3.2. A Technology and Standards for Geospatial Workflow

FACILITATING FLOOD RESPONSE USING GEOSOCIAL MEDIA

Submitted by:

VIGNESWARAN S,

DEV DINESH

M.TECH IN REMOTE SENSING & GIS

Specialization in Geoinformatics

Submitted to:
Shiva Reddy Koti
Scientist/Engineer-SE
Geoinformatics Department

SESSION: 2021-2022

Government of India
Department of Space
Indian Space Research Organisation
Indian Institute of Remote Sensing
Dehradun- 248001, Uttrakhand, India

July, 2022

TABLE OF CONTENT

ABSTRACT

There are several studies that have shown the ability of satellite remote sensing in monitoring floods, however, they cannot be immediately used during disaster due to acquisition delays. Immediate flood monitoring and mapping are demanding undertakings. This study is carried out to assess the feasibility of facilitating flood response using geosocial media. The Twitter API is particularly helpful in gathering information from users within the hazard zone, and this information is used to create geographical footprints. The data can be seen in a web application, and online users can validate the data themselves by marking it as so.

1. INTRODUCTION

The quantity of available Geoinformation from remote sensing, crowdsourced geosocial media (CGSM), and geo-computation techniques have greatly increased during the past two decades. There is a need to create a reproducible scientific workflow for processing the CGSM data for different applications. Determining affected or vulnerable areas of flood from CGSM is a crucial and challenging step in successful disaster response for which geosocial media as of now has still untapped potential[1].

Crowdsourced Geosocial Media (CGSM)

Participatory mapping and crowdsourcing have several uses in the crisis domain. Following a catastrophe, several communities will share information about the occurrence via social media like Twitter and online mapping. A crowdsourcing framework has been developed by CEGIS researchers to direct the creation of upcoming crowdsourcing projects for hazards science studies at the USGS. To investigate how diverse crowds might use remotely sensed aerial photographs to estimate sea level rise and hurricane damage to coastal locations in the wake of significant storms, one project is currently being developed. The examination of coastal photography by the general public has the potential to enhance coastal change prediction models. This study's main objective is to examine the benefits and drawbacks of combining official and crowdsourced geospatial data about dangers for both academic and practical emergency management applications.

Volunteered Geographic Information (VGI)

Government entities at all levels can improve their geospatial databases by leveraging volunteered geographic information, or geospatial content created by non-experts using mapping tools available online. VGI's alleged lack of accuracy is frequently mentioned as a deterrent to official mapping organisations using it more frequently. According to CEGIS research, certain participatory mapping initiatives can generate data that is just as precise as that generated by these organisations. Furthermore, because contributors have specialised local knowledge, the "eyes on the ground" of VGI occasionally outperform more expensive accuracy tests by official authorities. In an early stage of The National Map Corps, volunteers submitted data, which CEGIS researchers examined for accuracy. These volunteers' information on buildings—schools, hospitals, and the like—turned out to be accurate enough to be added to the official databases that make up The National Map. Researchers at CEGIS are also examining user motives, modifications to institutional cultures and economic models, the emergence of online communities, and legal restrictions on VGI.

Geospatial Scientific workflow

Environmental modelling requires the functionality of geospatial analytical systems. Software reuse and sharing become more and more crucial in integrated environmental modelling, testing, and analysis as large scale distributed geographic data becomes available. The Service Oriented Architecture (SOA) promotes reuse and interoperability of components on the Web, boosting the efficiency of assembly and lowering the cost of development. It also permits cooperation of data and process components among various organizational units. In recent years, a number of fields have seen an increase in the demand for flexible tools and interfaces for retrieving scientific data and conducting in-depth analysis on it.

These studies can be represented as scientific workflows where the data flow from one analytical step to the next is specified using a formal workflow language. Scientific workflows are often data- and/or compute-intensive, dataflow-oriented, and frequently entail data transformations, analysis, and simulations.

Reproducibility

Replicability refers to situations in which a researcher gathers new data to reach the same scientific conclusions as a previous study, but "reproducibility" refers to situations in which the original researcher's data and computer programmes are utilised to regenerate the results.

