected areas, we compared the above footprints with ground truth data. Active fire data related to the event of 25 July 2017 have been retrieved from the Fire Information for Resource Management System (FIRMS) archive, which provides data derived from satellite observations from both the MODIS and the Visible Infrared Imaging Radiometer Suite in shapefile format. MODIS active fire is selected for comparison.

On 25 July there were a total of 92 active wildfires, of which 88 in southern France. A visual qualitative comparison between the geospatial footprints resulting from the SWs and the MODIS active fire data is reported in Figure 6 for southern France.

For a quantitative comparison, we calculate the percentage of active fires that fall within the geospatial footprints of each SW. For SW1, SW2, and SW3 these are 34.1% (30 out of 88), 39.8% (35 out of 88) and 59.1% (52 out of 88), respectively. Although SW1 identifies bigger areas, these correspond to fewer fires.

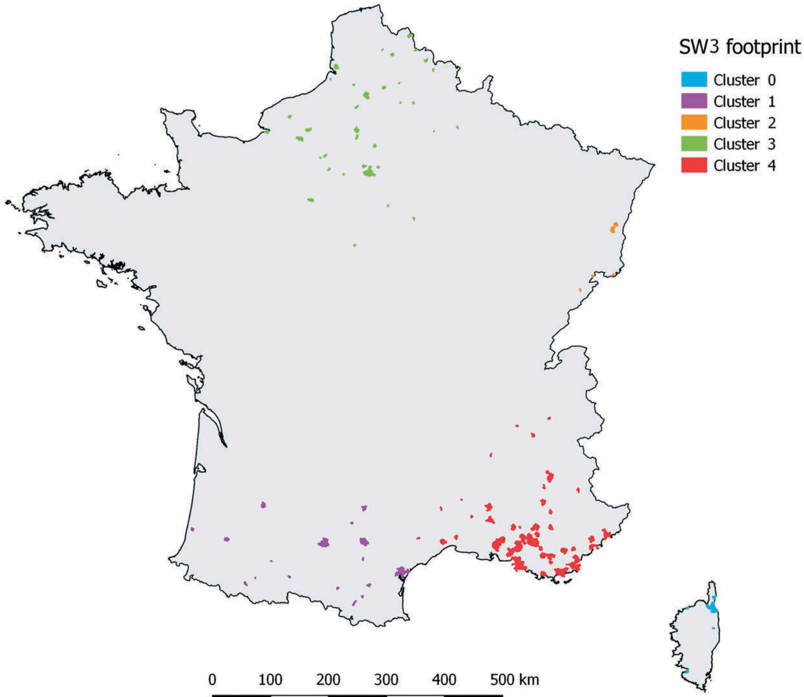

Figure 5. Resulting footprint of SW3.

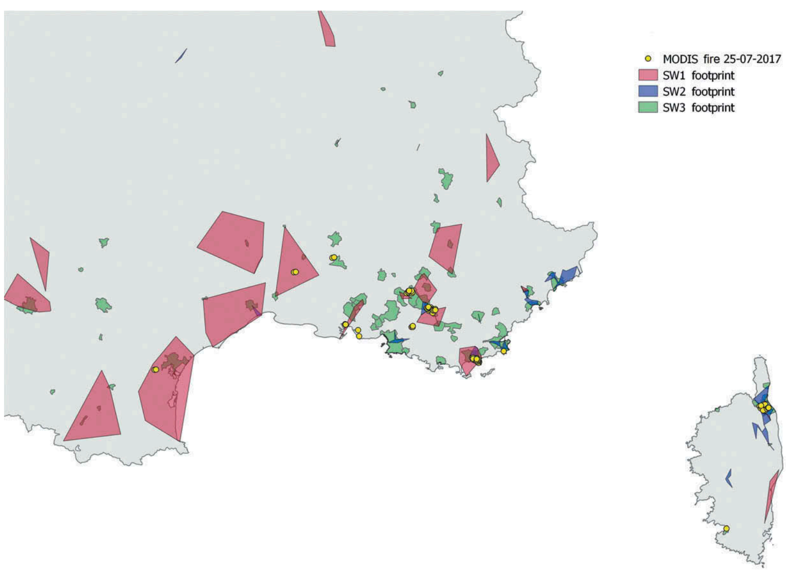

Figure 6. Comparison of resulting footprints from SW1 (red), SW2 (blue) and SW3 (green) with MODIS active fires (yellow dots) in the South of France for 25 July 2017.

