School of Computing

Innovative Big Data Integration and Analysis Techniques for Urban Hazard Management

Jedsada Phengsuwan

Submitted for the degree of Doctor of Philosophy in the School of Computing, Newcastle University, UK

December 25, 2020

Abstract

Modern early warning systems (EWS) require sophisticated knowledge of natural hazards, the urban context and underlying risk factors to enable dynamic and timely decision making (e.g., hazard detection, hazard preparedness). Landslides are a common form of natural hazard with a global impact and are closely linked to a variety of other hazards. EWS for landslide prediction and detection relies on scientific methods and models which require input from the time-series data, such as the earth observation (EO) and ancillary data. Such data sets are produced by a variety of remote sensing satellites and Internet of Things sensors which are deployed in landslide-prone areas. Besides, social media-based time-series data has played a significant role in modern disaster management. The emergence of social media has led to the possibility of the general public contributing to the monitoring of natural hazard by reporting incidents related to hazard events. To this end, the data integration and analysis of potential time-series data sources in EWS applications have become a challenge due to the complexity and high variety of data sources. Moreover, sophisticated domain knowledge of natural hazards and risk management are also required to enable dynamic and timely decision making about serious hazards. In this thesis, a comprehensive set of algorithmic techniques for managing high varieties of time series data from heterogeneous data sources is investigated. A novel ontology, namely Landslip Ontology, is proposed to provide a knowledge base that establishes the relationship between landslide hazard and EO and ancillary data sources to support data integration for EWS applications. Moreover, an ontology-based data integration and analytics system that includes human in the loop of hazard information acquisition from social media is proposed to establish a deeper and more accurate situational awareness of hazard events. Finally, the system is extended to enable an interaction between natural hazard EWS and electrical grid EWS to contribute to electrical grid network monitoring and support decision-making for electrical grid infrastructure management.

Publications

Published

1. Rajiv Ranjan, **Jedsada Phengsuwan**, Philip James, Stuart Barr, and Aad van Moorsel. 2017. “Urban Risk Analytics in the Cloud”. *IT Professional* 19, 2 (March 2017), 4-9. DOI: https://doi.org/10.1109/MITP.2017.20

2. **Jedsada Phengsuwan**, Nipun Balan TH, and Rajiv Ranjan ,“Onto-DIAS: Ontology-based Data Integration and Analytics System for Landslide hazard Early Warning”, *EGU General Assembly 2019*, 7-11 April, 2019, Vol. 21, EGU2019-11866, Vienna, Austria

3. **Jedsada Phengsuwan**, Tejal Shah, Philip James, Dhaval Kumar Thakker, Stuart Barr, Rajiv Ranjan , “Ontology-based Discovery of Time-Series Data Sources for Landslide Early Warning System”, *Springer Computing, S.I : Big time series data*, Computing 102, 745–763 (2020). https://doi.org/10.1007/s00607-019-00730-7

4. **Jedsada Phengsuwan**, Nipun Balan Thekkummal, Tejal Shah, Philip James, Dhaval Kumar Thakker, Rui Sun, Divya Pullarkatt, T. Hemalatha, Maneesha Vinodini Ramesh, Rajiv Ranjan , “Context-Based Knowledge Discovery and Querying for Social Media Data,” *2019 IEEE 20th International Conference on Information Reuse and Integration for Data Science (IRI)*, Los Angeles, CA, USA, 2019, pp. 307-314, doi: 10.1109/IRI.2019.00056.

5. **Phengsuwan, J**, Shah, T, Sun, R, James, P, Thakker, D, Ranjan, R. “An ontology-based system for discovering landslide-induced emergencies in electrical grid”, *Trans Emerging Tel Tech*, 2020;e3899. https://doi.org/10.1002/ett.3899

6. Zhenyu Wen, **Jedsada Phengsuwan**, Nipun Balan Thekkummal, Rui Sun, Pooja Jamathi Chidananda, Tejal Shah, Philip James and Rajiv Ranjan, 2020,

"Active Hazard Observation via Human in the Loop Social MediaAnalytics System" In Proceedings of the 29th ACM International Conference on Information and Knowledge Management (CIKM '20). Association for Computing Machinery, Virtual Event, Ireland, 3469–3472, DOI:https://doi.org/10.1145/3340531.3417430

7. D. Contreras, S. Wilkinson, N. Balan, **J. Phengsuwan**, P. James, "Assessing Post-Disaster Recovery Using Sentiment Analysis. The case of L'Aquila, Haiti, Chile and Canterbury", *The 17th World Conference on Earthquake Engineering (17WCEE)*, September, 2020;9c-0019, Sendai, Japan,

8. **Phengsuwan, J**.; Shah, T.; Thekkummal, N.B.; Wen, Z.; Sun, R.; Pullarkatt, D.; Thirugnanam, H.; Ramesh, M.V.; Morgan, G.; James, P.; Ranjan, R. Use of Social Media Data in Disaster Management: A Survey. Future Internet 2021, 13, 46. https://doi.org/10.3390/fi13020046

Acknowledgements

I want to acknowledge everyone who played a role in my academic accomplishments. My supervisor, Prof. Rajiv Ranjan, has provided patient advice, guidance, and all supports throughout the research process. Dr Tejal Shah, who has provided excellent advice and guidance for almost of my publications and research. Thank you, Prof. Philp James, who has advised and guidance on my research's real application.

I want to thank you all of my friends in Newcastle for the enjoyable moments while living in Newcastle.

Finally, my parents, who supported me with love and understanding, I could never have reached this current level of success without you.

Contents

List of Figures

List of Tables

Listings

Acronyms

LO	Landslip Ontology
EO	Earth Observation
EWS	Early Warning System
DPM	Department of Disaster Prevention and Mitigation
SSN	Semantic Sensor Network
WSN	Wireless Sensor Network

1

Introduction

1.1 Background and Motivation

Natural disasters are a consequence of natural hazard impacts on human settlement and their exposure [1]. The occurrence of a natural disaster in urban areas can cause the loss of life, the damage to properties and infrastructures, and economic loss and disruption. Efficient urban hazard management thus requires decision support systems to provide preparation and mitigation strategies for urban disaster risk reduction and disaster resilience. The advancement of Early Warning Systems (EWS) for natural disasters and urban vulnerabilities is playing a significant role in mitigation and minimising loss of life and damage to properties. Such systems require strong technical principles and sophisticated knowledge of the natural hazard, the urban context and risk management to enable dynamic and timely decision making against serious hazards. Landslides are a common form of natural hazard with global impacts. In particular, this is because landslides are closely linked with a variety of other natural hazards such as storms, earthquakes, flooding and volcanic eruptions. However, predicting individual landslide occurrence is problematic as it depends on many local factors and variables and anthropogenic inputs. Current Early Warning Systems (EWS) for landslides rely on scientific methods such as hyperlocal rainfall monitoring, slope stability models and sensor technologies such as the Internet of Things and Remote Sensing. Decision-makers analyse Earth Observation (EO) and ancillary data produced by sensors which are deployed in landslide-prone areas for landslide monitoring and prediction. The emergence of social media (e.g. Facebook, Twitter and Instragram) has led to the possibility of people also contributing to landslide monitoring by reporting warning signs related to landslide. Social media data produced by people can be considered as *"human sensors"* but these need to be coupled to efficient event detection systems to detect potential landslide events from social media streams. Based on this, knowledge about the interaction between warning signs from human sensors and landslides has to be realised to provide more accurate event detection and decision-making. Moreover, these events detected from social media need to be verified by decision-makers by analyzing IoT sensors or other corroborating data from the area of interest. Heterogeneity of sensor types (e.g. rain gauges, soil moisture, piezometers, inclinometers) and a wide spectrum of performance and accuracy characteristics and thus this leads to challenges in data integration and analysis to verify landslide precursors identified from social media. Ontology is a logic modelling technology which can be used to capture such knowledge from a domain expert to provide machine-understandable and parsable relationships and inferences. Generally, EWS analyses both historical and real-time time-series data to understand the changing pattern of hazardous events and provide decision making support from various perspectives. However, such data provided by multiple data sources exhibit different characteristics and heterogeneous representations. These raise a new challenge for existing EWS platforms as regards integrating heterogeneous time-series data across multiple data sources to enable multi-dimensional queries in various context (e.g. hazardous event, data sources, spatial and temporal).

In this thesis, a comprehensive set of algorithmic techniques for managing wide varieties of time series data from heterogeneous data sources is investigated. A novel ontology, namely the Landslip Ontology, is proposed to provide the knowledge base that establishes the relationship between landslide hazard and EO and ancillary data sources to support data integration for EWS applications. Moreover, ontology-based data integration and analytics systems that include human in the loop of hazard information acquisition from social media is proposed to establish a deeper and more accurate situational awareness of hazard events. The Landslip Ontology and the system are key contributions of this thesis for the Landslip project, a Natural Environment Research Council (NERC) funded project which aims to reduce the impacts of landslide multi-hazards in India. Finally, the system is extended to enable an interaction between natural hazard EWS and electrical grid EWS to contribute to electrical grid network monitoring and support decision making for electrical grid infrastructure management.

1.2 Early Warning System for Urban Hazard Management

Rapid population growth in cities demands effective plans to protect people from vulnerabilities, for example, natural disasters. However, most disaster-prone cities are unprepared for future disasters and ill-equipped to reduce associated risks [2]. The urban settlement extension of the Predeal town, a famous skiing resort located in the highest altitude in the Romanian Carpathians (975 - 1060 meters), on to a landslide-prone area presents a potential hazard for future urban development [3]. A tool that integrated scientific expertise for landslide susceptibility within the urban development framework was developed for local authorities to support effective urban planning. RAPIDS [4] is an Early Warning System developed to manage urban flooding and water quality hazards. Here, Artificial Neural Networks are used for real-time prediction of flooding based on weather radar and rain gauge rainfall data. This research was conducted using case studies for the town of Keighley, West Yorkshire, UK, to demonstrate the proof-con-concept. Alerta-Rio [5] is an Early Warning System operated by Rio de Janeiro city in Brazil to deal with rainfall and landslide. This system's major process is to monitor landslide and rainfall using a correlation model and rainfall threshold to issue a warning.

According to the ISDR definition, the early warning is *"the provision of timely and effective information, through identified institutions, that allows individuals exposed to a hazard to take action to avoid or reduce their risk and prepare for effective response"* [6]. An Early Warning System (EWS) for urban hazard management can play a significant role in enabling dynamic and timely decision-making for risk management in cities. The EWS is the process of analyzing vast amounts of urban data to understand and holistically model city vulnerability. Due to the complexity of risk management for cities, this process requires sophisticated techniques such as data integration, pattern detection, and data mining to manage and process big data from different sources using both real-time and batch-processing models. Thus, it has become a grand challenge problem in the computing science domain to support efficient Early Warning System development.

1.3 Internet of Things (IoT) and Time-Series Data for Urban Hazard Management

The concept of the Internet of Things (IoT) technology was first presented by *Kevin Ashton* at Procter & Gamble (PG) in 1999 [7]. Keven Ashton defines the IoT as unique identifiable entities which are connected and interoperable with radio-frequency identification (RFID) technology. The mature development of IoT technology has the potential to enhance the effectiveness of disaster management. The paradigm of IoT enables the effective ability of data collection and sharing. The technology thus becomes an enabling technology to improve the effectiveness of applications in disaster management which include disaster prevention and mitigation, emergency response, and disaster recovery [8].

The Early Warning System (EWS) is one of the most significant parts of disaster prevention and mitigation in disaster management as it provides decision support to decision-makers in the disaster preparation and response [9]. An Early Warning System for the prediction and detection of natural hazard and disaster relies on scientific methods [10] and models which require input from the time-series data, such as the Earth Observation (EO) and ancillary data. The time-series data is a collection of continuous observation and measurement obtained at consecutive time points [11]. Additionally, the EO data are produced by a variety of remote sensing satellites and IoT sensors (e.g. Wireless Sensor Network (WSN), ground-based or in-situ sensors) which are deployed in the natural hazard-prone areas. Examples of EO Data are satellite image time series, precipitation, temperature, and humidity. The ancillary data include the natural hazards-related data collected from other data sources or methods (e.g. manually recorded by a human, data mining techniques, and scientific models). Examples of ancillary data are the total number of dead, missing, and injured people, damage to specific locations, and hazard-related events.

The time-series data have been widely used in natural hazard early warning in urban areas. Balai Litbang Sabo (BLS) has developed a Landslide Early Warning System (LEWS) that utilises different time interval of cumulative precipitation data (e.g. 1-Day and 3-Day) obtained from various data sources to observe and predict the occurrence of landslides in Indonesia [12]. A study in [13] analyses a large volume of EO and ancillary data obtained from various sources to conduct a landslide hazard assessment in the urban area around Langat River Basin, Selangor, Peninsular Malaysia. Such data include slope, elevation, drainage density, erosion, soil series, land use, 100-years flood data, precipitation, buildings, road, and essential facilities demographic. The RAPIDS [4] has been developed as an Early Warning System (EWS) that applies a data-driven based Artificial Neural Network to analyse weather radar and historical rain gauge rainfall data for real-time urban hazard prediction in the UK.

The utilisation of time-series data in such applications reveals the research challenges of time-series data management in EWSs for urban hazard management. Additionally, the management of time-series data for disaster management involves the integration of heterogeneities and wide varieties of data sources, data ingestion, and data fusion [14]. For example, this could be the integration of time-series satellite images provided by a variety of remote sensing satellites and the integration of weather data produced by the IoT sensors which are deployed in a natural hazard-prone area. Based on this, the discovery of data sources has thus become a significant part of time-series data integration due to the variety of disaster or hazard applications and a large number of data sources. The efficient time-series data sources discovery could suggest a sufficient number of data sources relevant to a particular disaster or hazard of interest. Hence, this thesis investigates the efficient discovery of data sources to support time-series data integration for urban hazard management.

1.4 The Role of Social Media in Urban Hazard Management

The emergence of social networks and crowd-sourcing has opened up new opportunities in several areas such as marketing, Customer Relationship Management (CRM), information and knowledge sharing, collaborative activities, organisation communications, education and training, and disaster management [15]. In the context of disaster management, social media enables the application of human-centric approaches that allow the public to provide essential disaster-related information that can be used to enhance the effectiveness of disaster management in reducing the impact of natural disasters. Social media users are *"human sensors"* [16] that observe and measure real-world phenomena and generate different types of social media data. The social media data produced by the human sensors are time-series data that contain rich information about human activities, environmental conditions, and public sentiment, which geographic information scientists, computer scientists and domain scientists can use for data analysis [17–19].

According to recent research, social media data is being used in disaster management for disseminating hazard-related and early warning information to the public [20, 21], establishing situational awareness [22–24], and supplemental information for decision-making [25, 26]. However, the vast volume and wide variety of generated social media data create an obstacle in preventing disaster management by limiting the availability of actionable information from social media. Several approaches have, therefore, been proposed in the literature to cope with the challenges of social media data for disaster management. Natural Language Processing and classification models are standard techniques for extracting hazard-related information (e.g. hazard events, hazard-warning signs, geo-location, time) [23, 24]. These techniques are unable to reveal the relationship between the information extracted from social media data where decision-makers can utilise the extracted information with their relationship for more accurate decision; for example, heavy rain and cracks are observed and posted by different social media users who are in the same area.

Although the techniques are mature for extracting information and identifying hazard events from the social media data, they are unable to reveal the relationship between the extracted information. This issue is mainly because such relationships are complicated and require domain knowledge to identify the relationships. Additionally, natural hazard events usually comprise of several properties, e.g. types, location, time, and have interconnection with various hazard-related events, e.g. hazard interactions, triggers and warning signs. The extracted information from social media and their relationships can be used to form a knowledge base and help decision-makers to understand the overall situation based on social media data and enable more accurate decision making. This process requires domain experts in natural hazards to define the relationships and a methodology to organise the extracted information and their relationships and represent them as a knowledge base that can be used to enhance the effectiveness of decision-making.

Besides, a single post from a public user typically does not provide all the necessary information, and some pieces of information related to a potential hazard might be missing. The knowledge base can be used to recognise the missing information and can be applied to obtain the missing pieces of information from social media.

This thesis investigates the development of a conceptual model that contains the knowledge of natural hazards captured from domain experts and use the developed model with social media data to construct a knowledge base which can be used to support decision making for natural hazards. Moreover, the developed model can be applied in an application to obtain missing pieces of information from social media.

1.5 Urban Risk Analytics Framework for Urban Hazard Management

The urban risk analytics framework is a conceptual architecture for a cloud-based urban hazard Early Warning system. Figure 1.1 illustrates the main components of the framework, which comprises comprehensive components that satisfy the requirements of general urban risk analytics.

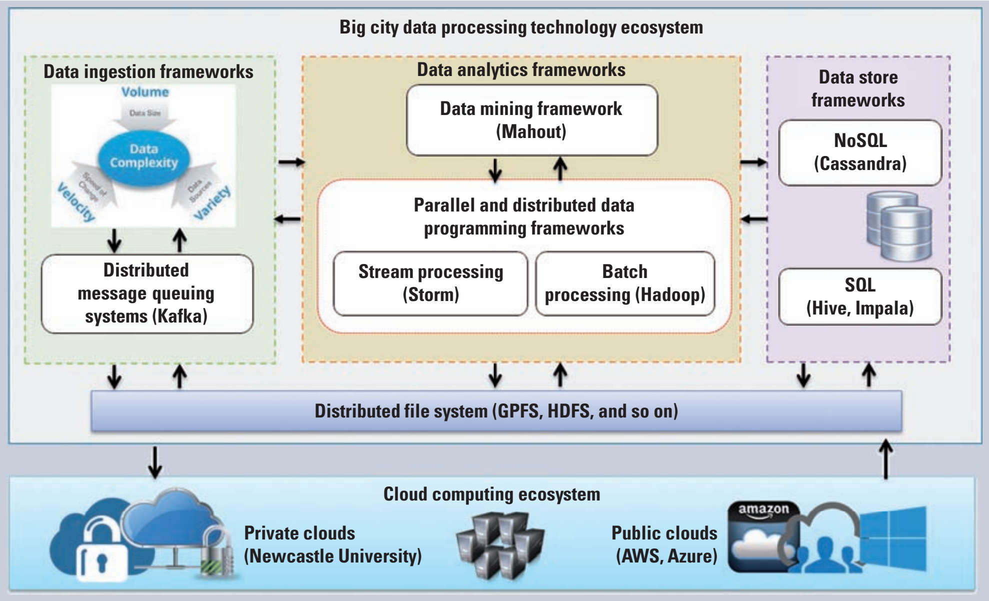

Figure 1.1: An urban risk analytics framework for processing heterogeneous time-series data in urban hazard management

1.5.1 Big City Data Processing Technology Ecosystem

This layer includes big data processing frameworks (BDPFs) that enable the creation of a big data application architecture. These frameworks can be classified as follows.