The terms "reproducibility" and "replicability" are used inconsistently or even contradictorily by various scientific disciplines and organisations. These phrases—as well as others, such repeatability—have long been associated with the idea that one experiment or study should confirm the findings of another. However, no terminologically coherent method of making distinctions within this general concept has arisen; instead, competing and contradictory words have proliferated. The lack of accepted standards for reproducibility and replicability makes it more challenging to evaluate these concepts.

Scientific workflow management system

A group of scientific jobs are typically streamlined as part of a scientific workflow that is focused with automating procedures in order to facilitate and hasten scientific discoveries. The main benefit of employing a scientific process is that it enables quick access to a variety of complex computer technologies without requiring knowledge of the underlying concepts. Every analytical step is presented visually, and we may easily reuse all or a portion of a workflow. Systems for managing scientific workflows define, control, and carry out the processes in diverse computing settings.

Geospatial analysis

When applied to geographic models, geospatial analysis is the collection, visualisation, and manipulation of imagery, GPS, satellite photography, and historical data that is either explicitly described in terms of geographic coordinates or implicitly in terms of a street address, postal code, or forest stand identifier. Crisis management, climate change modelling, weather monitoring, sales analysis, human population forecasts, and animal population management are just a few of the many uses for geospatial analysis.

Twitter API

Due to social media's widespread use, many social media platforms are becoming more and more well-liked as a source of data. Data collecting utilising APIs is evolving into a highly sought-after expertise in many data science positions as a result of the emergence of social media as a data source. With approximately 400 million active monthly users, Twitter currently has a vast amount of data that can be gathered, the majority of which is public. Additionally, the Twitter developer team just released the Twitter API v2, which was created from scratch and is available from second half of 2020.

KNIME Analytics Platform

Data science has recently dominated our daily lives, giving rise to a plethora of data analysis tools that the average Joe may use. Some of the most popular tools are KNIME Analytics Platform, Python, and R. The KNIME Analytics Platform's visual programming environment and user-friendly interface, which supports a wide range of technologies, are what give it its revolutionary nature. For the entire development, KNIME offers a graphical interface (a user-friendly GUI). Simply define the workflow between the different preconfigured nodes offered in KNIME's repository. KNIME offers a number of preset elements known as nodes for a variety of activities include reading data, using different ML algorithms, and presenting data in different forms. As a result, no programming experience is necessary to use KNIME.

1.1 PROBLEM DESCRIPTION

During disaster satellite data cannot be immediately used for mapping flood-affected areas. Alternately we can use geosocial media data for the creation of the geospatial footprint of flood-affected regions.

1.2 OBJECTIVES

To create a reproducible scientific workflow for processing the crowdsourced geosocial media data and disseminating the results over web

Sub-Objective

- Identification of areas affected by the flood.
- Data visualization over the Web.
- Use crowdsourcing for validation of results.

2. LITERATURE REVIEW

The potential for undesirable outcomes, such as spreading rumours, weakening authority, and supporting terrorist attacks, balances out the beneficial aspects of social media. This prompts a discussion on the morality of using social media during emergencies. Despite some hazards that are readily apparent, such as the invasion of privacy, it seems that ethical public consensus will typically prevail over dishonest attempts to manipulate the media. Social media is a powerful tool for revealing fraud and corruption, too. In conclusion, the widespread adoption and usage of social media by the general population around the world ushers in a new era in which emergency managers must immediately modify their operational procedures to meet the challenge and opportunities of this development[2].

Twitter is a popular microblogging site for the quick dissemination of opinions, ideas, and knowledge. It has been routinely utilised to disseminate appeals for assistance, convey evacuation plans, and aid in damage assessment during catastrophes. The accuracy of this information is crucial for making decisions in a crisis situation, yet it can be challenging to evaluate. The ability of quality evaluation techniques to be transferred from one geographic region to another has received little prior examination. This study's key contribution is its examination of the patterns of Twitter use by individuals based in various places during emergencies. We look into tweeting activity related to two earthquakes that occurred in Myanmar and Italy We evaluate the performance of Nave Bayes models for categorising Tweets when used on data from a different region than the one used to train the model, as well as the granularity of geographic references used, user profile traits that are related to credibility, and performance of user profiles that are related to credibility. Our findings indicate that the spatial granularity for the earthquake occurrences in Italy and Myanmar is comparable, but the information from surrounding locations is less for the earthquake event in Myanmar than for Italy. Despite the fact that user and usage characteristics fluctuate significantly and in complex ways, the Nave Bayes classifier performs well even when used on data from different geographical areas[3].