Another quantitative measure is to compare the complexity of polygons forming the footprints, in terms of area, number of vertices, amplitude and convexity.

Amplitude (*Ampl*) is the frequency of vibrations of the boundary of the polygon, intended as an increase of the boundary of the polygon compared to the boundary of its convex hull, and can be measured with the following formula (Brinkhoff *et al*. 1995):

$$Ampl(pol) = (boundary\ (pol)\ -\ boundary\ (convex\ hull\ (pol)))/boundary\ (pol)$$

If *Ampl*= 0, the polygon is convex. The higher the amplitude, the longer the boundary is compared to the convex hull.

The convexity *(Conv)* of a polygon is intended as shape deviation of the polygon from its convex hull (Brinkhoff *et al*. 1995):

$$Conv(pol)\ = (area\ (convex\ hull\ (pol)) - area\ (pol))/area\ (convex\ hull(pol))$$

Conv= 0 for convex polygons. The higher the convexity, the higher the deviation of the polygon from its convex hull and the smaller the area.

Both amplitude and convexity are in the interval [0,1]. Values close to 0 indicate convex objects.

The 36 polygons of SW1 have amplitude and convexity of 0 because they are convex and the number of vertices is low (between 4 and 10). Compared to the resulting polygons of SW2 and SW3, SW1 polygons have significant larger average area (578 km^2). SW2 has a similar number of polygons (29), but with a smaller average area (36 km^2). These polygons are more complex with an average convexity of 0.27, and an average amplitude of 0.073. SW3 has more polygons (158) with an average area of 33 km^2 (similar to SW2). SW3 polygons are also more complex: the average convexity is 0.23 and average amplitude is 0.173. These polygons have a high number of vertices which means they have a more detailed shape compared to polygons of SW1 and SW2. For details compare Table 4.

The complexity of the polygons is the result of the shape reconstruction algorithm adopted for SW1 (convex hull) and SW2 (concave hull) and of the geocoding method used for SW3 (which geocodes for municipality areas instead of location points), which also leads to SW3's clusters forming discontinuous areas, while for SW1 and SW2 each

Table 4. Descriptive statistics for the polygons forming footprints of SW1, SW2, SW3.

		SW1	SW2	SW3
Percentage of active fires		34.1%	39.8%	59.1%
Number of polygons		36	29	158
Polygons area (km2)	Average	578.3	36.3	33.1
	Total	20,821.96	1054	5235.26
Number of vertices	Min	4	4	12
	Max	10	63	1053
	Mean	6.47	11.41	102.67
Convexity	Min	0	0	0.059
	Max	0	0.895	0.933
	Mean	0	0.271	0.226
	Sd	0	0.315	0.108
Amplitude	Min	0	0	0.015
	Max	0	0.457	0.494
	Mean	0	0.073	0.173
	Sd	0	0.127	0.104

cluster corresponds to a continuous region. Both footprints resulting from SW2 and SW3 have similar convexity, but the shape of polygons in SW2 depends on the number of Tweets that are in a cluster and their spatial distribution. In SW3 the shape of the polygons corresponds to municipality units and it is independent of the input dataset.

Discussion

Geospatial footprints to represent areas affected by wildfires can be used for an initial evaluation and assessment of the impact of the event by emergency managers. Shape complexity needs to be considered when using the footprints as an emergency tool: the geospatial footprint resulting from SW2, even though it covers almost 40% of the active fires, identifies polygons with complex shapes that give non-meaningful insights in terms of affected areas, while SW1's footprint returns very large affected areas which can be misleading. In addition, we can also consider the impact of the number of Tweets in the polygons that contain active fires: the SW with higher number of Tweets in these areas performs better than the others. The count of Tweets is 509, 144 and 1569 in SW1, SW2 and SW3 polygons, respectively. For the case study of this paper, the resulting geospatial footprint of SW3 performs the best in identifying active fires and affected areas around them, even if these areas are quite discontinuous and correspond to municipality units, while the extent of bushfire events not always matches administrative boundaries. The number of Tweets used as input data and their spatial distribution have a significant impact on the size, shape and number of polygons generated by SW1 and SW2, so for a different dataset these two workflows may perform better than SW3.