- **Distributed message queuing frameworks** — Such frameworks provide a reliable, high-throughput, and low-latency system of queuing real-time data streams from social media and other streaming sources. Examples include Amazon Kinesis and Apache Kafka.
- **Data mining frameworks** — These frameworks implement a wide range of data analysis algorithms for analyzing massive datasets, from natural language processing (NLP, including latent Dirichlet allocation, regression, or naïve Bayes) to computational statistics (Bayesian networks or state vector machines). Examples include FlexGP, ApacheMahout, MLBase, and Apache SAMOA.
- **Parallel and distributed data programming frameworks** — These frameworks, such as Apache Hadoop, Apache Spark, and Apache Storm, provide a distributed system implementation of big data programming models that includes stream processing and batch processing. Distributed system resource manage-

ment complexities such as task scheduling, data staging, fault management, interprocess communication, and result collection are automatically taken care of in Apache Hadoop and Apache Storm. The large-scale data mining frameworks mentioned previously are generally implemented on top of Hadoop, Spark, or Storm.

- **Datastore frameworks** — These include SQL and NoSQL database frameworks in which message queuing, data mining, and parallel or distributed data programming frameworks persist the intermediate and final data. NoSQL frameworks (such as MongoDB, HyperTable, Cassandra, or Amazon Dynamo) support data manipulation based on nonrelational primitives. Such nonrelational data manipulation patterns lead to better scalability and performance for unstructured data (for instance, social media postings or mobile app data). On the other hand, SQL data stores (MySQL, SQL Server, or PostgreSQL) are based on relational data manipulation primitives in which SQL can be used to manipulate data (insert, delete, or update). Urban risk analytics frameworks will use both NoSQL and SQL data stores, driven by data variety and querying needs.

1.5.2 Cloud Computing Ecosystem

This layer comprises hardware resources (CPU, storage, and networking) provided by private (the Natural Environment Research Council data centers, for example) and public (Amazon Web Services) cloud data centers. The hardware resources at this layer provide computational and storage capabilities to the big data processing frameworks. The end-to-end lifecycle operations (including selection, deployment, monitoring, and runtime control) of big data programming frameworks on cloud resources can be dynamically controlled via research orchestration frameworks [27].

Current big data analysis frameworks (such as Apache YARN or Mesos) do not need to meet the requirement raised by new classes of applications. That is no workflows, no dynamic indexing of existing and new data sources, no cloud-based implementation, and no dynamic tuning of the performance of big data processing frameworks to meet users' decision-making requirements. Applications such as urban risk analytics, however, require support for holistically processing data emitted by multiple sources.

1.6 Big Data Integration for Urban Hazard Management

A variety of urban hazard and risk management applications can lead to new research opportunities in urban risk analytics. The following research challenges, including semantic-based data integration, data classification, data indexing, trajectory data, and edge analytics, arise when developing big data integration and analytics algorithms for urban hazard risk management.

1.6.1 Semantic-based Data Integration of Multiple Data Sources

The integration of EO data and ancillary data from multiple sources is one of the most challenging problems for natural hazard early warning systems. There is always the possibility that data from different data sources are provided in the different data models (e.g. text, CSV, XML, JSON). Also, there is the possibility of data conflicts among various data sources. Examples of data conflicts are vocabulary, date format, data units, data precision, spatial and temporal scale. These conflicts reveal the challenges in the integration of wide varieties of data sources in semantically meaningful ways. Besides, multi-hazard applications require knowledge of the relationship among the data sources and hazardous events to answer complex questions and to support critical decision making. Based on this, Semantic Web technology has thus played a significant role to solve the problem for data integration of multiple data sources. Here, a number of ontologies have been proposed in the literature to conceptualise the knowledge of EO Data and hazards. SSN [28] is an ontology that describes the concepts of sensors, observations and related concepts. *Landslide* [29] is an ontology that describes the knowledge of landslide process, trigger events, and related hazards. SWEET [30] is a collection of multiple ontologies that represent concepts and relationships in the domain of earth and environment. Even though these ontologies provide comprehensive concepts for sensor data and hazard event, there is a lack of ability to represent concepts of human sensors (e.g. social media data). Supplemental processes are required when applying these ontologies to EWS for multi-hazard applications. Moreover, integrating data from wide varieties of data sources by using a single ontology approach is difficult due to the conflicts of data.

1.6.2 Data Classification

Datasets from multiple sources (e.g. social media, mobile apps, Instagram, and sensor networks) flow at different speeds and volume, and in heterogeneous formats (e.g. text streams from social media or mobile apps and numeric streams from landslide sensors). This diverse nature of data sources leads to heterogeneous requirements in terms of developing computer algorithms for data classification (e.g. NLP for text streams, and continuous numeric computation, including finding the max, min, average, and standard deviation, over streams from landslide sensors) and event detection (e.g. detecting the occurrence of keywords from social media streams and detecting flooding, landslide, or tsunami signals from real-time sensor streams).

Furthermore, based on data characteristics (static versus real-time), these computer algorithms will need to be implemented in multiple big data platforms that support heterogeneous programming abstractions. For example, historical datasets are in general handled by frameworks such as Apache Hadoop and Apache Mahout (a machine learning library for Hadoop), which offer map and reduce functions. On the other hand, computer algorithms for classification and event detection (also known as sliding window analytics) over real-time data will need to be implemented in stream processing frameworks such as Apache Storm and Yahoo S4. It is well understood that programming computer algorithms in these big data platforms that can handle multiple data sources and data formats simultaneously while ensuring data processing efficiency is a challenging research problem [31, 32].

1.6.3 Data Indexing

Developing an indexing algorithm that can seamlessly integrate and establish relationships among static and real-time data across multiple data sources in a multidimensional querying context (spatial, temporal, semantics, source types, event types, and so on) remains a very challenging problem [33]. Although it is relatively straightforward to design relational or nonrelational schema to store the raw or classified data for a single source type (such as social media or sensor feeds), establishing a relationship and dependencies among the sources in a multidimensional querying context remains an unsolved problem.

1.6.4 Trajectory Data

Dealing with the trajectories of dynamic data produced by multiple sources is also a challenge (for example, the trajectories of taxis and buses are sequences of GPS samples, whereas the trajectories of smartcard ticketing devices are sequences of bus or subway stations). Notably, these trajectories differ in terms of data velocity, volume, and location accuracy.

1.6.5 Edge Analytics

Latency-sensitive data analytics tasks (such as analyzing streaming data from sensors) can benefit from "edge analytics" techniques, which have benefits including:

- reduced network congestion achieved by filtering non-relevant events at the edge; and
- reduced event-detection latency (such as detecting dangerous water flow levels by analyzing real-time images in on-board processors available in sensor gateways such as Raspberry Pi 3), as sensors and gateways no longer need to send data to far-off cloud data centers.

However, it remains an open challenge how to enact and provision data analytics tasks across edge and cloud data centers so that decision-making latency is minimised while event-detection precision and accuracy is maximised.

1.7 Landslip Project

India is one of the countries most affected by natural disaster, including landslides [34, 35]. Approximately 0.42 million square kilometres or 12.6% of land area in India is susceptible to landslide hazards [36]. Landslip[1] is a Natural Environment Research Council (NERC) funded project which aims to reduce impacts of landslide multi-hazards in India, especially the hydrological related landslide multi-hazards. The Landslip consortium is formed of scientists and researchers from nine organisations which include the British Geological Survey (BGS), King's College London, MetOffice, Newcastle University, Practical Action Consulting UK (PA-UK), Amrita University, Geological Survey of India (GSI), Practical Action Consulting India (PA-India), and Consiglio Nazionale delle Ricerche (CNR). The scientific research in the Landslip project focuses on the study of weather patterns, landscape systems, rainfall thresholds, and societal factors to enhance landslide-related hazard assessment in India at regional scales (e.g. > 5 km) in two main case-study regions, Nilgiris District and Darjeeling/East Sikkim Districts. The understanding of the processes and methodologies developed for the case-study regions can contribute to early warning systems in the regions and other landslide-prone regions in South Asia. Based on this, the project comprises of seven interlinked work packages with different research areas (e.g. meteorological, landscape, social and multi-hazard).

The research conducted in this thesis contributes to the Landslip project as a part of Work Package 5 (WP5: social dynamics and vulnerability). The main aim of WP5 is to investigate the potential use of social media in landslide hazard Early Warning Systems. There are several tasks in this work package which include Ontology development, social media data classification and event detection algorithms, and mobile phone application. The key contributions of this thesis for WP5 of the Landslip project are as follows:

1. *The Landslip Ontology* — the Landslip Ontology proposed in this thesis plays a significant role in representing the knowledge of the landslide hazard domain captured from scientists and researchers who are expert in landslide multi-hazard management and members of the Landslip project. This knowledge presents concepts and relationships of landslide hazard, social media, and time-series data sources. The Landslip Ontology is used to facilitate social media-based event detection and time-series data integration.

[1] http://www.landslip.org

2. *Human in the Loop Social Media Analytics System* — the system collects and extracts social media streams to detect landslide-related events. The system includes a dialogue system that can actively communicate with social media users to obtain missing pieces of information and provide rich information for decision-makers for more accurate decision making.

1.8 A Scenario of Big Data Integration and Analysis in Urban Hazard Management

This scenario focuses on the application of big data integration and analysis for urban hazard management in a city that is vulnerable to disaster risk. The city is located in a mountainous area and near a river basin and is prone to natural hazards such as landslide and flood. Infrastructures and utilities such as roads, railways, electricity supply, and water supply have been provided in the city. This city drives the economy within the country and has a high density of population due to the rapid migration of people from other towns and the countryside. Consequently, the city and population settlements are expanding on to the area previously covered by landslides without adequate planning from the local authorities. The site is vulnerable to landslide disaster due to the inadequacy of structural mitigation works on the unstable slopes.

Urban data collection systems derived from IoT sensors are being deployed in the city. These systems, while currently delivering moderate data volumes, will soon ramp up to produce significant data volumes for real-time data analysis and data mining-based analysis of historical data. People in the city are already creating additional streams of potentially valuable data via social media (e.g. Twitter, Instagram, Facebook) and other crowd-sourcing mobile applications. As a result, the city is overflown many times a day by various orbital platforms that provide near real-time Earth observation (EO) and ancillary data covering various metrics, from urban temperature to atmospheric conditions. Moreover, the city infrastructure systems underpinning the movement of people, electricity supply, water supply, and waste management around the city also provide real-time information on the state or flows to support urban emergency management.

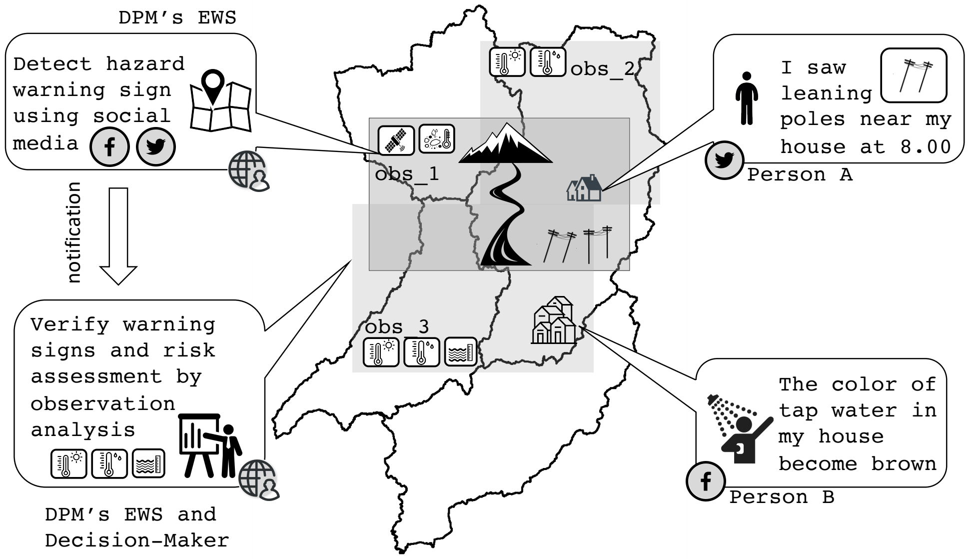

Figure 1.2: Urban hazard management scenario

The Department of Disaster Prevention and Mitigation (DPM) interested in the monitoring and forecasting of natural hazards in the city might wish to:

- analyse historical EO and ancillary data (slope, soil type and the impact of the flow of water in an area) to create a landslide susceptibility map and identify areas which are vulnerable to landslide and improve the urban expansion plan;
- analyse a historical archive of accumulated precipitation data to look at changing patterns of rainfall across the city;
- monitor the water level of the river in the city in real-time;
- analyse social media for relevant keywords or phrases to detect warning signs of landslides and floods;
- communicate with social media users who posted a message related to landslide and ask for more details; and
- develop real-time warning systems for landslides and floods.

Individually, each of these tasks requires a huge amount of data collection, data preparation, and subsequent analysis.

The city exhibits multiple levels of complexity across a large number of interacting domains (e.g. natural hazards, climate, transport, traffic control, electricity supply management) that operate on many temporal and spatial scales. This diversity of data and associated temporal and spatial variability has direct impacts on the ability to reliably and objectively monitor and characterise the environmental condition of cities. For example, although it might be possible to monitor slope movement in real-time using IoT sensors deployed in several locations in the city [37], it is unrealistic to undertake to do so in all areas in the city. However, images acquired by EO satellites and airborne remote sensing devices could let us extrapolate such measurements to the entire city. However, to gain maximum utility from such a diverse range of data, we require new integration approaches and associated analytics. This has been identified as a grand challenge problem in the computing science domain [31, 32].

1.9 The Research Challenges

This thesis mainly focuses on addressing the research challenges of big data integration and analysis for urban hazard management, especially the discovery of a wide variety of time-series data sources, and the organisation and construction of a knowledge base of hazard-related social media data. The complexity and the wide variety of time-series data sources with the absence of machine-understandable ontologies and knowledge bases are the major obstacles to data integration for urban hazard management. Data sources discovery is an essential part of the data integration for selecting potential data sources of EO and ancillary data efficiently. The data sources discovery associated with urban hazard knowledge is empowered to choose a sufficient number of data sources based on an understanding of the urban hazard context. For instance, when heavy rain has been detected in a particular landslide-prone area, the data sources discovery can suggest data sources providing EO and ancillary data for landslide monitoring in that area due to the knowledge that heavy rain is a warning sign for landslide hazards.

Besides, the variety and arbitrariness of hazard-related information extracted from the high volume of social media data are impediments to the organising and inferencing of the extracted information. Decision-makers are unable to understand the current situation of an urban hazard clearly or explore useful information to gain confidence in decision-making. Furthermore, well-organised information enables decision-makers to easily recognise where there are missing pieces of important information that should be further collected for more accurate decision making. This problem requires machine-understandable ontologies to facilitate the organisation and construction of a knowledge base of social media data for urban hazard management.

1.10 Research Questions

This thesis mainly investigates the problem of time-series data integration and analysis for urban hazard management. The landslide hazard domain is the primary focus of this thesis as it is closely linked to various other natural hazards with global impact. In particular, data sources discovery is the major obstacle to data integration in landslide hazards due to the complexity and wide variety of time-series data sources. Hence, the thesis focuses on the following research problems.

- How can we represent knowledge of the landslide domain that conceptualise the relationship between landslide hazard, social media, and wide varieties of time-series data to support landslide early warning and decision making?
- How can we utilise social media in context-based knowledge discovery to identify landslide events and discover potential time-series data sources to support time-series data integration for a landslide hazard Early Warning System?
- How can we perform landslide events information enrichment using social media to support early warning and decision making in landslide hazard management?

This thesis has addressed these research problems to enhance the efficiency of time-series data integration and analysis for urban hazard management. Specifically, the Landslip Ontology developed in this thesis has addressed the lack of a formal knowledge base of landslide domain concepts and relationships to facilitate time-series data source discovery. Furthermore, an ontology-based system for discovering landslide induced emergencies in the electrical grid has been developed to utilise social media and a knowledge base to identify landslide events and time-series data sources to monitor electrical grid infrastructure. Finally, the scarcity of information provided by social media users for landslide early warning and decision making has been addressed by a social media analytics system for active hazard observation developed in this thesis.

1.11 Scope and Contributions

This thesis investigates a comprehensive set of techniques to address the challenges of big data integration and analysis for urban hazard management. The investigation is conducted at different levels, ranging from conceptual models to concrete applications. The scope of the thesis is comprised of four main stages, as shown in Figure 1.3.

- *Landslip Ontology* — This stage aims at developing the Landslip Ontology (LO) to represent the knowledge of the landslide domain and provide a knowledge base that establishes relationships between landslide hazard, social media, and time-series data sources. The *LO* plays a significant role in this thesis to address the challenges in a wide variety of big data integration and analysis for urban hazard management. In addition, it is designed to facilitate time-series data sources discovery and social media event detection in the upper stage. The knowledge sources for the Landslip Ontology development are based on knowledge and experiences from scientists and experts who are members of the Landslip project.
- *Data Integration and Analysis Architecture* — This stage focuses on designing a common architecture that integrates essential components for ontology-based data integration and analysis system. The system is driven by the *LO* for data integration and analysis and is used to deploy concrete applications such as EWSs and Decision support tools.
- *Data Integration and Analysis Techniques* — This stage investigates a set of comprehensive techniques for two major components of data integration and analysis for urban hazard management, time-series data sources discovery and social media event detection. The time-series data sources discovery addresses the integration challenge of a wide variety of time-series data sources. The social

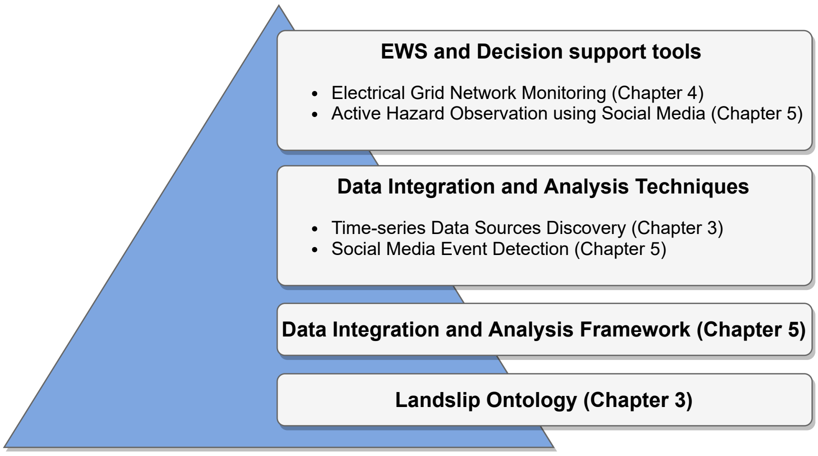

Figure 1.3: Scope of the thesis

media event detection integrates Machine Learning and Knowledge Representation and Reasoning techniques to provide context-based knowledge discovery of hazard events.