The way geographic information is gathered, shared, and used has altered as integrated mobile devices connected to social networks become more widely used. Numerous studies have already looked into how social media affects crisis situations. Although networks of volunteers have shown they can consistently curate a lot of material, there are problems with sustainability and scalability with this method. Therefore, the authors suggest a semi-automatic method for gathering geographic data that has been voluntarily shared on social media, assessing its quality, and then using it in a crisis situation. The system proposed is new in that it makes use of (geographic) context to assess quality and utility rather of concentrating on specific bits of information[5].

3. DATASETS AND METHODOLOGY

3.1 Pre-Requisites

- Twitter developer account with Elevated Access to Twitter API v2 endpoint.
- API Key for Google Maps Platform

3.2 Datasets Used

Twitter data which is freely accessible from Twitter API v2 endpoint is used for analysis. Twitter developer account was created prior to accessing the API.

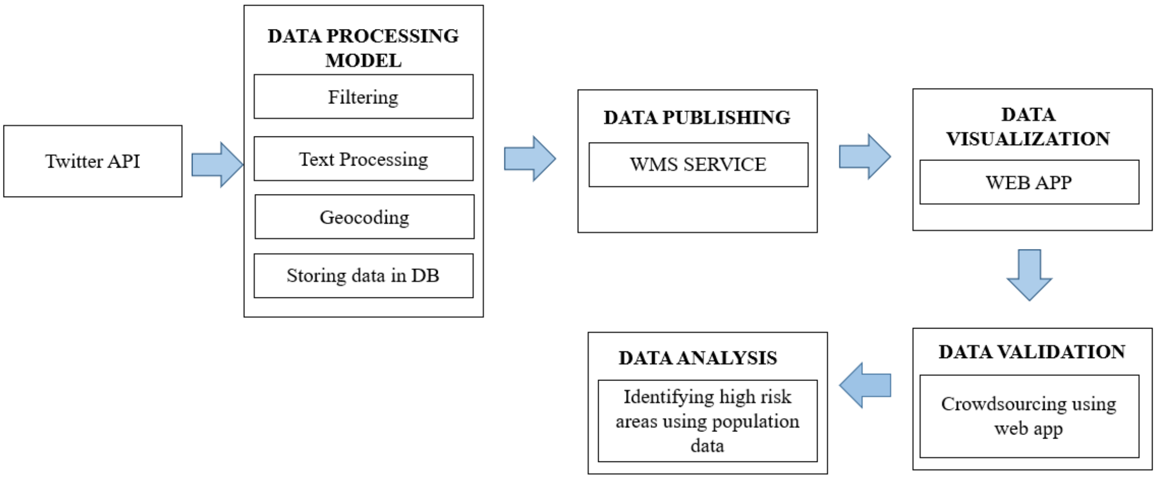

Figure 1: Methodology

3.3 Software and API used

API

- Twitter API v2
- Google Geocoding API

Software

- KNIME Analytics Platform with extensions
- PostgreSQL + PostGIS
- Geoserver

Leaflet Javascript library was used to create web application

3.4 BPMN Diagram

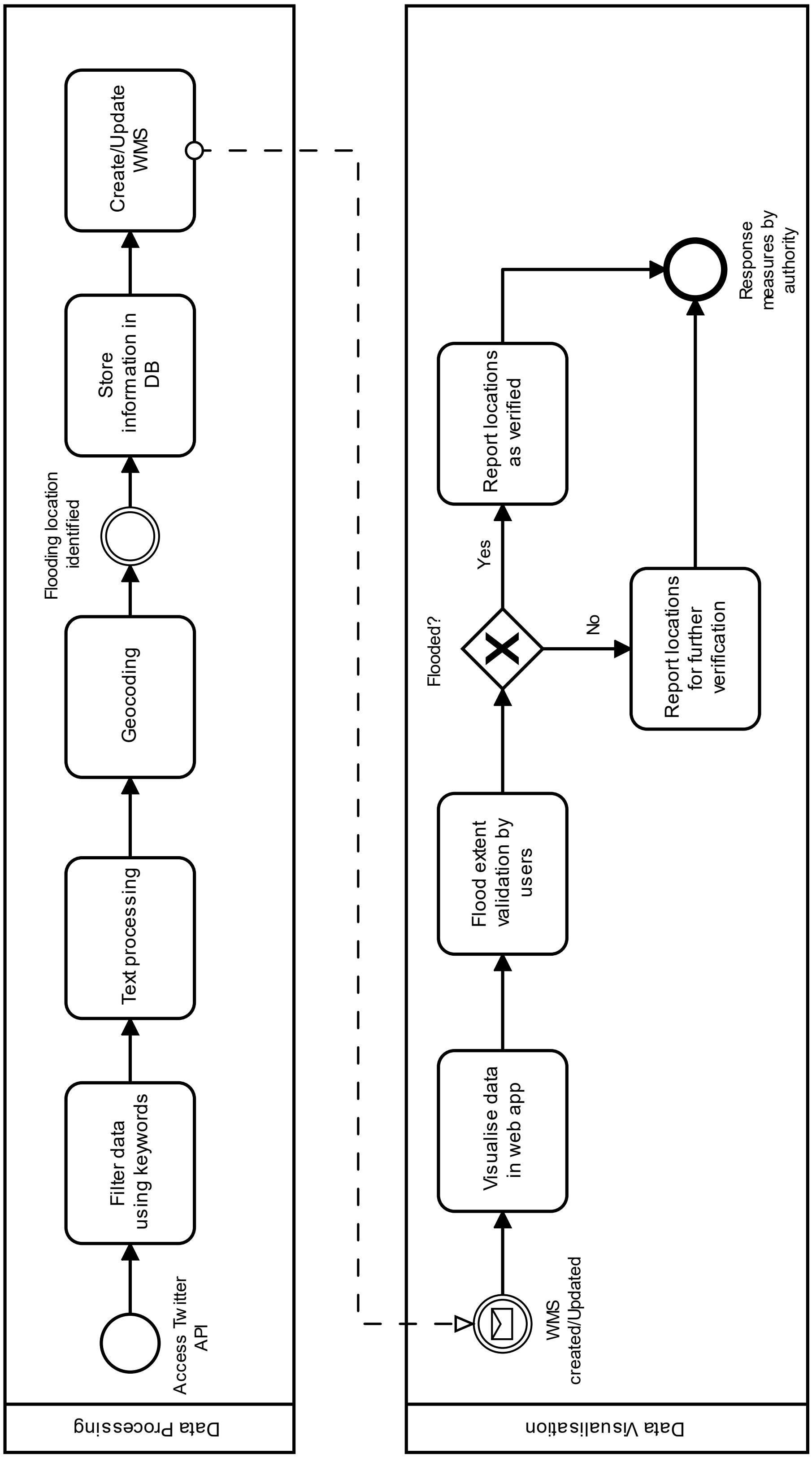

Figure 2: BPMN Diagram

3.5 Data Processing with KNIME Analytics Platform

The data processing flow created from KNIME is shown below.