The quantitative comparison of resulting footprints with MODIS active fires represents a first step for evaluating the accuracy of the results. Further investigation into the utility of the generated footprints could involve a more detailed quantitative comparison of footprint area (i.e. match between footprint and ground truth of burned area), and an extensive user evaluation in which practitioners test and comment on the generated footprint. Both are out-of-scope for this paper, but based on our results we can already propose that any analysis of footprint surface and burned area would require additional datasets, because the CGSM is likely to be too sparse and heterogeneously distributed to allow precise delineation of burned areas.

Recent studies such as Spinsanti and Ostermann (2013) have already shown that CGSM can contribute to detection of wildfire events. CGSM may not be used as standalone source of information for precise identification of events boundaries but can still be a useful resource to determine hotspots, most affected areas or areas at risk that do not yet appear on any earth observation products.

Social media represent a communication channel that cannot be ignored in the analysis of disaster events. Personal communication with disaster managers during events organised by the GEO-SAFE project (GEO-SAFE 2016, 2017, 2018) has revealed that they monitor social media networks in order to identify and reduce the spread of misinformation, which is rarely intentional, but still poses a significant threat. In case of higher data volumes, this would also require NLP techniques to pre-filter the incoming information based on content. Additional content analysis may also provide useful details about the event (e.g. blocked roads, shelters, etc.).

In order to determine which additional datasets are needed for a comprehensive footprint, a wider user evaluation is required. Such a user evaluation should be facilitated by the results of the present study, as our aim to increase reproducibility would also ease comparable user evaluations. Such user evaluations would be an important step towards producing ready-to-use footprints, which are out of this paper's scope.

Conclusion

Research on geospatial analysis of CGSM uses many different data sets and sources, analysis tools, and methods at different geographic scales. To improve a researcher's or practitioner's ability to benefit from the diverse options available, we need to facilitate reproduction and replication of studies. To contribute to improved reproducibility, we demonstrated the replication of three different geospatial disaster analytics from the published literature in SWMS. We then showed how the resulting replicated scientific workflows obtain geospatial footprints of bushfire events from Twitter data, followed by a systematic, qualitative and quantitative comparison and evaluation of key common operations and results between workflows and with respect to ground truth data.

Reproducibility and replicability of scientific studies is fundamental to ensure the advancement of knowledge. As described in Nust *et al*. (2018), for a work to be fully reproducible, input data, analysis methods including code, and detailed results need to be available. None of the original selected papers provide such full reproducibility, lacking in input data or details about used tools and software. This led us to replicate the studies, i.e. we had to substitute or alter the workflow from the original. The spatio-temporal extent of the case study data affected the choice of the clustering algorithm used in SW2 and SW3 and allowed us to replace ST-DBSCAN – which is currently not available in KNIME – with DBSCAN. Two of the papers (Spinsanti and Ostermann 2013, Saravanou *et al*. 2015) used Twitter data, for which Twitter's terms of service prohibit free sharing. In this case, sharing only the tweets IDs provides a step towards better reproducibility.

The adaptability of the chosen workflow management system (KNIME) to different input data formats, databases and real-time connections to the Twitter Streaming API and the Twitter Search API allows it to deal with the different types and formats of crowdsourced data. The possibility to aggregate nodes into metanodes allows the user to organise tasks into modules that can be easily reused when creating a new workflow. The numerous available nodes for data mining in KNIME represent an additional potential for testing different data filtering methods. In addition, the system lets the user visualise the intermediate and final results internally without the need to export them into other systems; this allows the user to easily modify parameters and adjust steps as they are created.

While our results are promising, much work remains before scientists and practitioners alike can benefit from shared, reproducible workflows to improve disaster response. In principle, the approach is system agnostic, i.e. our use of KNIME is the result of matching functionality and convenience. Its free and open source character makes it in principle available to everyone. However, this paper also highlights the fundamental problem of using data sets which are publicly available but for which terms of service prohibit free re-distribution. This hints at the need to have reference implementations using common benchmarks for comparison, so that potential users can make an informed decision based on their own available data, which workflow promises the best results. Such an open repository has yet to be developed.