- *EWS and Decision Support Tools* — This stage provides concrete applications that utilise the techniques and designs derived from the previous stages to support the decision making in urban hazard management. There are two main applications present in this thesis, electrical grid network monitoring and active hazard observation using social media. The first application demonstrates an application of the *LO* and ontology-based data sources discovery technique in order to enhance the efficiency of electricity supply management in the areas that are prone to natural hazard. The latter application demonstrates a system that can observe natural hazards via the monitoring of the social media stream. Here, a Dialogue System is included in the system to provide an active way of natural hazard decision making by having a human in the loop of hazard information acquisition from social media. The proposed system addresses the challenge of social media information enrichment where missing hazard-related information can be obtained from specific social media users.

The major *contributions* of this thesis are as follows.

1. A formal knowledge base of landslide domain concepts to enable the integration of time series data from multiple and heterogeneous data sources for the early

prediction of landslide events, in Chapter 3.

2. An ontology-based data integration and analysis techniques to enable the discovery of time-series data sources and the detection of social media events for a landslide Early Warning System, in Chapter 4.

3. A process for harmonising the landslide knowledge base and electrical grid information services for monitoring of electricity grid network, in Chapter 4.

4. Data integration and analysis system for active hazard observation using social media, in Chapter 5.

1.12 Thesis Structure

This thesis is comprised of six chapters; the organisation of the thesis chapters is presented in Figure 1.4.

- **Chapter 1, Introduction** — This chapter explains the general background of big data integration analysis for urban hazard management. The background includes (i) the roles of Internet of Things technology, time-series data and social media data in urban hazard management; (ii) urban risk analytics framework; (iii) several topics on big data integration in hazard management; and (iv) the overview of the Landslip project. It reveals challenges and research questions for time-series data management, including IoT and social media data, in the context of hazard management. Novel techniques of big data and analysis for urban hazard management are introduced in this chapter, along with the thesis contributions.

- **Chapter 2, Literature review** — This chapter is a literature survey of the methodologies for social media data management and analysis for disaster management which is the key contribution of the thesis. A research taxonomy for the analysis and management of social media data is proposed to provide a systematic literature survey. This chapter also includes Research gaps analysis on big data integration in hazard management.

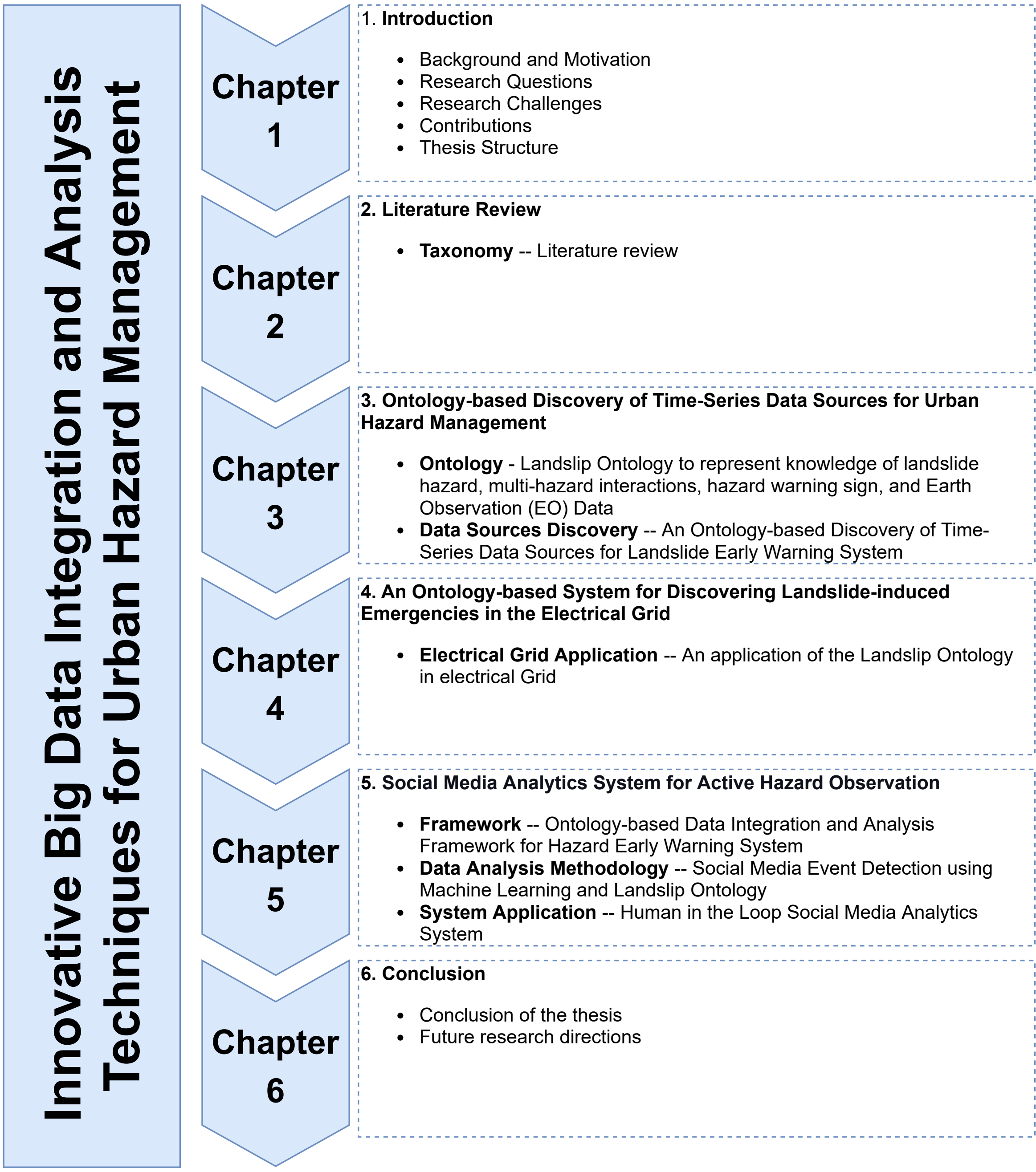

Figure 1.4: Summary of the thesis structure

- **Chapter 3, Ontology-based Discovery of Time-Series Data Sources for Urban Hazard Management** — This chapter presents a methodology for the integration of time-series data in the context of hazard management. The Landslip Ontology (*LO*), a novel ontology providing a formal knowledge base of landslide domain concepts, is proposed in this chapter to enable the integration of time-series data from multiple heterogeneous data sources. The purpose of the *LO* is to facilitate the discovery of time series data sources, an essential component of time-series data integration, for timely verification and prediction of landslide hazards. Ontology-based data sources discovery for landslide hazard Early Warning System is presented in this chapter to demonstrate the data integration methodology.

- **Chapter 4, An Ontology-based System for Discovering Landslide-induced Emergencies in the Electrical Grid** — This chapter presents an application of *LO* that harmonises the knowledge base of the landslide domain and electrical grid information services for monitoring of electrical grid network. The application demonstrates the utilisation of data integration across domains using the technique proposed in Chapter 4.

- **Chapter 5, Social Media Analytics System for Active Hazard Observation** — This chapter presents a big data integration and analysis system for active hazard observation using social media. It shows a concrete system as a prototype that utilises the *LO* for bi-directional interaction between social media users and the system to provide an active way of hazard information acquisition from social media that includes the human in the loop.

- **Chapter 6, Conclusion** — This chapter concludes the thesis, along with the discussion of future works.

2

Literature Review

This chapter provides a survey of how social media data contribute to disaster management and the methodologies for social media data management and analysis in disaster management. This survey includes the methodologies for social media data classification and event detection as well as spatial and temporal information extraction. Furthermore, a taxonomy of the research dimensions of social media data management and analysis for disaster management is also proposed, which is then applied in the survey of existing literature and discussion of the core advantages and disadvantages of the various methodologies.

2.1 Introduction

Disaster management has played a significant role in mitigating and minimising loss of life and damage to properties and infrastructure. Effective disaster management demands intelligent infrastructure for the collection, integration, management, and analysis of a variety of distributed data sources including ground-based sensors, video streaming, and satellite imagery [38]. The emergence of social network and crowdsourcing enables the application of human-centric approaches that allow the public to provide essential disaster-related information that can be used to enhance the effectiveness of disaster management in reducing the impact of natural disasters. Social media data contain rich information about human activities, environmental conditions, and public sentiment, which geographic information scientists, computer scientists and domain scientists can use for data analysis [17–19]. Social media not only generates massive volumes of data but also a wide variety of data types such as text, images, and videos. In 2020, there were 3.5 billion social media users worldwide, equivalent to about 45% of the world's population [39]. Facebook had over 2.6 billion monthly active users (MAUs) and 1.73 billion daily active users as of March 31 2020. Twitter had 330 million MAUs and 145 million daily active users [40] in 2019. 500 million tweets are sent by Twitter users every day, equivalent to 5,787 tweets per second [41]. Moreover, there were 1 billion MAUs on Instagram as of June 2018, and 500 million daily active users updated their stories. The large number of posts generated by these active users exemplifies the variety of dimensions of social media data. For instance, Twitter data comprises of several types of information, including account IDs, timestamps, user tweets (e.g. texts, images, videos), coordinates, retweets and so forth. The volume, velocity, and variety of data thus make it increasingly difficult for disaster managers to extract relevant and timely information from such data.

A comprehensive taxonomical framework is presented in this article to effectively explore, assess, contrast and compare existing approaches that use social media data for disaster management. Previous surveys in this subject area have mostly focused on specific aspects, hence, they appear to be narrow and fragmented. Topics include: (i) digital volunteerism [42]; (ii) disaster management lifecycle [43] including Warning, Impact, Response, and Relief; and (iii) disaster response [44] and [45]; considering aspects such as space, time, content, and network reach. Although the above-mentioned papers classified the literature based on taxonomies that covered aspects related to the effect of emergency occurrences on social media, social media data gathering and processing, and social media's effect on post disaster management, they are limited in scope with only a "broad" significance. In contrast, we present a holistic and comprehensive taxonomical framework. We propose a taxonomy that is much more exhaustive with additional (sub-)dimensions that contribute to an "in-depth" understanding of end-to-end challenges (e.g., data source, application, methodology, information dimension, and language) related to managing social media data for detecting, predicting, and responding to natural disasters. To date, this level of investigation has received little attention, and this article aims to alleviate this gap. While the existing surveys mainly discuss the novel techniques used to analyse social media data , their proposed classifications cannot represent the overall perspectives for applying social media data in disaster management. Overall, this survey proposes a novel taxonomy designed for understanding all significant aspects of social media data management and analysis for disaster management challenges, ranging from data sources to social media applications. Our taxonomical framework's main advantage is that it can provide the guidance of data management processes required in the context of using social media data for disaster management.

The focus of the work presented in this chapter is on understanding how social media data contribute to disaster management. We do this by surveying the literature for methodologies for social media data management and analysis for disaster management. We classify social media data based on their sources, language, information dimension, methodologies for data management, analysis, evaluation, and applications. The aim of this work is to provide a useful classification that could potentially be used to improve decision-making by enabling disaster managers to identify the appropriate data sources and the corresponding methodologies for analysis and management.

The main *contributions* of this chapter are as follows:

1. identifying the research challenges involved in using social media data for disaster management and the methodologies for data analysis and management

2. a research taxonomy for analysis and management of social media data

3. application of the proposed taxonomy to survey the existing literature of data analysis and management.

The rest of this chapter is structured as follows: motivation for this work is presented in Section 2.2 followed by the details of the survey method in Section 2.3. Classification details are presented in Section 2.4 including a categorisation of the data sources, language analysis, and identification of types of users. Languages presented in social media data that are used for social media data analysis are discussed in section 2.5. In Section 2.6, the approaches to how social media includes Spatial and Temporal Information are discussed. The methodologies for social media data management and the application of this data in disaster management are discussed in Sections 2.7 and 2.8 respectively before summarising the chapter in Section 2.9.

2.2 Motivation

During natural disasters, social media can play an essential role in the emergency response and provide a complete picture of situational awareness during and after the disaster. There are several challenges in acquiring and extracting hazard-related information from social media including volume, unstructured data sources, signal to noise ratio, ungrammatical and multilingual data, and fraudulent message identification and removal.

The massive amounts and variety of data generated by social media lead to a different level of information being extracted from the social media data. For instance, geographical information (geo-tagging) attached to a tweet about a roadblock on a hilly road provides more useful contextual information than a similar tweet without geo-tagging. Similarly, a tweet with attached images could potentially provide more situational awareness. For example, a tweet with photos of a roadblock in a hilly road can help people who are driving on the road nearby to understand the current situation of the roadblock and change to a new route away from the blocked area.

Due to the volume and complexity in such a large amount of social media data, it is crucial to have tools and systems that can automatically classify and extract information which could turn data into meaningful, actionable information to those attempting to manage the situation. This information has to be systematically managed and made available upon request and to be queried based on different query conditions. The main dimensions of query include geo-location/geo-fence, keywords and their disambiguations, user type (e.g. government, NGOs, news agencies, public etc.), and type of message (e.g. warning, news, SOS, request for supplies, general posts/tweets about an ongoing or impending situation). It is also important for the system to remove common false-positive patterns. For instance, the word “Landslide” in a tweet talking about a landslide victory of a sports team could potentially be classified as a tweet about a landslide hazard. To support this, the use of a tool such as an ontology can be applied to yield meaningful information from complex data. For instance, an ontology of landslides would represent the domain of landslide hazard through relevant terms and relationships between them. These relationships provide formal definitions to the domain terms thereby enabling machines to understand and analyse them. Thus, the knowledge represented in an ontology enables machines to perform intelligent tasks such as interactively communicating with social media users to extract contextual information related to an event of interest, identifying the relation of this information with the hazard of interest and providing this to the decision maker as a complete picture to enable informed decision making.

An ontology-based approach is thus more sophisticated than traditional data management approaches since it combines the data model with the associated domain knowledge that can be processed by machines to obtain semantically rich and meaningful information [46]. Systematic extraction of important information and semantic meaning of the free text in social media will help make the systems intelligent enough to organise and present data in an actionable form. Similarly, Natural Language Processing (NLP) is an important technology to understand and extract information from user-generated text content. We reviewed several natural language processing methods and case studies [47–49].

In this survey, the applications described above, including ontological support, NLP and data mining, are reviewed in the context of social media and natural hazard response and recovery.

2.3 Scope and Survey Procedure

In this section, we present a taxonomy of the existing research in social media data management and the procedure for selecting the publications discussed in this chapter.

2.3.1 Scope

Social media data provide a rich footprint of real-world events that can be used to facilitate the management of disasters. Several research works have proposed methods for exploiting social media data for the efficient management of disasters. The scope of our survey is determined by the common issues discussed in existing research works. Based on these issues, we have developed a taxonomy of social media data management and analysis for disaster management. Figure 2.1 depicts the taxonomy that shows six different aspects discussed in existing research including data source, language, social media user, information dimension, methodology, and application.

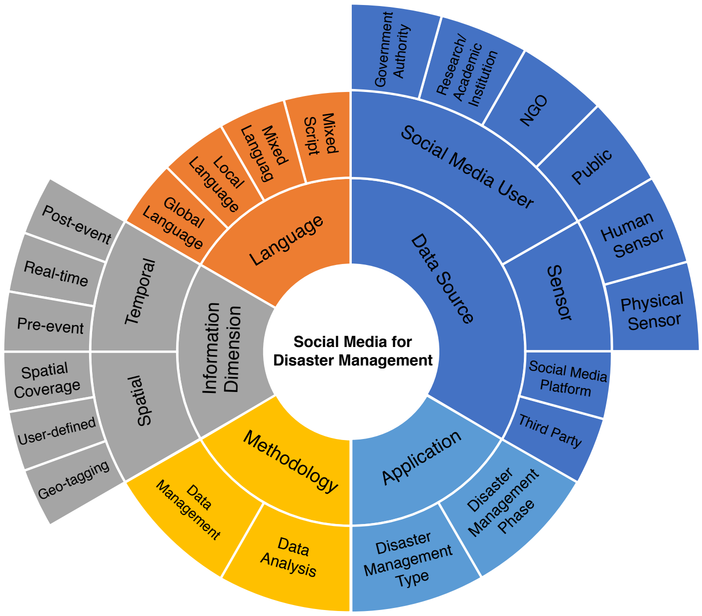

Figure 2.1: Taxonomy of Social Media Data Management

The taxonomy elements are described below.

- *Data Source* —refers to sources of social media data and ancillary data provided by other sources (e.g. physical sensors, WSN, and Web Services) that are used to facilitate social media data analysis in disaster management. We frame the dimension of data sources into four sub-classes, *Sensor*, *Social Media User*, *Social Media Platform*, and *Third Party*, based on the common attributes of data sources mentioned in the selected papers. The Sensor class is divided into *Physical Sensor*, *Human Sensor*, and the Social Media User is divided into four types of social media users including *Government Authorities*, *Research/Academic Institutions*, *Non-Governmental Organisations (NGOs)*, and *Public*.

- *Language* — refers to the language used for making a post on social media. Language is classified into *Global Language*, *Local Language*, *Mixed Language*, and *Mixed Script* types.

- *Information Dimension* — There are two major dimensions of information available in social media content, *Spatial* and *Temporal*. The methodologies for analysing social media content to extract spatial and temporal information are also included in this category. The spatial dimension refers to the representation of geographical information in social media and the methodologies for geo-location identification and analysis. The geographical information is present in several ways, including *Geo-tagging*, *User-defined*, and *Spatial Coverage*. The *Temporal* dimension refers to the utilisation of temporal information describing disaster-related events in the existing system for event detection. The temporal dimension is further classified as *Pre-event*, *Real-time*, and *Post-event* based on the temporal categories of real events.

- *Methodology* — refers to the methodologies and algorithms used to analyse social media data, especially the spatial and temporal information. We categorise the methodology based on data analysis stages that include *Methodology for Data Management* and *Methodology for Data Analysis*. The evaluations for each methodology are also summarised.

- *Application* — refers to the current uses of social media data for disaster management that are classified into two aspects of *Disaster Management Phases* and *Disaster Management Types*.

2.3.2 Research Gap Analysis

The recent emergence of cloud services, for instance, Microsoft Azure, Google App Engine, and Amazon Web Services, provides virtualised hardware resources and Big Data Processing Frameworks (BDPFs) to facilitate the development of Early Warning System for urban hazard management. However, state of the art in efficiently undertaking multi-source and multi-dimensional big data analytics for urban hazard domains is still relatively primitive. For example, BDPFs such as Apache Mahout and Apache SAMOA provide a platform for developing and executing classification and event detection algorithms (based on machine learning algorithms for NLP) over Apache Hadoop and Apache Storm, respectively. However, they do not guide how to define and model "events" relevant to a particular data source type or how to train the existing NLP algorithms to automatically detect and query [50] these events from the real-time and historical data. Moreover, BDPFs do not know the underlying machine learning algorithm and the overarching data analytics application. Hence, they are unable to adapt to the algorithm's performance based on application requirements and cloud resource availability. Furthermore, there is still a gap in the development of unsupervised machine-learning approaches that can help match a given data source to the best and most accurate machine-learning algorithm based on application-level goals [51] — for example, by maximising event detection accuracy and precision, minimising querying latency across multiple data sources. Furthermore, the Spark project at the University of California, Berkeley, released a new heterogeneous data querying engine called Spark SQL [52]. Spark SQL's DataFrame API is able to manage a distributed collection of data organised into named columns [52], which is similar to a traditional database. Multiple data sources from both external databases (JavaScript Object Notation, relational database management systems, or Apache Hive) and internal Spark data collections can be manipulated and processed through this API. Additionally, the mechanisms of multi-dimensional querying and ad hoc analysis are important to urban risk analysis frameworks. Integrating online analytical processing—a business intelligence technique—with DataFrame is one potential challenge for big data integration. Although Spark SQL can query multiple structured data sources, it cannot automatically integrate and resolve dependencies across those data sources in a multi-dimensional querying context, as noted.