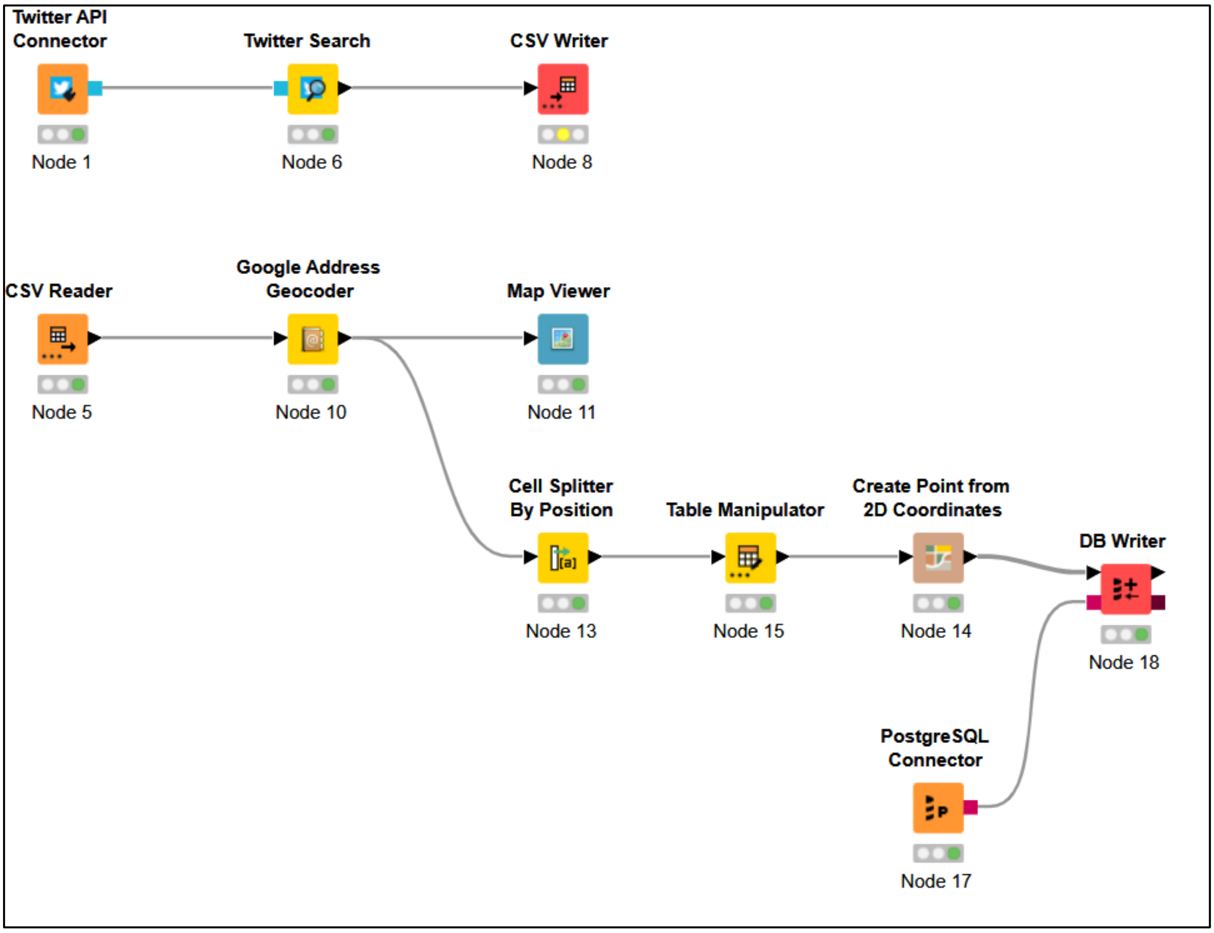

Figure 3: KNIME Workflow

KNIME software consists of readily available nodes for different functionalities. The description for each node is provided below.

1. **Twitter API Connector**: Creates a connection to access Twitter's API.
2. **Twitter Search**: Search over Twitter's API.
3. **CSV Writer**: Writes out the input data table into a file or to a remote location denoted by an URL
4. **CSV Reader**: Reads CSV files
5. **Google Address Geocoder**: This node allows geocoding of street addresses, such as “1600 Amphitheatre Parkway, Mountain View, CA”, to geographic coordinates, such as (37.423021, -122.083739) using Google’s geocoding API.

6. **Map Viewer**: This node allows to display geographical locations on a map
7. **Cell Splitter By Position**: Splits the content of a selected column into several separate new columns.
8. **Table Manipulator**: Allows to perform several column transformations on any number of input tables such as renaming, filtering, re-ordering and type changing of the input columns.
9. **Create Point From 2D Coordinates**: Takes two columns containing coordinates and changes it into Point geometries.
10. **PostgreSQL Connector**: This node creates a connection to a PostgreSQL server via its JDBC driver.
11. **DB Writer**: Inserts data rows into the database based on the selected columns from the input table.

3.6 Data Visualization over web

The data saved in the database is published as a new WMS service using Geoserver. Then, Leaflet application was created to display the WMS service in an interactive map.