Acknowledgements

We acknowledge the Marie-Curie GEO-SAFE (Geospatial Based Environment For Optimization Systems Addressing Fire Emergencies) project (691161).

Disclosure statement

No potential conflict of interest was reported by the authors.

ORCID

V. Cerutti http://orcid.org/0000-0002-9612-1581
C. Bellman http://orcid.org/0000-0003-1679-8216
A. Both http://orcid.org/0000-0003-4021-9632
M. Duckham http://orcid.org/0000-0002-7249-6709
B. Jenny http://orcid.org/0000-0001-6101-6100
R. L. G. Lemmens http://orcid.org/0000-0001-5269-6343
F. O. Ostermann http://orcid.org/0000-0002-9317-8291

References

Alani, H., Christopher, B.J., and Tudhope, D., 2010. Voronoi-based region approximation for geographical information retrieval with gazetteers. *International Journal of Geographical Information Science*, 15 (4), 287–306. doi:10.1080/13658810110038942

Alexander, D.E., 2013. Social media in disaster risk reduction and crisis management. *Science and Engineering Ethics*, 1–17. Available from: http://www.ncbi.nlm.nih.gov/pubmed/24306994

Amenta, N., Choi, S., and Krishna, R., 2001. The power crust. *Sixth ACM symposium on solid modeling and applications*. ACM, 249–260.

Arampatzis, A., *et al*. 2006. web-based delineation of imprecise regions. *Computers, Environment and Urban Systems*, 30 (4), 436–459. doi:10.1016/j.compenvurbsys.2005.08.001

Bellman, C., *et al*., 2018. Open spatial analytics. Available from: https://sites.rmit.edu.au/openspatialanalytics/osa-node-rep/ [Accessed 19 June 2019].

Berthold, M.R., *et al*. 2009. KNIME – The Konstanz Information Miner: version 2.0 and beyond. *ACM SIGKDD Explorations Newsletter*, 11 (1), 26–31. doi:10.1145/1656274.1656280

Birant, D. and Kut, A., 2007. ST-DBSCAN : an algorithm for clustering spatial – temporal data. *Data & Knowledge Engineering*, 60 (1), 208–221. doi:10.1016/j.datak.2006.01.013

Brinkhoff, T., *et al*., 1995. Measuring the complexity of polygonal objects. *Proceedings of the 3rd {ACM} international workshop on advances in geographic information* systems, *Baltimore*, Maryland, *in Conjunction with {CIKM}*, 1–2 December 1995. 109.

Capineri, C., *et al*., 2016. *European handbook of crowdsourced geographic information*. London: Ubiquity Press.

Cerutti, V., *et al*., 2016. Identification of disaster-affected areas using exploratory visual analysis of georeferenced tweets : application to a flood event. *Association of geographic information laboratories in Europe (AGILE) proceedings*, Helsinki, Finland, Short Papers.