Integrating and analysing heterogeneous sensor data from multiple data sources in an urban risk analytics framework is very hard due to the variety of data formats and sources. An effective urban risk analytics framework is driven by enabling technologies, which can range from IoT sensors technology to remote sensing technology. Moreover, with the high volume and extremely high rate of the data streams generated by heterogeneous sensors, ontology and Semantic Web technologies have emerged as one possible solution for integrating heterogeneous data. In other words, to develop an effective mechanism for urban risk data integration, there is a strong requirement to provide a formal description of the relationships among the variety of data sources.

Ontology engineering is a widely used technique in data integration, in which a knowledge base is captured from multiple sources such as domain experts, articles, processes, and knowledge is modelled using some standardised ontology language (for instance, the Web Ontology Language). Recently, several methodologies have been proposed for developing multisource data integration ontologies [53]. METHONTOLOGY [54] is one method that is widely used to develop ontologies in several domains. This method provides completed processes that cover the whole lifecycle of ontology development. Based on this, ontology engineering has become important in establishing a common understanding among experts from different areas that are working toward urban risk data analytics frameworks.

The Semantic Sensor Network Ontology (SSN) [1] [28] is a W3C standard for describing the concepts of sensors and observations. These concepts include sensor and sensor network modeling, measuring capabilities, sensor data, constraints, processes, deployments, and so on. SSN is widely used in sensor-based applications, including satellite imagery, scientific monitoring, and industrial infrastructure. SSN is a key ontology used for integrating varieties of sensor data and analysing disaster events. However, these comprehensive concepts do not cover descriptions related to specialised urban risks such as flooding, tsunamis, landslides, and so on.

2.3.3 Survey Procedure

The taxonomy presented in 2.3.1 is defined based on the common research issues mentioned in the selected publications in this survey. Our process for choosing the publications is divided into three main steps: (i) taxonomy and keyword determination; (ii) publication search; (iii) publication review; and (iv) publication selection.

- ***Step i. Taxonomy and keyword determination*** — We created a primary taxonomy and identified keywords based on the application of social media data in disaster management. The taxonomy and keywords were determined based on the requirements of our ongoing Landslip project[2]. The keywords were further used in step ii to search for publication candidates. Further, the set of keywords and taxonomy were iteratively refined throughout the process. We also used the

[1] www.w3.org/TR/vocab-ssn/

[2] http://www.landslip.org

keywords provided in the selected papers (in step iv.) to identify more keywords and analyse the critical issues in the selected papers to refine the taxonomy.

- ***Step ii. Publication Search*** — We searched for publication candidates from several potential repositories based on the set of keywords identified in step i. The main publication repositories and search engine used in this step included IEEE Xplore[3], ACM Digital Library[4], SpringerLink[5], ScienceDirect[6] and Google Scholar[7]. In addition, search results from Google Scholar directed to the sources of the publication candidates, which included ResearchGate, JMR, ScienceAdvances, PLOS, MDPI, and Tandfonline. The keywords can be classified into three main classes: *Social Media* (e.g. Social Media, Social Network, Crowd-sourcing, Twitter, Microblogs), *Disaster* (e.g. Disaster Management, Emergency Management, Landslide, Earthquake, Flood, Rainfall), and *Data Management and Analysis* (e.g. Data Analysis, Data Management, Data Mining). The search for publications was based on the combinations of keywords from these classes, which helped to narrow down the search leading to more focused and relevant results.

- ***Step iii. Publication Review*** — The search results of publications returned from repositories and search engines were reviewed and selected as candidates based on information provided in the title, keywords, and abstract of the publications and relevant to one of the scopes defined in the taxonomy (Section 2.3.1). As a result, *200 publications* from the repositories and search engines were selected as publication candidates.

- ***Step iv. Publications Selection*** — We scanned through the contents in the publication candidates and selected the publications for this survey based on their relevance to terms in the taxonomy. For example, [16] is one of the selected papers that provides information relevant to all terms defined in the taxonomy shown in Section 2.3.1. Accordingly, *40 publications* from several repositories

[3]https://ieeexplore.ieee.org
[4]https://dl.acm.org
[5]https://link.springer.com
[6]https://www.sciencedirect.com
[7]https://scholar.google.com

Table 2.1: Number of selected publications from repositories and search engines

Repository	Num of publications
IEEE Xplore	12
ACM Digital Library	9
ScienceDirect	7
SpringerLink	4
Others (via Google Scholar)	8

and search engines were selected. These are presented in Table 2.1. Next, we analysed the selected publications to extract more keywords and update the list of the keywords identified in step i. Furthermore, critical issues mentioned in the selected papers were used to refine and update terms and sub-terms in the taxonomy.

2.4 Data Source for Social Media Data Analysis

Effective disaster management demands high quality and rich data from many data sources that are related to the disaster of interest. Data sources could be any sensors and data services that provide data to a data consumer. This section presents four main data sources for social media data analysis classified in the taxonomy as Sensor, Social Media User, Social Media Platform, and Third Party. We analyse the characteristics of data sources for each aspect used for disaster management.

2.4.1 Sensor

The subclass Sensor includes data sources that produce original data for social media data analysis. Such data sources include *physical sensors* (e.g. remote sensing, in-situ sensor, wireless sensor network) and *human sensors* (e.g. social media, blogs and crowd sourcing). A physical sensor is a set of physical sensing devices that observe and measure physical phenomena and transform observation and measurement into a human-readable form. On the other hand, a social or human sensor comprises of human activities and interactions to observe real-world events and produce information in the social network[55].

- **Physical Sensor** — Earth Observation (EO) and ancillary data generated from physical sensors can enhance the effectiveness of social media data analysis for disaster management. There are several types of sensors generating EO and ancillary data. The in-situ sensor is a basic sensor that is deployed in the place of interest to observe and measure the physical phenomena directly. Examples of in-situ sensors include temperature sensors, rain gauges, and soil moisture sensors. In-situ sensing is suitable for analytics that requires high accuracy observation. Furthermore, Wireless Sensor Network (WSN) [56] is an advanced in-situ sensor system that consists of spatially distributed sensors, called nodes. Each node is usually equipped with wireless connectivity, a microcontroller, a power source and multi-type sensors. Based on this, data observed by each node can be exchanged among nodes within the system. With the computing capability, WSN can be applied in many applications, including industrial process monitoring and control, machine health monitoring, natural hazards and fire detection. Even though the in-situ sensing method can provide highly accurate data, the deployment of in-situ sensors to cover a wide area is difficult and extremely expensive. Remote sensing technologies (e.g., radar, satellite and airborne) have thus been used to remotely sense physical and environmental conditions and generate observation data that cover a wide area.

- **Human Sensor** — Human sensors utilise people to observe and measure real-world phenomena and generate different types of observation data including social media data. The emergence of social networks and mobile applications has enabled people to report about observed events. These activities are considered as human sensing [16] and can be a significant data source for effective urban risk analytics. The data sources include RSS feed, social media, Instagram, Twitter, Facebook, SMS, and online news. Similarly, crowdsourcing is a process that encourages people to give their contributions with regard to certain tasks in a specific context. This process is widely used in disaster management applications where people can report a disaster event they observed. For example, in 2010, people used Ushahidi, a web-based and mobile crowdsourcing application, to report about the earthquakes in Haiti [57].

The integration of physical sensor and human sensor data sources can enhance early warning and decision making in hazard management. In addition, social media data analysis utilises data produced by the human sensor to detect warning signs and other hazard-related events. However, the warning sign observed by social media users who are unknowledgeable about hazards requires additional processes to verify the potential of hazards. Based on this, Earth Observation (EO) and ancillary data generated from physical sensors play an important role in the verification and prediction processes. Spatial (e.g. geo-locations) and temporal (e.g. observation times) attributes are the most common attributes of the physical sensor and human sensor data sources to select potential physical data sources for the verification.

2.4.2 Social Media User

Messages originating from different accounts in social media have different quality and trustworthiness [58]. For instance, official accounts used by government agencies are likely to have more trustworthiness than public users with personal accounts. However, although government agencies that are responsible for the management of disasters use social media to disseminate disaster-related information, they still play a limited role in the communities. Instead, it is the public users that play a significant role in contributing to information networks during disaster events. Paper [59] shows different distributions of Twitter users participating in various disaster events with public users having a clearly greater percentage of participation. To summarise, different types of social media users play different roles in disaster management, each providing different context, quality and trustworthiness of social media data. In this chapter, we classify types of social media users as government authorities, research/academic institutions, Non-Governmental Organisations (NGO), and the public.

- **Government Authority** — refers to government organisations involved in disaster response and support. These organisations are authorised to: (i) disseminate official announcement and actionable warning information to people in a disaster risk area e.g., National Disaster Management Authority (NDMA) and (ii) provide supporting information for disaster management e.g., Geological Sur-

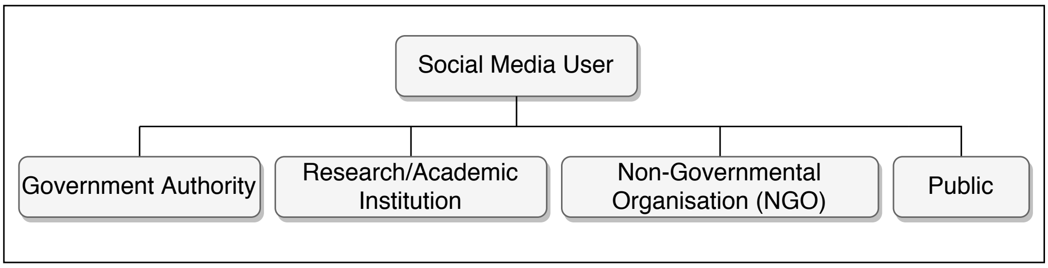

Figure 2.2: Social Media User

vey of India (GSI), British Geological Survey (BGS) and national Meteorological Offices.

- **Research/Academic Institution** — refers to institutions or research groups who are conducting research on disaster management.

- **Non-Governmental Organisation (NGO)** — refers to private-sector organisations that are disseminating disaster-related information on social media. This user type contributes a greater percentage of information than the government authorities and provides higher quality of information compared to the information provided by individual users. Examples of NGOs include Save the Hills, CNN, and ANI.

- **Public** — refers to individual users with personal social media accounts. This user type makes the most contribution to social media by sharing disaster-related information. With a huge number of users in this category, it constitutes the greatest percentage in information networks compared to other types of users. Most research [23, 60–64] relies on information contributed by public users even though the information may be of uncertain quality and trustworthiness. As a consequence, social media data preparation techniques (e.g. data filtering, data classification, and data extraction) to improve data quality and enhance the accuracy of social media data analysis proves challenging.

2.4.3 Social Media Platform

Generally, social media data are directly accessible from social media platforms (e.g. Facebook, Twitter, and Instagram). These platforms are considered as major data sources for social media data analytics in disaster management. Most social media platforms provide HTTP-based APIs for data consumers to access their social media services (e.g. data service and analytics services). Data consumers can use their tools to communicate with the respective APIs to collect and store social media data for their purpose [65]. For example, Twitter provides search APIs which enable consumers to find historical or real-time data by using keywords or hashtags. Much research on using social media data for disaster management utilises such APIs to access social media data directly from the social media platform [23, 60–62]. Due to the unstructured characteristics of social media data and indeterminacy of originated sources, the quality and trustworthiness of the collected social media data become significant issues [58]. Based on this, additional processes (e.g. data filtering, data classification and data extraction) for data preparation are required. Due to privacy concerns, some of the social media platforms (e.g. Facebook and Twitter) have put several restrictions on data access.

2.4.4 Third Party

Social media data are also collected and organised by organisations and institutes for specific purposes. Due to the benefits of Open Data, some of them have been interested in opening their collected social media data for others [58]. These organisations are considered as alternative data sources for social media data. This section discusses the different methods of accessing social media data. Third parties who provide their collected social media data are considered as alternative data sources for conducting research on social media data analytics for disaster management. This social media data is collected and organized in a specific way to be used for a specific purpose. For example, CrisisLexT26 [63] provides crisis-related tweets during emergency events which are collected from Twitter by using crisis-specific keywords. CrowdFlower [66] provides the Figure Eight platform for free open datasets. These include tweets relevant to various kinds of disasters. Most social media data collected by third-party data sources is usually prepared using additional processes to provide higher quality datasets. As well as using datasets from third-party data sources for disaster management, such datasets can be used as training datasets and evaluation for data analysis in many research on disaster management. For example, [64] utilises datasets from CrisisLexT26 and CrowdFlower as training datasets for identifying disaster-related tweets.

Table 2.2 depicts some existing works that use social media data sources for disaster management. Twitter is a major source of social media data which can be accessed via the Twitter API.

Table 2.2: Data Source of Social Media for Disaster Management

	Data Source		
References	Physical Sensor	Human Sensor	Access Method
[16]	USGS, NOAA	Flickr crowdsourcing	Web, Web Services
[57]	-	Ushahidi crowdsourcing	Mobile App, Web Services
[55]	-	Twitter	Direct access via Twitter APIs
[23]	-	Twitter	Direct Access via Twitter APIs
[60]	-	Twitter	Direct Access via Twitter APIs

2.5 Language

Language refers to the language used for making a post on social media. Languages of the social media data are investigated and classified into four categories: global language, local language, mixed language, and mixed script.

- **Global Language** — refers to social media posts that are in English.
- **Local Language** — refers to social media posts in languages other than English.
- **Mixed Language** — refers to social media posts that are a combination of two or more languages.

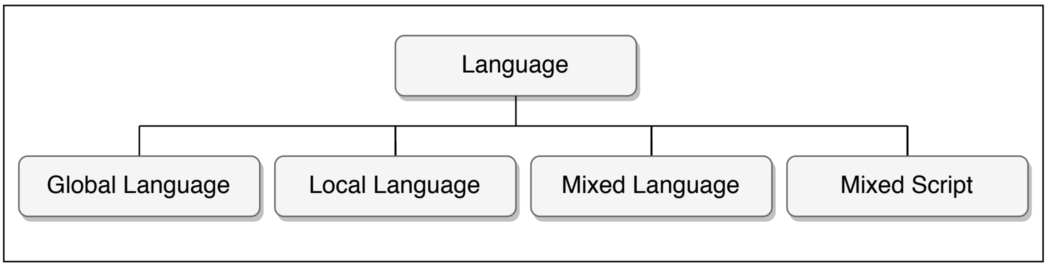

Figure 2.3: Language

- **Mixed Script** — refers to social media posts that are in a stylistic or linguistic variation of two or more languages. For example, a Twitter user may tweet in Hindi language using English script.

Table 2.3 illustrates the variety of languages presented in social media data. Social media posts in English are used in most research, and English is also a common language for social media posts in mixed language. Moreover, multiple local languages can be seen in the social media post in some research. Such variety has become challenging in the understanding of text based information produced by social media. Here, Natural Language Processing (NLP) has played an essential role in understanding and extracting useful information from the text information and facilitating disaster management. Research works in NLP are discussed in section 2.7

Table 2.3: Language Used in Social Media for Disaster Management

	Language			
References	Global	Local	Mixed Language	Mixed Script
[23]	✓	-	-	-
[60]	✓	Arabic	English, Arabic	-
[67]	✓	-	-	-
[68]	✓	Filipino	English, Filipino	-
[69]	✓	Hindi	English, Hindi	Hindi
[70]	✓	Spanish, German	-	-

2.6 Information Dimension

This section presents two major dimensions of information, spatial and temporal, which are essential parts of social media data for disaster management as we explain below.

We also investigate the variety of methodologies for analysing social media contents to extract the spatial and temporal information.

2.6.1 Spatial

Although spatial representation of social media data such as geo-location plays an essential part in social media-based event detection or event analysis, there is little social media data that provides information about users' location [71]. Furthermore, there is a variety of location information represented in social media ranging from a very precise location using geographic coordinates (e.g. *longitude* and *latitude*) to a very fuzzy location using descriptive language (e.g. city name).

Geographical information is represented in several ways on social media as shown in figure 2.4.

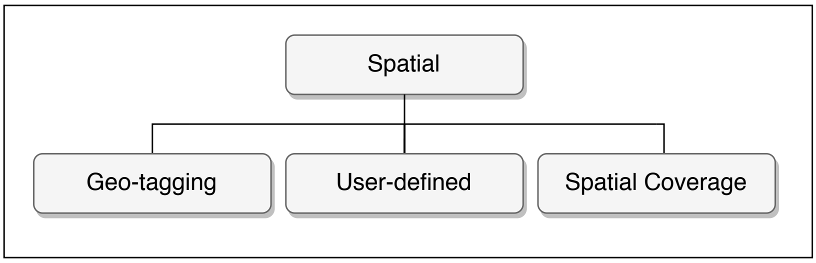

Figure 2.4: Spatial Representation of Geographical Information in Social Media

- **Geo-tagging**. The social media systems attach geographical information automatically or manually (by users) when the users post a message.
- **User-defined**.The user will mention the location in the post, either as the place name or as geographic coordinates.
- **Spatial coverage**. Many posts only mention the geographical extent such as the town/village/locality, or a district, or a country/province, or the continent or similar information.

Several approaches have been proposed in the literature to address the issues of spatial representation in social media. Here, we present the state-of-the-art methods to *identify* and *analyse* the spatial information of social media data.

Geo-location identification The research in [71] shows that 0.42% of all tweets have the latitude and longitude function to tag their geo-location, and out of 1 million Twitter users, only 26% have listed a city name. In most cases, general expressions (such as California), or nonsensical expressions (such as wonderland) are used. This research aimed to detect the location of the tweets that do not clearly mention geographic information. To this end, the authors proposed a method to compute the probability that a word links to a city. In order to improve accuracy, the authors introduce a model of spatial variation for analysing the geographic distribution of words in tweets. [72] mentioned that geo-location information in tweet data may have noisy signals. For example, a user in the UK tweets about a Houston Rocket's game or his vacation in India. To overcome this, the authors integrate two types of signals (user's friend and user's tweet's nearby location) from a social network to predict a user's location.

Geo-location analytics The authors in [73] discussed the effect of the earthquake on the East Coast of the United States (US) on August 23, 2011 by analysing the collected tweet data. The main finding of the paper is the patterns between the distance from the epicentre and the time after the earthquake. [74] used sensor data to identify the flood-affected regions. The authors performed some statistical analysis of the collected data to find the general spatial patterns and to explore the differences between the spatial patterns among the relevant tweets. On the other hand, methods such as Kernel density estimation (KDE) have been widely used in clustering the activities of Hurricane Sandy [75] and spatial hotspot detecting for the 2012 Beijing rainstorm [18].