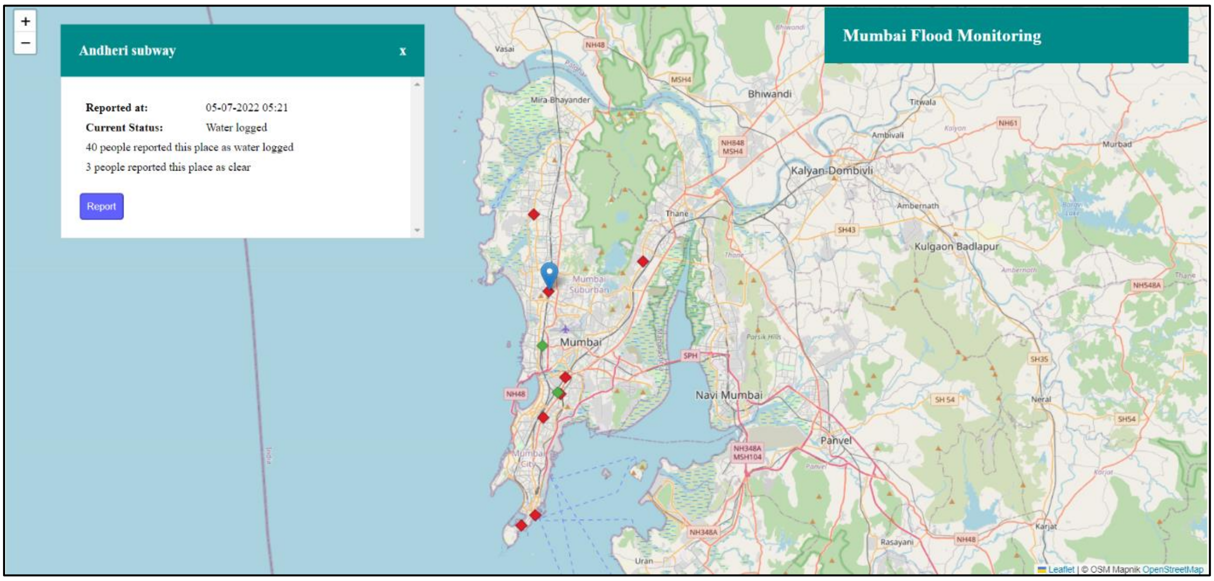

Figure 4: Web Application

Users of the web application have a functionality to verify the authenticity of the map data. So crowdsourcing is required for validating the results of the geospatial process.

4. RESULTS AND DISCUSSIONS

In this work, a reproducible scientific workflow for processing the crowdsourced geosocial media data was created which can be used during disaster or emergency to obtain valuable insights. A web app was also developed for information dissemination over the web. This approach can be refined and used by different users all over the world. Almost all the tasks in the workflow are automated and require minimal human intervention. This will be helpful during any emergency. KNIME Server, a commercial offering from KNIME also supports tasks scheduling which can be used to retrieve live tweets and process them in real time. This can be further explored in future studies.

5. REFERENCES

[1] V. Cerutti *et al.*, "Improving the reproducibility of geospatial scientific workflows: the use of geosocial media in facilitating disaster response," *https://doi.org/10.1080/14498596.2019.1654944*, vol. 66, no. 3, pp. 383–400, 2019, doi: 10.1080/14498596.2019.1654944.

[2] D. E. Alexander, "Social media in disaster risk reduction and crisis management," *Sci. Eng. Ethics*, vol. 20, no. 3, pp. 717–733, 2014, doi: 10.1007/S11948-013-9502-Z.

[3] K. Zahra, F. O. Ostermann, and R. S. Purves, "Geographic variability of Twitter usage characteristics during disaster events," *http://www.tandfonline.com/action/journalInformation?show=aimsScope&journalCode=tgsi20#.VsXpLiCLRhE*, vol. 20, no. 3, pp. 231–240, Jul. 2017, doi: 10.1080/10095020.2017.1371903.

[4] C. Granell and F. O. Ostermann, "Beyond data collection: Objectives and methods of research using VGI and geo-social media for disaster management," *Comput. Environ. Urban Syst.*, vol. 59, pp. 231–243, Sep. 2016, doi: 10.1016/J.COMPENVURBSYS.2016.01.006.

[5] F. Ostermann and L. Spinsanti, "Context Analysis of Volunteered Geographic Information from Social Media Networks to Support Disaster Management: A Case Study on Forest Fires," *https://services.igi-global.com/resolvedoi/resolve.aspx?doi=10.4018/jiscrm.2012100102*, vol. 4, no. 4, pp. 16–37, Jan. 1AD, doi: 10.4018/JISCRM.2012100102.