Cheng, Z., Caverlee, J., and Lee, K., 2010. You are where you tweet: a content-based approach to geo-locating twitter users. *Proceedings of the 19th ACM international conference on information and knowledge management*. ACM, 759–768.
Edelsbrunner, H., Kirkpatrick, D.G., and Seidel, R., 1983. On the shape of a set of points in the plane. *IEEE Transactions on Information Theory*, 29 (4), 551–559. doi:10.1109/TIT.1983.1056714
Editorials, 2016. Reality check on reproducibility. *Nature*, 533 (7604), 437. doi:10.1038/533437a
Ester, M., *et al*., 1996. A density-based algorithm for discovering clusters in large spatial databases with noise. *In Kdd,* 96 (34), 226–231.
Flanagin, A.J. and Metzger, M.J., 2008. The credibility of volunteered geographic information. *GeoJournal*, 72 (July), 137–148. doi:10.1007/s10708-008-9188-y
Galton, A. and Duckham, M., 2006. What is the region occupied by a set of points ? *Lecture Notes in Computer Science*, 4197, 81–98.
GEO-SAFE, 2016. EU-Australia GEO-SAFE think-tank. Available from: https://sites.rmit.edu.au/rmitopt/2016/11/14/eu-australia-geo-safe-think-tank/.
GEO-SAFE, 2017. *GEO-SAFE mid-term review meeting*, Bastia, France.
GEO-SAFE, 2018. Summary of the workshop on stochastic mapping of bushfires. Available from: http://geosafe.lessonsonfire.eu/2018/02/07/summary-of-the-workshop-on-stochastic-mapping-of-bushfires-29-30-january-2018/ [Accessed 26 June 2019].
Grajales, F.J., *et al*., 2014. Social media: a review and tutorial of applications in medicine and health care. *Journal of Medical Internet Research*, 16 (2), 13.
Granell, C. and Ostermann, F.O., 2016. Beyond data collection: objectives and methods of research using vgi and geo-social media for disaster management. *Computers, Environment and Urban Systems*, 59, 231–243. doi:10.1016/j.compenvurbsys.2016.01.006
Haworth, B., Bruce, E., and Middleton, P., 2012. Emerging technologies for risk reduction : assessing the potential use of social media and vgi for increasing community engagement. *Australian Journal of Emergency Management*, 30 (3), 36–41.
Kongthon, A., *et al*., 2012. The role of twitter during a natural disaster : case study of 2011 Thai flood. *Technology Management for Emerging Technologies (PICMET): 2012 proceedings of PICMET '12*, Vancouver, BC, Canada, 2227–2232.
Macqueen, J., 1967. Some methods for classification and analysis of multivariate observations. *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, 1 (233), 281–297.
McPhillips, T., *et al*., 2009. Scientific workflow design for mere mortals. *Future Generation Computer Systems*, 25(5), 541–551. doi:10.1016/j.future.2008.06.013
Miller, H.J. and Goodchild, M.F., 2015. Data-driven geography. *GeoJournal*, 80(4), 449–461. doi:10.1007/s10708-014-9602-6
Nust, D., *et al*., 2018. Reproducible research and giscience: an evaluation using agile conference papers. *PeerJ*, 6, e5072. doi:10.7717/peerj.5072
Ostermann, F. and Spinsanti, L., 2012. Context analysis of volunteered geographic information from social media networks to support disaster management. *International Journal of Information Systems for Crisis Response and Management*, 4 (4), 16–37. [Accessed 12 Jan 2015]. doi:10.4018/jiscrm.2012100102
Ostermann, F.O. and Granell, C., 2017. Advancing science with VGI: reproducibility and replicability of recent studies using VGI. *Transactions in GIS*, 21 (2), 224–237. doi:10.1111/tgis.12195
Saravanou, A., *et al*., 2015. Twitter floods when it rains: a case study of the UK floods in early 2014. *WWW '15 companion proceedings of the 24th international conference on world wide web companion*. 1233–1238. Available from: http://dl.acm.org/citation.cfm?id=2740908.2741730
Scassa, T., 2013. Legal issues with volunteered geographic information. *Canadian Geographer*, 57 (1), 1–10. doi:10.1111/j.1541-0064.2012.00444.x
Schade, S., *et al*., 2012. Semantic observation integration. *Future Internet*, 4 (4), 807–829. Available from: http://www.mdpi.com/1999-5903/4/3/807/

Scheider, S., Ostermann, F.O., and Adams, B., 2017. Why good data analysts need to be critical synthesists. determining the role of semantics in data analysis. *Future Generation Computer Systems*, 72, 11–22. doi:10.1016/j.future.2017.02.046

Shelton, T., Poorthuis, A., and Zook, M., 2015. Social media and the city: rethinking urban socio-spatial inequality using user-generated geographic information. *Landscape and Urban Planning*, 142, 198–211. doi:10.1016/j.landurbplan.2015.02.020

Spinsanti, L. and Ostermann, F.2013. Automated geographic context analysis for volunteered information. *Applied Geography*, 43, 36–44. Available from: http://linkinghub.elsevier.com/retrieve/pii/S0143622813001185 [Accessed 16 Dec 2014].

Zahra, K., *et al*., 2017. Geo-spatial information science geographic variability of twitter usage characteristics during disaster events. *Geo-Spatial Information Science*, 5020(November), 1–10. doi:10.1080/10095020.2017.1371903

Zhong, X., *et al*., 2016. Real-time estimation of wildfire perimeters from curated crowdsourcing. *Scientific Reports*, 6, 24206. Available from: http://www.nature.com/srep/2016/160411/srep24206/full/srep24206.html