Table 2.4: Spatial

References	Spatial			Pros and Cons
	Geo-tagging	User-defined	Spatial coverage	
[71]	-	Twitter user's city-level location estimation	Twitter user's city-level location	The proposed approach can improve the accuracy to identify the location of hazard events. The temporal aspect of the identified location needs to be further investigated for more accurate decision making.
[72]	-	profiling users' home locations	profiling users' home locations	The approach can be applied to disseminate disaster early warning messages to the social media users where their home location is in risk
[73]	spatial pattern analysis	spatial pattern analysis	-	The analysis of pattern between the distance of epicentre and time using social media can enhance situation awareness. However, the proposed method requires a sufficient amount of tweets to avoid data quality issues.
[74]	spatial pattern analysis	spatial pattern analysis	spatial pattern analysis	The approach to combinesocial media and sensor data for statistical analysis provides a more precise assessment of tweets spatial pattern in a hazardous area.
[75]	KDE-based clustering for social media analysis from Hurricane Sandy	KDE-based clustering for social media analysis from Hurricane Sand	KDE-based clustering for social media analysis from Hurricane Sandy	Associating social media data with hurricane damage data for spatial pattern analysis enable qualified assessment of rapid damage.
[18]	-	Hot-spots detection	Hot-spots detection	The proposed method for spatial analysis enables timely decision making to emergency response and full awareness of public concern.

2.6.2 Temporal

Most of the social media applications attach a time-stamp to the data posted. The temporal relation between events can be derived from the time-stamp of the event and the content. We studied how temporal information is used in the existing systems for event detection. In the context of event detection, we categorised the temporal information into three categories as follows:

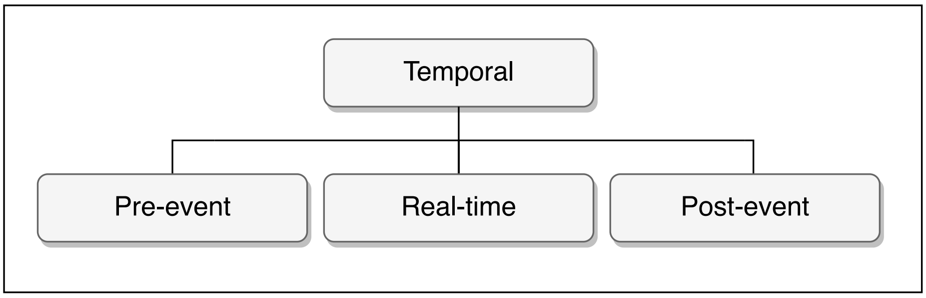

Figure 2.5: Temporal

- **Pre-event:** Represents the time period before the occurrence of the event of interest. In general, a social media message that is posted before the event occurrence can be analysed to derive the following information: (i) warnings - e.g. a post about bad weather from Met-Office before heavy rainfall or a cyclone alert etc. serves as a warning message for an impending natural disaster, (ii) precursor event detection - e.g. social media post about leaning electric pole in a location can serve as a precursor for landslide event detection, and (iii) temporal offset - pre-event posts from social media are analysed to determine the offset between the time of the post and time of the actual event such as, for instance, the time taken after the leaning pole post and the actual landslide in that locality. Pre-event posts from social media can thus be utilised for serving the mitigation and preparedness phases of emergency management.
- **Real-time:** Represents the time span during which the event is happening. In the real-time of the event occurrence, social media may be widely used for information sharing about the incidents related to the event. Generally, the real-time posts from social media, during the occurrence of the event, can be analysed for: (i) obtaining situational awareness - e.g. “trains cancelled, schools closed in

Kerala due to heavy rains"; a social media post about "roadblock due to landslides on NH-8", (ii) deriving/issuing warnings about the after-effects/impacts of a disaster - e.g. "high tides are expected in coastal areas after the tremors", and (iii) response, relief and recovery - e.g., a tweet during Kerala flood 2018: "shortage of bubble wrap and ready to eat items in Sanskrit College Palayam".

- **Post-event:** Represents the time period after the occurrence of the event of interest. Often after disasters, social media is widely used to communicate about supplies required, information about missing people, death toll, property loss, relief operations planned by the government and NGOs, protective measures to be undertaken while returning home, funds donated by various authorities, etc. Thus, the post-event data can generally be analysed for: (i) warning of further events, (ii) deriving information on the impact of the event, (iii) identifying the relief and recovery measures required, and (iv) determining the temporal offset between the time of the post and time of the actual event.

It is important to analyse the behaviour of the public/communities before, during and after disasters for bringing in effective disaster response, management, planning, and mitigation. Since social networks serve as the easiest and most common way for sampling public opinion, we can make use of the time-stamped, geo-tagged data from social media for this purpose. Chae et al. [76] explain about the temporal analysis of Twitter data related to hurricane Sandy wherein they analyse the Twitter user density distribution two weeks before and after the date of the event as well as for a time period on the day of the event, right after the announcement of the evacuation order. A similar study on the spatio-temporal analysis of Twitter data for the same disaster event was performed by Kryvasheyeu et al. [77], according to whom, the persistence of the Twitter activity levels in the time frame immediate to the occurrence of the event (post-event) is a good indicator to determine which areas are likely to need the most assistance. Further, during a disaster, normalised activity levels, rates of original content creation, and rates of content rebroadcast must be considered to identify the hardest-hit areas in real-time. In [78], the number of tweets during the Christchurch, New Zealand earthquakes were analysed over time for a window of five minutes. The analysis indicated that when a 4.2 magnitude or stronger earthquake occurred at a particular time, it correlated with the spike in the number of tweets over that time frame.

Another crucial factor to be considered while choosing the time frame for social media data collection is the type of disaster. For disasters like landslides, floods, and storms, we may be able to capture some of the warning signs for these events from the social posts before the actual occurrence of these events whereas for other events such as wildfires and earthquakes, the posts relevant or related to them may surface only after the occurrence of these events. Wang et al. [79] analysed wildfire-related tweets of some of the major wildfires that occurred in San Diego County, US, with respect to space, time, content, and network by collecting Twitter data from the day when the first wildfire occurred to the date when most of these wildfires were 100% contained. The temporal evolution of wildfire related tweets obtained using different keywords, with and without the location, gave an insight into the time-lag taken for the spreading of the information. Furthermore, as Granell and Ostermann mention in [80], the duration of the impact of these events also affects the temporal and contextual variation in the data related to these events. For instance, real-time and post-event data can be utilised for disaster response and recovery whereas pre-event data can be utilised for preparedness and planning. A case study to analyse the social media text during and after the 2012 Beijing Rainstorm is described in [18], where the authors performed time-series decomposition of the data to identify the overall trend and variations with respect to different development stages of the event as well as the cyclical trends of microblogging activity. They concluded that the trend analysis of text streams for different topics over time corresponded well to different development stages of the event. For example, texts related to the event increased in the week after the rainstorm following which, they began to subside slowly, and finally faded out.

The classification of these social media messages into different contextual categories and their analysis over time helps to identify the transition between various phases of disaster management and supports effective decision-making for disaster preparedness, response, and recovery. [81] presents a classifier based on logistic regression, which automatically classifies the gathered social media data into various topic categories during various disaster phases and the temporal trend of these topic categories in different phases. The experimentation using tweets related to hurricane Sandy revealed that: i) tweets regarding preparedness reached their peak on the day before the event when the emergency declaration was issued, ii) a large proportion of tweets related to impact are observed within a few days of the event occurrence, and iii) the largest peak of tweets related to disaster recovery was observed five days after the event.

Table 2.5: Temporal

References	Temporal			Pros and Cons
	Pre-event	Real-time	Post-event	
[76], [77]	Analysis of temporal distribution of tweets for Hurricane Sandy	Analysis of temporal distribution of tweets for Hurricane Sandy	Analysis of temporal distribution of tweets for Hurricane Sandy	The spatiotemporal visualization can bring a significant benefit for decision making. However, how to obtain spatial information from the miro-blogs is a big challenge which is not discussed in the papers.
[79]	-	Spatial and temporal analysis of wildfire Twitter activities using kernel density estimation (KDE)	-	The idea of using social media for wildfire hazards analysis is very interesting, but the analysis method and evaluation are too simple.
[18]	-	Trend analysis of 2012 Beijing Rainstorm using Time-series decomposition of social media data	-	This is the first study that utilises Chinese social media to uncover the emergency events in the Urban city. This paper introduces a useful data cleaning method of Chinese texts and can be used for purposes. However, the analysis of how to use the extracted spatial information is too simple.
[81]	-	Temporal trend analysis using logistic regression for tweets related to Hurricane Sandy	-	This paper makes a very detailed and complete categorizing for social media in disaster management.This expert knowledge can be reused by other researchers. However, the data analysis is too simple.
[78]	-	Extraction of situation awareness information using Burst detection technique and Online Clustering	-	This paper provides a solution to processing social media data in real-time. However, the evaluation is not sufficient to prove that the proposed method is practical and applicable to real-world applications.

2.7 Methodology

In the previous section, we discussed the state-of-the-art methodologies and algorithms used in research for extracting the spatial and temporal information from social media. However, methodologies and algorithms are used in all stages of social media data analytics. In this section, we categorise the most popular methodologies applied in social media data based on the data analysis stages.

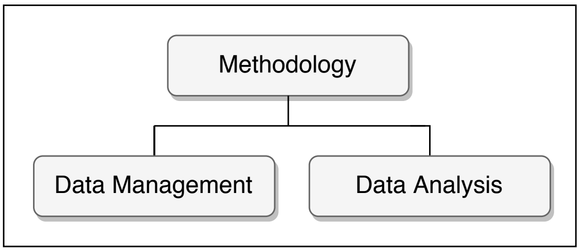

Figure 2.6: Methodology

2.7.1 Methodologies Used for Data Management

Data management for social media includes collecting, indexing, storing and querying of social media data for accessibility, reliability and timeliness of the data. Social media is generating a large volume of data everyday. For instance, according to [82], Facebook generates around four petabytes of data every day. The sheer amount of data itself poses a significant challenge in social media data management, making it a Big Data problem. Data management and analysis systems for social media data must therefore be able to handle the four Vs of Big Data analytics namely, volume, variety, velocity and veracity. In this section, we present the state-of-the-art in various systems and in research involving social media analytics for disaster management. From the data management perspective, we reviewed how data are collected, filtered, pre-processed, localised, stored, indexed and queried.

Maynard et al. [49] present a framework for real-time semantic social media analysis, which is based on the popular open-source framework for natural language processing GATE [83]. For the evaluation of the framework, they used the Twitter streaming API for data collection. Both streaming and batch processing approaches have been evaluated. GATE Cloud Paralleliser (GCP) [84] was used to perform batch processing of text, which supports execution of NLP processing pipelines with millions of documents. It performs the pre-processing and transformations required to load into the main information management system in the GATE pipeline, Mímir (Multi-paradigm Information Management Index and Repository). It also supports indexing of text, annotations and semantics. In real-time stream analysis, the Twitter client is used to capture data from the Twitter streaming API to feed into a message queue. Separate semantic analysis processors analyse and annotate the text and push into Mímir, which in turn enables semantic search using the knowledge encoded in knowledge graphs or ontologies[85]. This enables the indexed documents to form semantic relationships and thus making it easy to perform complex semantic searches over the indexed dataset. GATE Prospector [48] is used for exploring and searching datasets in the Mímir system. This system is reviewed in the next section.

Kim et al. propose a conceptual framework [86] for social media data collection and quality assessment. The framework's strategy consists of three major steps to develop, apply and validate search filters. Retrieval precision and retrieval recall are measured. Quality assessment of data collection is an important aspect of analysing a large amount of data such as social media contents. This is very much relevant in the disaster management scenario. Search filter development is performed with keyword selection, which includes disambiguations and slang words, and this procedure was generally performed manually by domain experts. Search filters are developed using standard logical operators like AND, OR, NOT and by involving data pre-processing techniques such as n-gram analysis and proximity operator. D-record [87] utilises three data sources: Twitter, OpenStreetMap and satellite images. A set of keywords for a needed concept was expanded using topic modelling learned using an SVM-based classifier with SMOTE. Goonetilleke et al. in ”Twitter Analytics: A big data management perspective” [47] reviewed several open source and commercial tools for data collection, management and querying for Twitter, many of which have been used in disaster management applications. Wang et al. 2013 and Wang et al. 2010 [88][89] developed a scalable CyberGIS for analysing large amounts of social media content in a natural disaster context. The system employs data fusion techniques to fuse social media data with census data and remote-sensing imagery. Slamet et al. proposed a system design [90] to find a Secure Place Locator (SPL) which covers a system information engineering aspect. It involves combining multiple data sources like location databases, governmental information and information from the community as it uses a relational database model to store and process the data.

Yates et al. performed a case study on emergency knowledge management and social media technologies [91]. The study investigated social media and related tools for effective knowledge management. It discusses how US Government agencies use social media data as an informal information dispersal mechanism and also studies how visual information layering helped the disaster management scenario.

Apart from text information, the use of multimedia data, such as images, audio, and video, in extreme event management [92] remains challenging due to the variety and complexity of the social media contents. Such events require sophisticated techniques to represent and analyse the multimedia contents to understand extreme events better. Research work in [93] proposes a novel data model based on a hypergraph structure to manage the massive amount of multimedia data produced by social media. The proposed data model comprises three different entities, users, multimedia objects, and annotation objects, to represent the variety and complexity relationships of the multimedia contents. This approach enables merging social media contents from different social media platforms in a single data structure. Here, the influence diffusion algorithm [94] has been proposed to investigate social media users who have significant interactions on a particular social media object.

Table 2.6: Methodology

References	Methodology: Data Management	Methodology: Data Analysis	Pros and Cons
[49] [95]	GATE Mimir is used for indexing and Querying. GATE Cloud Paralleliser is used for non-live processing. For real-time processing a message queue based system is used.	3 cases of tweet studies: 2015 UK election, EU referendum, and political reactions to climate change. Sematic Analytics Pipeline with NER (using TweetIE), NEL (using YODIE). Topic detection from tweets and sentiment analysis	GATE is a matured natural language processing toolkit. The proposed system is not domain-specific & the study is done using Twitter data only.Both data management and data analysis methodologies are GATE framework dependant
[87]	Consists of a pipeline of tools to geotag, classify text and tag “Needs” from tweets. Elastic Search is used to store and query data. Data sources consist of Twitter, OpenStreetMap & Satellite Images	CrisisNLP & CrisisLexT26 used to train the classifier Location-specific D-Record ontology is used which consists of concepts ”Needs” & ”Availability”. SVM based text classifier	Geo-locating tweets using OpenStreetMap and location mentions. Matching locations with “Needs” to locations with supplies. The system is limited to Twitter data. Relatively smaller set of data.
[96]	Semantic analyses are used for extracting tweets related to earthquake from Twitter and detect events in real-time	Uses Support Vector Machine (SVM) for the tweet classification and probabilistic model for location estimation	An example of real-time earth quake event detection using semantic analysis and real-time twitter feed where a user functions as sensor. Precision and recall of the system based on a number of tweets are evaluated to choose the better performance. The model assumes a single instance of a target event.
[97]	A topic modelling technique ‘Latent Dirichlet Allocation’ is used to retrieve information from a set of social media messages.	Seasonal-Trend Decomposition based on locally-weighted regression (Loess) known as STL is used to identify abnormal events	The abnormal event detection is based on probabilistic topic extraction and time series decomposition. Interactive visual analytic system. Analytics and processing approach is not of real-time nature
[98]	Candidate retrieval algorithm for retrieving events, Scoring and ranking procedure for ranking documents of a particular event.	Event detection based on SVM classification. The proposed system is trying to address the event identification[99] problem. The system do clustering of documents by events.	Precision, Recall, and F-Measure for evaluation. The system makes use of multiple social media sources which is an advantage. The Event candidate retrieval requires historical analysis. The system uses traditional SQL database which is not scalable for large social media datasets

2.7.2 Methodologies Used for Data Analysis

In [96], researchers have shown an example of an earthquake event and early warning using a social approach. It is accomplished by integrating semantic analysis and real-time data from Twitter. They have made two primary assumptions that each Twitter user is a sensor, and each tweet is associated with a time and location. Semantic analysis is used for classifying tweets into positive and negative classes. Tweets related to earthquake events are classified as a positive class, while tweets unrelated to earthquake events are classified as a negative class. Furthermore, they use the machine learning algorithm Support Vector Machine (SVM) for tweet classification.

On the other hand, Latent Dirichlet Allocation (LDA), a topic modelling technique in the information retrieval domain, is used in [97]. LDA is used to extract the inherent topic structure from a set of social media messages, and the extracted topic refers to an event (e.g. 2011 Virginia Earthquake).

The authors have given an example of topics and their proportion of each topic to all messages and showed how earthquake events, captured from the topics, constituted a small proportion of messages. Using the LDA topic model approach, meaningful topics are discovered with many iterations. Abnormal events, captured from extracted topics, do not happen frequently and cover only a small fraction of the social media data stream. In order to identify such abnormal events, the authors use Seasonal-Trend Decomposition based on locally-weighted regression (Loess) known as STL. In STL, the reminder component is used to implement control charts. The detected anomaly events are compared with other social media data to confirm the anomalies.

A candidate retrieval algorithm is used in [98] for retrieving events from the database. The authors implement feature extraction for extracting spatial, temporal and textual information and then used scoring and ranking to determine which document belongs to what event. SVM based classification is the methodology used in this paper for event detection.

An architecture for a public health surveillance process using SMART-C is presented in [100]. The architecture explains the data sources with their modalities, users, and services provided by the underlying system to enable enhanced situational awareness and informed decision making during all phases of disaster management. The authors discuss the requirements for implementing the following services: event classification/-grouping, semantic reasoning, location determination, event extraction, speech analysis, text analysis, video analysis, sensor analysis, geospatial analysis, response planning and generation, and alert dissemination service. They also present a discussion on security and privacy, event detection, and correlation.

Classification and information extraction from Twitter is carried out in [67]. The authors use free software parts of speech tagging for Twitter and Weka data mining tools. For classification purposes, they first broadly classify the tweets into personal, informative and other tweets. They further classify the informative tweets into: (i) caution and advice, (ii) damage, (iii) donations, (iv) people, and (v) other. They use a Naive Bayesian classifier for feature extraction and use unigram, bigrams, and Part-of-Speech (POS) tagging to provide a rich set of features in the classifier. Once a tweet is classified, a sequence labelling task identifies relevant information using conditional random fields.

A participatory sensing-based model is discussed for mining spatial information of urban emergency events in [101]. The researchers conduct simulations on the typhoon event, Typhoon Chan-hom. They propose a hierarchical data model with three different layers: (i) Social user layer, (ii) Crowdsourcing layer, and (iii) Spatial information layer. In the Social user layer, the proposed method collects data related to emergency events. In the Crowdsourcing layer, the positive samples are collected, and the address and GIS information are mined. Information related to the same emergency events is clustered in this layer. In the Spatial information layer, the spatial information of the emergency event is mined. Semantic analysis on the geo-tagged microblog data helped obtain a public opinion from the spatial perspective and assistance could be offered where it was required. From the collected data, it was observed that the risk was high in Beijing, Zhejiang, Jiangsu, and Shanghai.

In [71], the authors propose a probabilistic framework for identifying the location of a Twitter user based on the content of their tweet. The authors use a simple cart classifier to classify the tweets with strong geo-scope and then use a lattice-based neighbourhood smoothing model to refine user location. They also show that, with an increase in the number of tweets, the location estimation process converges.

The authors in [72] propose a unified discriminative influence model to solve the problem of profiling users' home locations on Twitter. They adapt probabilistic methods for local prediction and global prediction for profiling user location. Local prediction-based profiling uses the user's friends, followers and their tweets to profile the user's location efficiently, whereas global prediction, in addition, uses unlabelled users to accurately profile user location. In D-record [87], text sentences are vectorised to capture their semantics. Before featuring, the text is pre-processed by stemming, case folding and removing noisy lexical elements using SVM classifier with a lexicon-based feature, TF-IDF vectors and gensim's word2vec embedding.

2.8 Application

In this section, we investigate the contribution of a social media application in the context of disaster management strategy, which is a discipline to deal with disasters or avoid disasters where possible. In general, disaster management strategies consist of four phases: *Mitigation*, *Preparedness*, *Response*, and *Recovery* [102]. These four phases demand supporting tools and technologies for effective disaster management. Several recent research works have utilised social media data to address problems in different types of disasters and phases of disaster management. According to EM-DAT [103], there are two general groups of disasters: natural disasters and technological disasters with several types of disasters within each of these groups. Figure 2.7 depicts two significant dimensions of social media applications in disaster management studied in this chapter, *disaster management phase* and *disaster management type*. The disaster management phase represents the stage in life-cycle of disaster management contributed by social media applications whereas the disaster management type represents a disaster group of applications. Based on these dimensions, we investigate the current coverage of existing social media applications for disaster management and the overall picture of existing applications.

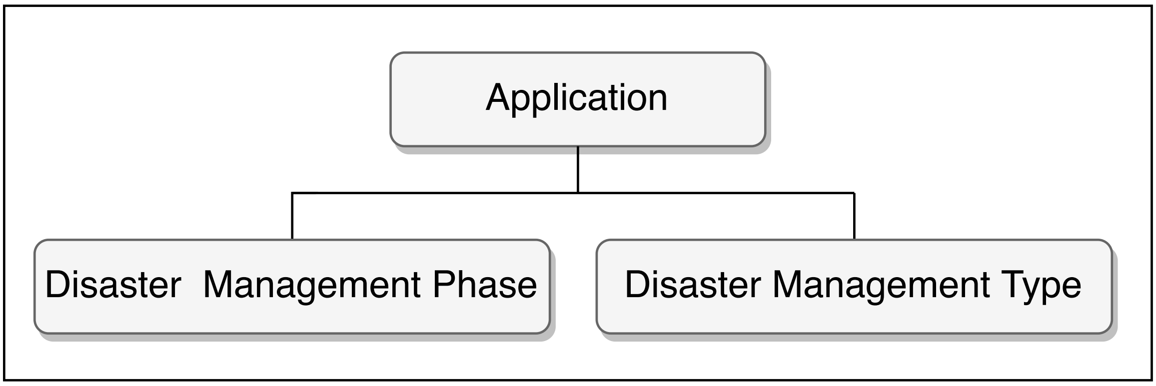

Figure 2.7: Dimensions of Social Media Applications in Disaster Management

2.8.1 Disaster Management Phases

Disaster management phases describe the normal life-cycle of a disaster and provide a useful framework for response [102]. As we have already established, there has been an increased use of social media in different phases of disaster management. Several techniques for the application of social media for disaster management have been proposed in the literature. As presented in Sections 2.6.1 and 2.6.2, social media data is usually generated with spatial and temporal information. Such information can be used to facilitate disaster management in different phases.

- **Mitigation** — the actions to minimise the cause and impact of hazards and prevent them from developing into a disaster.
- **Preparedness** — the action plans and educational activities for communities to confront unpreventable hazard events.
- **Response** — the actions to protect people's lives and properties during hazards or disaster events.
- **Recovery** — the actions to restore damaged properties and communities' infrastructures and to cure people of their illnesses.

2.8.2 Disaster Management Types

Disaster management types refer to groups of disasters, which are classified based on the root cause of the disaster. According to the International Disaster Database (EM-

Table 2.7: Application of Social Media Data in Disaster Management

Publications	Application	
	Disaster Management Phase	Disaster Management Type
[16]	Preparedness, Response	Technological, Natural
[76]	Preparedness, Response, Recovery	Natural disaster
[77]	Preparedness, Response, Recovery	Natural disaster
[18]	Response	Natural disaster
[81]	Preparedness, Response	Natural disaster

DAT) [103], there are two main groups of disasters, natural disasters and technological disasters.

- **Natural Disasters** — are natural events that emerge from natural processes or phenomena and may cause loss of people's lives and properties. Natural disasters are further divided into six sub-groups: Biological, Geophysical, Climatological, Hydrological, Meteorological and Extra-terrestrial disasters. Some examples of natural hazards are flood, landslide, earthquake, and tsunami.

- **Technological Disasters** — these are disasters that are a consequence of technological processes or human activities. Some examples of technological disasters are industrial and transport accidents.

Table 2.7 lists the application of social media data in disaster management. We identify the sub-class of both disaster management phase and disaster management type to the application of social media proposed by each publication. Consequently, it reveals the current coverage of the existing application of social media in disaster management. The research in [16] proposes a novel approach to view social media data as a human sensor and use social media to observe technological disasters (sightings of oil) and natural disasters (earthquake and air quality). Geo-location is extracted and used as a boundary for the prediction of an oil spill. Authors in [76] and [77] analysed Twitter data to identify public behaviour patterns from both, spatial and temporal perspectives during the natural disaster, Hurricane Sandy. An investigation of the emergency information distribution using social media during an emergency event is studied in

[18]. This work analyses the social media stream during the 2012 Beijing Rainstorm by using classification and location models. The authors in [81] analysed tweets about Hurricane Sandy to find temporal trends using a classifier based on logistic regression. The applications of social media mentioned in Table 2.7 are proposed to address problems in a different phase of disaster management. The work in [16, 76, 77, 81] addresses problems in the Preparedness phase while the outcomes in [16, 18, 76, 77, 81] are utilised for the Response phase. The applications outlined in [76, 77] are used for the Recovery phase.

It can be seen that most research has focused on the application of social media data for natural disasters rather than technological disasters. However, the approach presented in most of these works can be applied to multiple phases of disaster management. Interestingly, the Response phase is the most popular aspect to exploit social media data while there are no publications available to apply for the mitigation phase.

2.9 Summary

In this chapter, we reviewed research publications to investigate the contribution of social media data and the techniques for data management and analysis in disaster management. We studied the various dimensions of the contributions based on our proposed taxonomy that includes data sources, languages, spatial and temporal information, methodology, and applications. Human-centric approaches (e.g. social media, blogs, and crowdsourcing) have become a significant data source that provides the observation data of real-world events and contributes to disaster management. Several research publications have proposed exploiting social media data for disaster management with Twitter being one of the most significant social media data sources used for disaster management. The temporal and spatial information extracted from Twitter is critical information to support decision-making in disaster management. Geo-location identification and analysis are key research challenges of the spatial perspective in disaster management. Even though several methodologies have been proposed in the literature, these challenges remain unresolved. However, social media content, along with temporal information, including posting time, and event time, can be used to facilitate disaster management in several ways. Many research works used such information to detect precursor events or support decision making during disasters. Furthermore, several approaches for managing, analysing, and evaluating social media data have been proposed in the literature. It is evident that Big Data technology is a key technology for social media data management due to the high volume of generated social media data. Moreover, machine learning and information retrieval algorithms are widely used to collect, classify, and extract essential information from social media. Such information includes temporal and spatial information and disaster events. F-Measure, precision, and recall are common techniques for evaluation of the proposed methods for data collection, classification and extraction. Finally, the application perspective of this survey has shown evidence that social media plays a significant role in every phase of disaster management and the generated data has been extensively used in such management.

3

Ontology-based Discovery of Time-Series Data Sources for Urban Hazard Management

The research work in this chapter proposes a novel ontology, namely the Landslip Ontology, to provide the knowledge base that establishes the relationship between landslide hazard and EO and ancillary data sources. The Landslip Ontology (*LO*), a key contribution in this chapter, aims to facilitate time-series data source discovery to verify and predict landslide hazards. The *LO* is evaluated based on scenarios and competency questions to verify the coverage and consistency. Moreover, the *LO* can also be used to realise the implementation of a data sources discovery system which is an essential component in EWS that needs to manage (store, search, process) rich information from heterogeneous data sources.

3.1 Introduction

The analysis of big time series data has been a a grand challenge in several domains including health healthcare [104–107] and natural hazard management [6]. The advancement of Early Warning Systems (EWS) for natural hazards and urban vulnerabilities is playing a significant role in mitigation and minimising loss of life and damage to infrastructure. EWS systems require strong technical underpinning and sophisticated knowledge of the natural hazards such as the urban context and risk factors to enable dynamic and timely decision-making. Landslides, the main focus of this paper, are a common form of natural hazard that has global importance. Landslides are closely linked with a variety of other natural hazards such as storms, earthquakes, flooding and volcanic eruptions. The prediction of individual landslide occurrence is complex as it depends on many local factors, variables and anthropogenic (caused or produced by human beings) activities. Current EWS for landslides rely on scientific methods such as hyperlocal rainfall monitoring, slope stability models and analysis of remotely sensed images. With the emergence of Internet of Things (IoT), decision makers are also analysing observation and measurement data produced by sensors (e.g., soil moisture, soil movement, rainfall, humidity, wind speed) which are deployed in landslide prone areas.

Moreover, the emergence of social media (e.g. Facebook, Twitter and Instagram) has lead to the possibility for general public to also contribute to landslide monitoring by reporting warning signs related to landslide events. Before EWS can optimally utilise information from multiple, heterogeneous time series data sources (e.g., social media, IoT sensors), it is essential to realise a common knowledge base for capturing the core conceptual information and the cross co-relationship between events (that could be potentially discovered by analysing those data sources). Moreover, cross co-analysis of time series data sources is not only useful for the discovery of event co-relation but also allows for additional event verification. For example, landslide early warning sign detected by processing Twitter streams (e.g., by monitoring tweets relevant to landslides) can be verified by analyzing IoT sensor data or other corroborating data (e.g., news feed, remotely sensed satellite data) obtained from the area of interest. However, discovering such cross co-relationship of events from heterogeneous time series data sources has many challenges including lack of common terminology and presence of implicit relationships that are difficult to manually identify and analyse.

The main *contribution* of this chapter is a formal knowledge base of landslide domain concepts to enable the integration of time series data from multiple heterogeneous sources for real-time analysis and early prediction of landslide events. Underpinning this knowledge base is the Landslip ontology that captures the relationships between landslides, multi-hazards, warning signs, sensor data and other time series data sources. The purpose of the ontology is to facilitate data discovery which will be used to find potential data sources for landslide verification. The proposed Landslip ontology is evaluated based on scenarios and evidence from landslide hazard in Southern India (an area prone to landslide activity) [62]. The experimental results show the accuracy of the data discovery mechanism and indicate the benefits of using social media (along with other time series data sources) as a potential warning mechanism to bolster the potential of landslide early warning.

The rest of this chapter is organised as follow: related work is discussed next in Section 3.2, followed by a Landslip scenario in Section 3.3. The detail of Landslip ontology is described in Section 3.4, followed by the design of data sources discovery system in Section 3.5. The evaluation of Landslip ontology is discussed in Section 3.6. Finally, we summarise this chapter and future work in section 3.7.

3.2 Related Works

3.2.1 Data Utilisation in Multi-Hazard Early Warning System

Multi-hazard refers to a collection of multiple major hazards that a country faces [6] . There is a possibility that several hazardous events occur simultaneously and are interrelated. Tropical storms, for example, is one of the most common environmental hazards (in the tropics), which can trigger multiple hazards such as heavy rainfall that in in turn can induce flash flooding. Furthermore, heavy rain and flooding increase the moisture content of soil in a mountainous area and this may induce landslide. To minimise the loss of life and property damage from these inter-related hazards a comprehensive strategy for hazard management is required. In general, a strategy for hazard management is comprised of four phases [108]: (i) *mitigation* — the actions to minimise the cause and impact of hazards and prevent them from developing into full-blown disaster; (ii) *preparedness* — the action plans and educational activities for communities to confront with unpreventable hazard events; (iii) *response* — the actions for emergency situations to protect people life and properties during hazard or disaster events; and (iv) *recovery* — the actions to restore damaged properties and community's infrastructures and to cure people from their illnesses. These four phases demand supporting tools and technologies to enhance the effectiveness of hazard management.

Several modern multi-hazard early warning systems take advantage of the data explosion on the social media. Authors in [64] proposes using a twitter data analysis framework for identifying tweets that are relevant to a particular type of disaster (e.g. earthquake, flood, and wildfire). Several techniques, including matching-based and learning-based, to identify relevant tweets are also evaluated. The work in [109] studies the potential of using social media data to identify peatland fires and haze events in Sumatra Island, Indonesia. A data classification algorithm is used to analyse the tweets and the results are verified by using hotspot and air quality data from NASA satellite imagery. A data classification algorithm is also used in [57] to automatically classify tweets and text messages (from Ushahidi crowdsourcing application) generated during the Haiti earthquake in 2010. The goal of their work is to provide an information infrastructure for timely delivery of appropriately classified messages to the appropriate responsible departments. Work in [110] proposed a decision support system that integrates crowd sourcing information with Wireless Sensor Networks (WSN)to improve the coverage of monitoring area in flood risk management in Brazil. This research introduces the Open Geospatial Consortium (OGC) standards to facilitate the data integration among crowd sourcing information and WSN.

3.2.2 Semantic Web Technologies and High Variety Data Management for Multi-hazards

Earth Observation (EO) and urban data provided by multiple data sources are accessible by different methods ranging from direct download to various standard Web Services APIs (e.g. Web Map Services, Web Feature Services, Sensor Observation Services, RESTful API, SOAP-based API, etc.). In addition, there are heterogeneities among EO and urban data provided by different data sources [111] including: (i) syntactic heterogeneity — the difference in data format or data model for presenting datasets (e.g. plain text, CSV, Excel, XML, JSON, O&M, SensorML, etc.); (ii) structural heterogeneity — the difference in data schema for describing the same types of datasets (e.g. describing soil moisture using different XML Schemas); and (iii) semantic heterogeneity — difference in meaning or context of the content in datasets. These heterogeneities reveal the challenging problems brought forth by the high variety data availability in multi-hazard applications. Semantic Web Technologies have thus played a significant role by providing languages and tools for modelling domains including describing the concept and relationship among the data and hazardous events. According to W3C definition [112][113], the Semantic Web is a web of data that provides a common framework for data sharing and reuse across applications, enterprises, and communities.

Ontology, a key element of the Semantic Web, is a specification of a conceptual model for describing knowledge about a domain of interest. A basic concept in a form of ontology can be described by an Resource Description Framework (RDF) triple [114] which is comprised of a subject, a predicate and an object. Concepts described by RDF can be extended by Web Ontology Language (OWL) [115] to construct an ontology for representing rich and complex knowledge about things. In the case of multi-hazards application, an ontology can be used to: (i) represent domain knowledge through concepts, their attributes and relationships between data sources, data and hazards; and (ii) facilitate data integration across multiple data sources that represent varieties, velocity and volume characteristics of big data.

Ontologies are widely used in hazard management to model knowledge about hazards and use it to manage actual data derived from EO and urban sources. Hazard assessment and urbanisation analysis are two of the common application areas where ontologies are used. The Semantic Sensor Network Ontology (SSN) [28] and the Semantic Web for Earth and Environmental Terminology (SWEET) [30] are two significant ontologies that are commonly applied for hazard management. Authors in [30] reuse SWEET to conceptualise knowledge and expertise of several areas, such as buried assets (e.g. pipes and cables), soil, roads, the natural environment and human activities. Additionally, the Ontology of Soil Properties and Process (OSP) is proposed in their work to describe a concept of soil properties (e.g. soil strength) and process of soil (e.g. soil compaction). The OSP and other concepts are used to express how they affect each other in asset maintenance activities. Furthermore, [28] and [116] present the application of SSN for wind monitoring. The first work uses SSN with Ontology for Kinds and Units (QU) [117] to conceptualise wind properties (e.g. wind speed and direction) while the later uses SSN and SWEET to model the concepts of wind sensors and data streams of wind observations. The Landslides ontology [29] extends

SSN to organise knowledge for the landslides domain such as the concepts of landslides, earthquake, geographical units, soil, precipitation and wind. Even though these ontologies provide comprehensive concepts for sensor data and hazard event, and provide a reusable, widely used semantic underpinning, they do not cover conceptual aspects on human sensors (e.g. social media data). Hence, currently additional processes are required when applying these ontologies to EWS for multi-hazard application.

The related literature in the context of multi-hazard management can be classified based on the following three perspectives, *data sources*, *hazardous event analytics*, and *EO and urban time series data management*. It can be seen that effective multi-hazard management demands high quality and rich data from vast amount of data sources that are related to the hazard of interest. Data sources utilised by multi-hazard management applications can be any sensors and/or data services that provide EO and urban data. Such data sources include *physical sensor* (e.g. remote sensing, in-situ sensor, wire- less sensor network) and *human sensor* (e.g. social media, blogs and crowd sourcing). Recent data analytics research for multi-hazard management focused on hazardous event analysis, which are conducted into three main directions, *event identification*, *event verification*, and *event prediction*. These research reveal the challenging problems in the EO and urban time series data management, especially the discovery of potential time series data sources over the complexity and high variety of such data sources in multi-hazard management applications. Ontology is a common method for not only modeling knowledge about hazard but also managing EO and urban data. Recent work around developing the ontology in this domain are classified as *standardizing ontology* and *reusing ontology*. They have shown that current standard ontologies for data sources discovery do not exist. In addition, existing applications of ontology in this domain mostly investigate specific problems, in other words these approaches are not generalized. They fail to model the relationship between data sources and the domain knowledge which is an important factor for efficient data integration and data sources discovery.

3.3 Landslip Scenario

The development of Early Warning System and Decision Support System for multi-hazards can be accomplished in several approaches [118], depending on (i) the rules stakeholders engage in hazard risk reduction, (ii) geographical conditions of hazard prone area, and (iii) EO and urban data provided by responsible organizations. To achieve the goals of risk prevention and mitigation on landslide multi-hazards, scenario-based approach [119] is thus used in order to specify the scope of landslide problems and landslide hazard management activities. Additionally, the scenario-based approach is defined as a narrative story that represent expected uses of a system in the domain of interest from both domain experts and ontology developers viewpoints. Therefore, the scenario helps to identify the scope of the domain ontology to be designed.

3.3.1 Scenario

The Landslip scenario is co-created with domain experts who are members of Landslip. In addition, the Landslip is an NERC funded project involving the analysis of observation data and social media data provided by several institutions to contribute to the reduction of landslide impacts in the risk area of India. The scenario focuses on the *preparedness phase* of disaster management where *warning signs* of landslide from social media and observation data are detected before the occurrence of landslides. The landslide warning sign is an incident that indicates the potential of landslide hazard. It can be observed by human or an early warning system. The examples of landslide warning signs are the physical changes of utilities or infrastructure (e.g. blocked road, Leaning telephone poles, retaining walls or fences), a movement of soil from foundation, and a change of color in a river. In addition, a warning sign observed by people who are unknowledgeable about landslide hazard (e.g. social media users) requires additional process to verify the potential of landslide. Designing the scenario, historical event of landslide and domain experts experiences are considered. Figure 3.1 illustrates a situation before landslide that happens in a place located in an urban area. This area encompasses both natural environment (e.g. river, and mountain) and built environment (e.g. schools, hospitals, road, water supply and electricity). Locating in a high slope area, the place is prone to landslide and is monitoring by the National Disaster Management Authority (NDMA). Here, an expert from NDMA analyses satellite images to detect warning signs and informs decision makers for the potential landslide hazard. Meanwhile, *person A* who lives in the place is enjoying his leisure time by walking around his house. Living in the landslide prone area, it raises his awareness on possibility of landslide hazard. Then he is keen to contribute to his community by reporting any incident observed in his daily life via his social media account. While walking around his place, he has observed a leaning pole nearby his house. Thus, he takes a photo of the leaning pole and reports the incident to social network using his Twitter account. He also gives additional information such as observed time and place to the tweet. Besides, *person B* who lives in a nearby area has observed that the color of tab water in his house become brown. He thus uses hist Facebook account to report this incident. These messages from social media are collected by NDMA Early Warning System (EWS) to detect landslide warning signs. Here, an NDMA's staff who is a member of decision making team receives a notification message from EWS with regard to the leaning telephone pole. The incident is considered as a warning sign for landslide hazard. Receiving such information from social media users, the decision maker needs to verify the information before making further decision. Base on this, the decision maker who is a domain expert in landslide hazard risk assessment performs data analysis using an appropriate landslide analytical model. This process requires adequate historical events of landslide, EO and urban data provided by various data sources to get more accuracy of the data analysis. Hence, the decision maker searches for potential data sources from the Data Sources Discovery Service (DS) and gathers EO and urban data from the discovered data sources. The gathered data is then used to verify the social media information. Furthermore, the decision maker utilizes the EWS to assess the risk and impact of the landslide event using EO and urban data and take timely actions against the event. An example of such actions is disseminating actionable warning information to people in the landslide prone area.

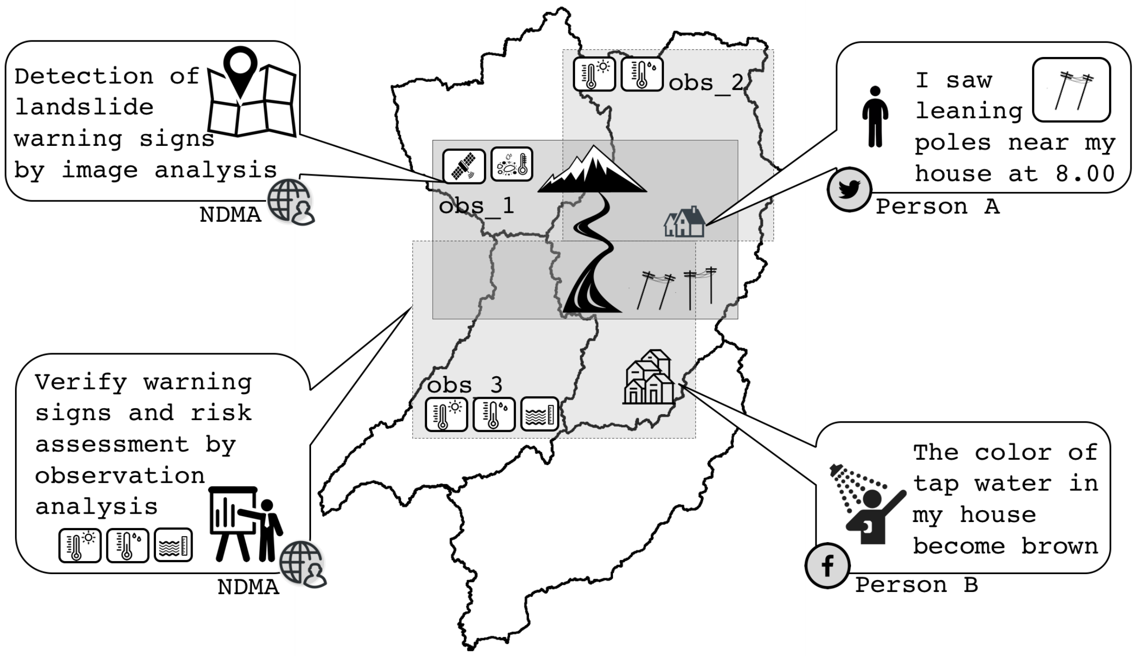

Figure 3.1: Overall of Landslip Ontology.

3.3.2 Overall Concepts

The above scenario reveals the essential role of data-driven early warning system for landslide hazard management which comprises of 5 main components.

- *Exposure* — refers to people and environment which are living or located in landslide hazard prone area and are affected by landslide multi-hazard. In addition, environment can be classified to natural environment and built environment. The natural environment is all living and non-living things that occurred naturally (e.g. animals, river, forest, mountain, etc.). On the other hand, the built environment [120] is a combination of infrastructures and facilities produced by people as a core foundation in the community (e.g. house, school, road, bridge, electricity, water supply, etc)

- *Stakeholder* — refers to people or organizations who have a stake in the landslide event. In the scenario, stakeholders are: (i) *social media users* who report landslide warning signs through their social media (e.g. Facebook, Twitter, Instragrams, etc.); (ii) *data collectors and providers* who deploy sensor devices in landslide hazard prone area and provide EO and urban data collecting from such devices to EWS for analysis. Data providers also include the third parties who collect data from sensor devices owned by the others. (iii) *Decision makers* who have responsible for conducting landslide hazard risk assessment using available social media data and EO and urban data. They make a decision based on result from Decision Support System and hazard risk management plan in order to inform people in risk area before the occurrence of landslide hazard.

- *Event* — refers to an occurrence which is related to a hazard. Additionally, hazard itself is also consider as an event. The hazard-related event is classified as pre-hazard event, post-hazard event and event during hazard. Since Early Warning System analyse EO and urban data to predict the potential of hazard in the area of interest, *warning signs* and *anthropogenic processes* are the majority of events in this scenario. In addition, a warning signs is an event that can indicate a possibility of hazards. An example of the warning sign is broken underground utilities which can be a warning sign for landslide hazard. An anthropogenic process refers to human activities which can induce hazards. An example of such activities is vegetation removal which induce landslide.

- *Data Sources* — refer to any sensors and data services that provide data to data consumers. These data sources have different capabilities to provide data. Sensor is a component that observes and measures physical phenomena and transform the observation and measurement into a human readable form. There are two types of sensor, *physical sensor* and *human sensor*. The data service is an application software that collects, stores and provides data from multiple devices. Several types of data sources are currently available to provide EO and urban data for multi-hazard applications.

- *Decision Support Applications* — refer to an integrated system that provide functionalities for stakeholders to monitor, forecast and predict, validate and assess hazardous events. In this scenario, EO and urban data collection system, data sources discovery services, hazardous event detection system and Early Warning System (EWS) are major components of Decision Support Applications. As a consequence, these applications enable stakeholders to take timely actions to reduce impacts of landslide hazard in advance. For example, once a landslide hazard is likely to be happened, a decision maker can make a decision based on information and knowledge from EWS to disseminate actionable warnings information to people in the landslide prone area.

The Data-driven early EWS analyses landslide-related data to enable dynamic and timely decision making against landslide hazard. Such data includes historical landslide events, historical and real-time observation data generated by physical sensors, and social media data. In addition, a number of sensor devices have been deployed in the landslide hazard prone area by organizations who are in charge of landslide hazard management. the organizations collect EO and ruban data from their sensors to monitor landslide hazard events in real-time. Besides, the collected data is stored in their local repositories and is provided as data sources to their co-ordinated organizations for further analysis. Here, the data sources metadata is published to a *Data Sources Discovery Services (DS)* which is an application of Decision Support System. The DS enables data publishers to advertise their data sources by registering data sources metadata to a metadata registry service. Moreover, It allows data consumers to search for their potential data sources to be used in their applications.

3.4 Landslip Ontology

Our proposed Landslip Ontology is developed based on NeOn [121] methodology. We define scope and purpose of the ontology based on the scenario mentioned in section 3.3. The ontology is implemented in OWL and is built in Protege. According to the scenario mentioned in the previous section, Landslip Ontology is designed and developed to conceptualize the knowledge of landslide hazard and its warning signs. Moreover, knowledge of data sources is also necessary to facilitate data sources discovery and landslide precursor verification. Based on this, Landslip ontology is comprised of two main modules, Landslip Common and Landslip-DataSources.

The main contribution of the Landslip Ontology to domain knowledge is to represent the knowledge of landslide domain experts and provide machine-understandable and parsable relationship and inferences. The Landslip Ontology utilises its landslide conceptual model and knowledge base to facilitate efficient data integration and data sources discovery. In addition, the relationship between landslides, other hazards and their interaction, social media, time-series data sources is used to discover a sufficient number of data sources based on an understanding of the land hazard context.

Scope and Purpose — The development of Landslip Ontology is driven by the goal of Landslip project to mitigate the impact of landslide hazard. Thus, the ontology focuses on the preparedness phase of disaster management where landslide warning signs play an important role to indicate the potential of landslide. The scope of Landslip Ontology is defined based on the scenario in section 3.3. Based on this, the ontology conceptualizes knowledges of landslide hazard specifically causes of landslide hazard and multi-hazards interactions which can trigger landslide hazard. Furthermore, the ontology conceptualizes landslide-related incidents which can be observed by people in landslide prone area. These incidents are considered as warning signs for landslide hazard. The concepts of landslide hazards are linked to EO and urban data which are set of properties for landslide observation. The ontology focuses on Landslide multi-hazard domain. The level of granularity is determined to the competency questions and terms identified.

Knowledge Sources — The ontology is design based on knowledge and experiences from four scientists and experts, from Landslip project, who are specialists in landslide hazard management with average 10 years experiences. Specifically, One scientist works for British Geological Survey (BGS) with focus on multi-hazard management. Other One is a scientist from Geological Survey of India (GSI) who are working on landslide hazard management in India. The two others are academic staffs who are specialist in natural hazard and geoscience.

Besides, publications [122], [123]and standard specifications [28, 30, 124–126]involving multi-hazards and geo-spatial data models are also used as additional knowledge sources to design the ontology.

The main goal of Landslip Ontology design is to capture necessary concepts and their relationships of the landslide hazard domain. To achieve the design goal, we have organised a workshop and interview to collect information for the design. A workshop session was organised in the 3rd Annual Partners Project Meeting for LANDSLIP to

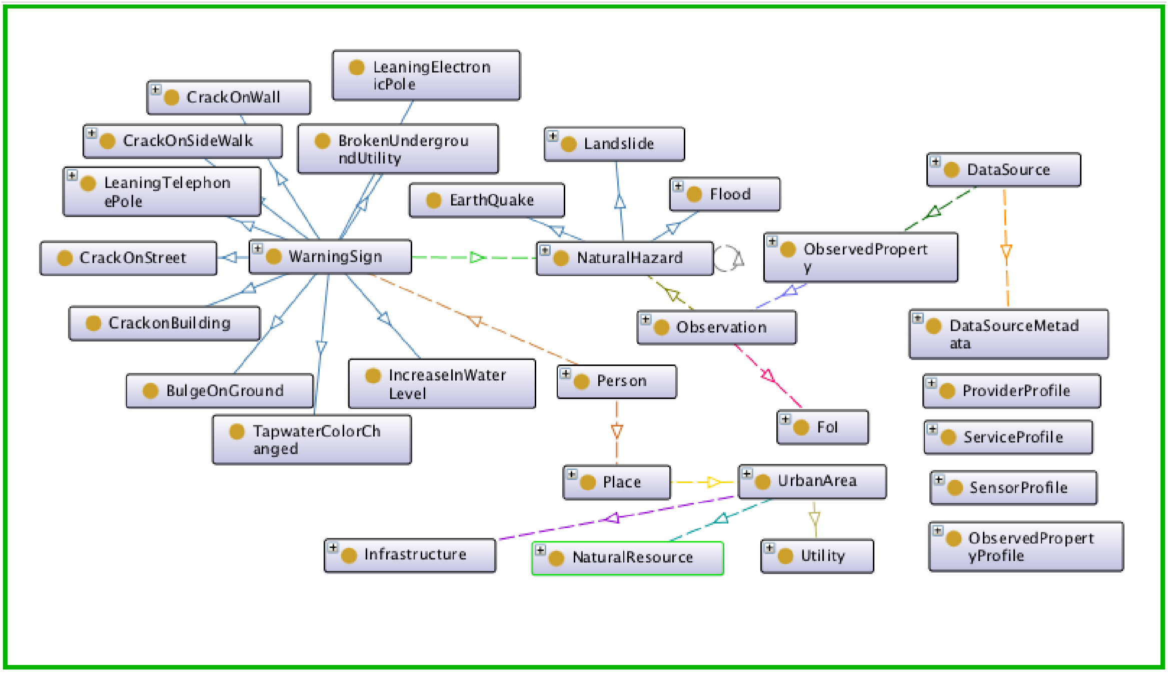

Figure 3.2: Overall of Landslip Ontology.

discuss how social media has been used in natural hazard EW systems. There were 35 domain experts in natural hazard management, including scientists, social scientists, local authorities, and NGOs. The domain experts are from several organisations, including Amrita University, BGS, CNR, KCL, UK MetOffice, Newcastle University, Keystone, PAC-India, GSI, Save the Hill, and Siligun College. The participants were divided into three groups to discuss "questions" they would like to know from social media users in different period of the disaster occurrence, pre-disaster, during a disaster, and post-disaster, respectively. Moreover, we organised an interview with landslide domain experts to discuss the questions obtained from the workshop. As a result, important questions have been selected by the domain experts to use in the design of Landslip Ontology. Furthermore, the concepts of data sources have to take into account for data sources discovery under the landslide hazard context. In addition, the design goal of data sources concepts is to provide metadata for accessing a wide variety and geographically distributed EO and ancillary data sources. Based on this, Landslip Ontology's overall design comprised major concepts, landslide domain concepts, and data sources concepts. These two major concepts are parts of the Landslip Ontology and can be used dependently for other purposes in the future.

Figure 3.2 depicts overall concepts of our proposed Landslip Ontology. The ontology is comprised of two modules: (i) *Landslip Common Ontology* – defines concepts about landslide hazard and its interaction to another hazards and anthropogenic process; and (ii) *Landslip Data Sources Ontology* – defines concepts about observation and data sources for landslide hazard risk assessment. The Landslip ontology reuses SSN ontology and terminology defined in OGC standards (e.g. Observation and Measurement [124], SensorML [125] and SOS [126]).

3.4.1 Landslip Common Ontology

The purpose of Landslip Common Ontology is to provide a conceptual knowledge model of landslide domain. In addition, the Common Ontology is a combination of theoretical knowledge and human experiences to identify warning sign before landslide. Basically, landslide is one of the most significant multi-hazards which can be found in many places around the globe [127]. Such hazard has interactions or can be triggered by another hazards [122]. Based on this, the Landslip Common Ontology conceptualizes knowledges of landslide and its interaction with other multi-hazards [122, 123] and knowledge of warning signs that can be observed by human and use such knowledge to indicate landslide event before the occurrence of landslide. The knowledge defined in the ontology can be used to facilitate landslide early warning based on warning signs observed and reported by people in social network. Figure 3.3 (a) illustrates the concept of the Landslip Common Ontology which comprises of four main concepts as follow:

- *UrbanArea* — defines concepts about urban area that prone to landslide including its basic elements. The urban area encompasses both natural resources (e.g. river, and mountain) and built environment, including infrastructure (e.g. road and railway), utility (e.g. electricity and tab water) and place (e.g. school, hospital, house and flat). Located in landslide prone area, these elements can be affected by landslide and other multi-hazards.
- *NaturalHazard* — defines a set of multi-hazards which can trigger landslide hazard. This concept captures knowledge mainly on the interactions between landslide hazard and other multi-hazards (e.g. flood, earthquake, tsunami, and

drought). In addition, the interactions between other multi-hazards are able to indicate landslide hazard.

- *AnthropogenicProcess* — defines a set of human activities that produce negative effects to landslide [122]. This concept also captures knowledge about the interaction with in the processes to provide direct and indirect indications of landslide hazards. In addition, the direct indications are the processes that are a trigger of landslide while the indirect indications are the processes that trigger other processes which trigger landslide. Moreover, the major indicators for anthropogenic processes are warning sign observed by a person.

- *WarningSign* — defines a set of incidents that can be an indications of landslide hazard, other multi-hazards and anthropogenic processes. The concept of warning sign is mainly focus on incidents which can be observed by a person. Such incidents are useful for landslide EWS in order to detect landslide precursors based on incidents reported in social network.

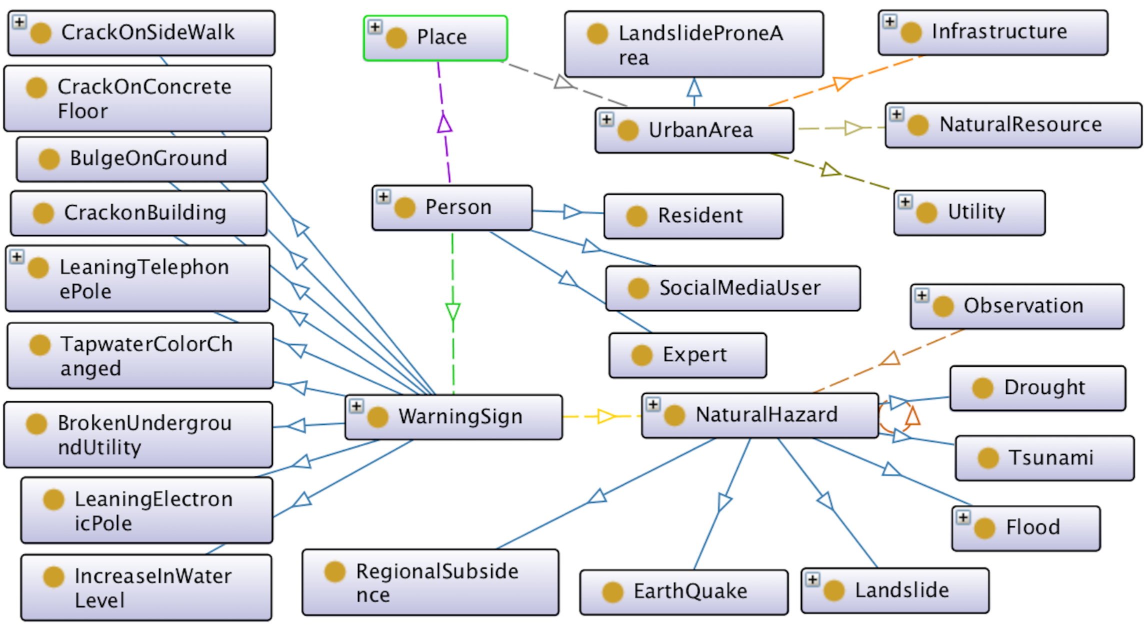

Figure 3.3: Common Ontology

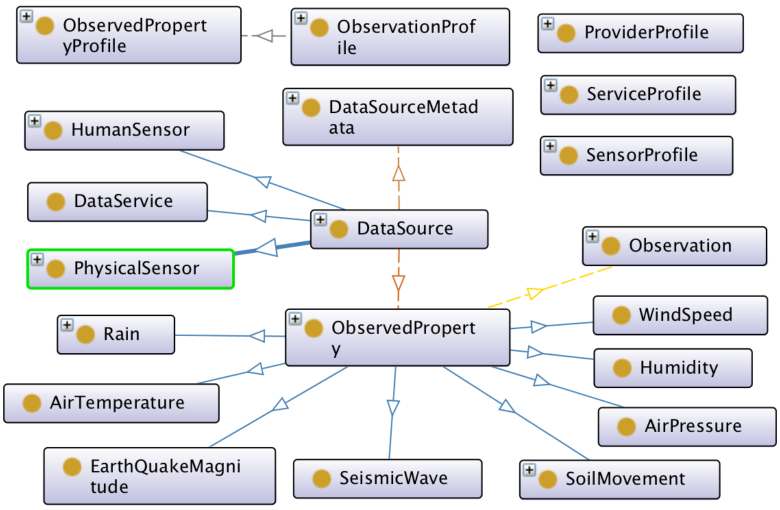

Figure 3.4: Data Source Ontology

3.4.2 Landslip Data Sources Ontology

EO and urban data observed by sensors indicate events or changes of landslide phenomena. Such data (e.g. rain, temperature, soil moisture. etc.) from a variety of data sources is collected by data provider and provides for stakeholders to be used in their landslide hazard applications [128]. Due to the high variety and geographically distributed nature of OE and urban data sources, effective data sources discovery [129] is thus required in order to provide sufficient amount and quality of data for landslide hazard risk assessment. Landslip Data Sources Ontology is developed to enable semantically discovery of data sources. In addition, the ontology describes concepts and relationships of EO and urban data, data sources, sensor devices, and data providers. With the combination of this ontology and Landslip Common Ontology, data sources discovery mechanism utilizes knowledge of landslide hazards to discover data sources which are related to the hazard of interest. Specifically, the knowledge of landslide warning sign can be used to identify appropriate observed properties and data sources for landslide precursor verifications. This capability enable EWS to provide dynamic and timely decision making against landslide hazards. Figure 3.4 (b) illustrates the Landslip Data Sources Ontology which comprises of three main concepts and reuse SSN Ontology [28] and OGC standard [124–126] for the concepts of observation and sensors.

Table 3.1: Landslip Ontology features

Feature	Value
No of classes	98
No of properties	26
No of individuals	30
No of axioms	462
DL expressivity	ALCH(D)

- *DataSource* — is the main concept of Landslip Data Sources Ontology. A data source is any sensors or data services that provide observation data (e.g. physical sensor, human sensor and data service). DataSource defines a set of comprehensive information related to observation and data sources metadata which are the details of data sources.
- *Observation* — defines a set of observed properties (EO and urban data) which are used to observed features of interest related to landslide hazard. The examples of observed properties are soil moisture, soil movement, rain, earth quake magnitude, temperature, humidity, and wind speed. These observed properties are accessible to EWS via data sources.
- *DataSourceMetadata* — defines a set of information which are necessary for data acquisition process. This concepts is comprised of four groups of information profile: (i) observation profile – a set of observed properties provided by a data source; (ii) observed property profile – provides information about data type, feature of interest, and phenomenon time; (iii) sensor profile – provides information about type of sensor, feature of interest, and list of event to be observed; (iv) service profile – provides information which can be used to access a service (e.g. service type, endpoint, provider); and (v) provider profile – provide the information about data provider (e.g. provider name, contact address).

3.4.3 Ontology Metrics

Table 3.1 shows a summary of the ontological features of Landslip Ontology in terms of size (number of classes, properties, and individuals), expressivity, and complexity of the core knowledge captured by axioms.

3.5 System Architecture

To realise the ontology-based data sources discovery system, we have designed the architecture which comprise of three main layers: (i) data sources layer; (ii) data discovery layer; and; and (iii) hazard applications layer. Figure 3.5, depicts the overview architecture of our proposed data sources discovery services.

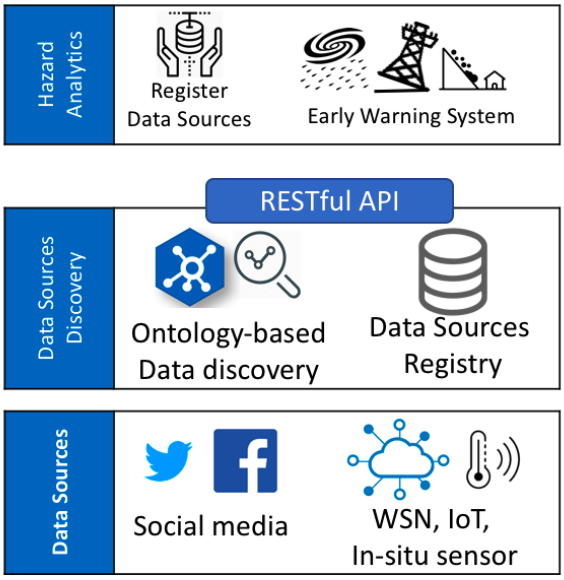

Figure 3.5: Landslip Data Sources Discovery Service Architecture.

- *data sources layer* — consist of a number of data sources provided by various data providers. Data sources collects EO and urban data from physical sensors deployed in landslide prone area. These sensors observe or measure properties of landslide and other earth observation which can be use to indicate landslde hazard. The data sources are accessible through a variety of methods (e.g. REST API, RDBMS, WSN) depending on data source providers. Moreover, data from social medias is also considered as data soures in this layer.

- *data sources discovery layer* — maintains the Lanslip ontology which represents knowledge of landslide and data sources in a triplestore. It also provides data sources registry which store data sources metadata, including metadata for sensor, service and observation. Furthermore, there are a number of functionalities provided by this layer which allow uers to (1) publish data sources; (2) search for potential data sources; and (3) indicate landslide hazard using warning signs. These functionalities are accomplished based on knowledge of landslide and data sources provided by the Landslip Ontology. In addition, the functionalities provided by this layer is accessible through data sources discovery service APIs which are available in form of RESTful Web Services API.

- *hazard applications layer* — privides client APIs to access the functionalities offered by the data sources discovery layers. In addition, the client APIs are design for both data provider and data consumer. Here, data provider can user the client API to register their data sources along with data sources metadata. On the other hand, data consumer uses the client API to search for potential data sources based on landslide warning sign.

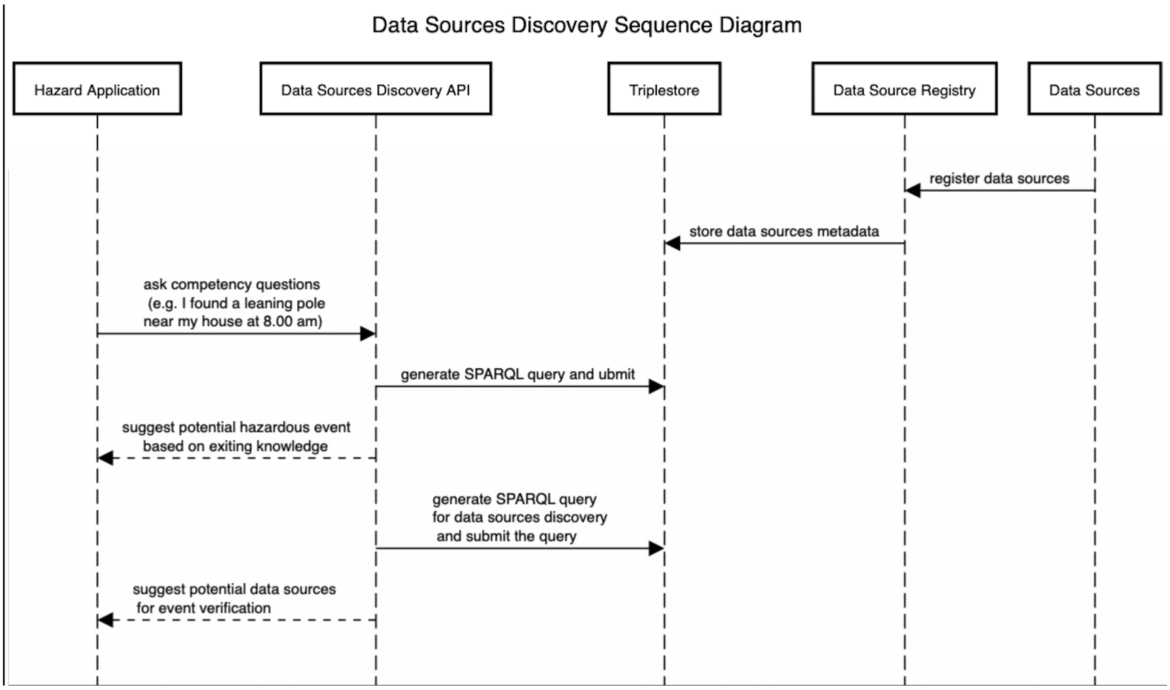

Figure 3.6: Overall of Landslip Ontology.

Figure 3.6 illustrates the interactions among the three layers. Initially, multiple data sources provided by different providers are registered to the data sources registry. In addition, the actual knowledge of landslide is constructed based on Landslip Ontology and information extracted from social media. Both data sources metadata and landslide knowledge are stored in Triplestore. Here, a hazard application utilizes the system by invoking the Data Sources Discovery API to ask a competency question which is related to landslide multi-hazard. The API then generates a SPARQL query which correspond to the selected competency question and submit the Triplestore for reasoning query. As a result, the API suggests potential hazardous event based on existing knowledge. Built from social media, the knowledge requires further analysis to verify the correctness of the suggested event. The API generates additional SPARQL query for the discovery of potential data sources. Finally, data sources metadata providing the detail of the potential data sources is returned to the hazard application. The application then use the information to access actual data sources and retrieve EO and urban time series data for hazard event verification and other data analytics.

3.6 Evaluation

An evaluation was conducted to verify the coverage of the Landslip Ontology and its application in landslide early warning. Whilst various approaches for evaluating an ontology exist, competency questions remain the most common approach [130, 131]. This approach stipulates that an ontology must be able to represent the competency questions using its terminology and answer these questions using the axioms [132]. According to the use case mentioned section 3.3, some of competency questions are developed as shown in table 3.2. We arranged an interview with domain experts who are members of the Landslip project. Those domain experts include two academic staffs who are specialist in natural hazard and geoscience, and a scientist from British Geological Survey (BGS). Based on the interview, 12 competency questions were received and some of main competency questions are developed as shown in table 3.2 for the evaluation.

The competency questions were used for ontology validation and evaluation. The evaluation was conducted using a set of synthesised data that represent the use case of landslide hazard mentioned section 3.3. We manually added information of natual hazards and EO and urban data to our knowledge base. The information includes landslide hazard, hazard triggers, warning signs, EO and urban data, and data sources. We performed validation over the dataset using Pellet to check for ontology consistency, concept satisfiability, classification, and realisation. Based on the competency

Table 3.2: An example of Competency Questions

	Competency Questions
Q1	What other hazards are likely to happen when hazard H has happened?
Q2	What is the probability of an event E occurring when warning sign W has been observed?
Q3	What is the probability of an event E occurring when a set of warning sign,$W_1, W_2, W_3, \ldots, W_n$ have been observed?
Q4	Is warning sign W an indicator for landslide L?
Q5	What are observed properties that can be used to verify landslide when a warning sign W is observed?
Q6	Identify the data sources and their metadata required to observe a set of hazards $(H_1, H_2, H_3, \ldots, H_n)$

questions, we performed preliminary experiments by querying over the knowledge base.

In order to write competency questions and to demonstrate that the LANDSLIP ontology can be used to ask and answer these questions we use the semantic query language called SPARQL Protocol and RDF Query Language (SPARQL).Using SPARQL query language, we defined query for each competency question to get answers from our knowledge base. Figure 3.7 and 3.8 show snapshots of SPARQL query for *Q2* and *Q6* and output for the competency question *Q6* on running the query in protégé. By executing the query based on the competency questions *Q1* - *Q6*, we could verify the coverage of the Landslip ontology. From the results it can be seen that the ontology is able to identify hazard events based on an observed warning sign. Furthermore, the ontology can suggest potential data sources and their metadata which can be used by domain experts to perform timely decision making against hazards.

Q1

```
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX :
<http://www.semanticweb.org/ncl/ontologies/2018/6/landslip#>

SELECT ?hazard
         WHERE { :landslide_2 :triggers ?hazard }
```

Q2

```
SELECT ?hazard
WHERE { :leaning_telephone_pole_1 :isWarningSignFor ?hazard }
```

Q3

```
SELECT ?warningSign ?hazard
WHERE { ?warningSign :isWarningSignFor ?hazard .
            VALUES (?warningSign) {
               (:crack_on_wall_2)
               (:crack_on_sideWalk_2)
            } .
}
```

Q4

```
ASK WHERE { :leaning_telephone_pole_1 :isWarningSignFor
:landslide_1 }
```

Q5

```
SELECT    ?observedProperty
WHERE {
         ?observation :isObservationFor :landslide_1 .
         ?observedProperty :isObservedPropertyFor ?observation .
}
```

Q6

```
SELECT  ?hazard ?observation ?observedProperty ?dataSource
?metadata ?profile ?p ?value
WHERE {
         ?observation :isObservationFor ?hazard .
         ?observedProperty :isObservedPropertyFor ?observation .
         ?dataSource :isDataSourceFor ?observedProperty .
         ?dataSource :hasDataSourceMetadata ?metadata .
         ?metadata :hasProfile ?profile .
         ?profile ?p ?value .
         VALUES (?hazard) { (:flood_1) (:landslide_1) } .
         FILTER (?p != rdf:type)
}
```

Figure 3.7: SPARQL Query for Competency Question Q1 to Q6.

hazard	observation	observedProperty	dataSource	metadata	profile	p	value
flood_1	obs_2	rain_1	dataSource_3	ds_metadata_3	sensor_profile_2	featureOfInterest	"foi_karela_bbox_1"@
flood_1	obs_2	rain_1	dataSource_3	ds_metadata_3	sensor_profile_2	sensorType	"in-situ"@
flood_1	obs_2	rain_1	dataSource_3	ds_metadata_3	sensor_profile_2	eventList	"Flood"@
flood_1	obs_2	rain_1	dataSource_3	ds_metadata_3	provider_profile_2	providerName	"MetOffice"@
flood_1	obs_2	rain_1	dataSource_3	ds_metadata_3	obs_profile_21	observedPropertyType	"Rain"@
flood_1	obs_2	rain_1	dataSource_3	ds_metadata_3	obs_profile_21	featureOfInterest	"foi_karela_bbox_2"@
flood_1	obs_2	rain_1	dataSource_3	ds_metadata_3	obs_profile_21	phenomenonEndTime	"2018-07-21T00:00:00"
flood_1	obs_2	rain_1	dataSource_3	ds_metadata_3	obs_profile_21	phenomenonBeginTime	"2004-01-01T00:00:00"
flood_1	obs_2	rain_1	dataSource_3	ds_metadata_3	obs_profile_21	observedPropertyName	"rain_1"@
landslide_1	obs_1	soil_moisture_1	dataSource_1	ds_metadata_1	service_profile_11	serviceURL	"http://127.0.0.1/rest/m
landslide_1	obs_1	soil_moisture_1	dataSource_1	ds_metadata_1	service_profile_11	serviceProvider	"Amrita"@
landslide_1	obs_1	soil_moisture_1	dataSource_1	ds_metadata_1	service_profile_11	serviceAdapter	"Rest_adaptor_11"@
landslide_1	obs_1	soil_moisture_1	dataSource_1	ds_metadata_1	service_profile_11	serviceType	"REST"@
landslide_1	obs_1	soil_moisture_1	dataSource_1	ds_metadata_1	sensor_profile_1	featureList	"foi_karela_bbox_1"@
landslide_1	obs_1	soil_moisture_1	dataSource_1	ds_metadata_1	sensor_profile_1	eventList	"Landslide"@
landslide_1	obs_1	soil_moisture_1	dataSource_1	ds_metadata_1	sensor_profile_1	sensorType	"in-situ"@

Figure 3.8: SPARQL Output for Competency Question Q6.

3.7 Summary

This chapter proposed the Landslip Ontology to address the research problem of discovering a wide variety of data sources. The *LO* represents the knowledge of the landslide domain and provides a knowledge base that establishes relationships between landslide hazard, social media, and time-series data sources. This thesis's essential contribution is used as a formal knowledge base of landslide domains throughout the thesis. Additionally, the *LO* is utilised to harmonise with electrical grid information services to monitor the electrical grid network's failure during a landslide, as described in the next chapter (Chapter 4). It is also utilised in Chapter 5 to facilitate monitoring the potential occurrence of landslide events and generate questions for communicating with social media users to obtain more detail of the potential event.

4

An Ontology-based System for Discovering Landslide Induced Emergencies in Electrical Grid

Early Warning Systems (EWS) for electrical grid infrastructure have played a significant role in efficient electricity supply management in natural hazard-prone areas. This chapter prototypes the system using landslides as an example of a natural hazard for electrical grid infrastructure monitoring. Essentially, the system consists of background knowledge about landslides derived from the Landslip Ontology proposed in Chapter 3 and information about data sources to facilitate the process of data integration and analysis. Using the *LO*, the prototype system can report potential landslide occurrence and suggest potential data sources for the electrical grid network monitoring. This chapter's main contribution is a process for the harmonisation of the knowledge base and electrical grid information services for monitoring of the electrical grid network.

4.1 Introduction

People around the globe rely heavily on electrical energy, provided by the electrical grid system, often more than other sources of energy. The electrical grid is a complex network of electrical power system which includes electricity generation, transmission and distribution (TD), and consumption. It provides a variety of operation to deliver electrical power from the place where it is generated to the consumers [133]. The infrastructure of the electrical grid system comprises of several key components to support the delivery of electricity to consumers: (i) *Generating Plants* – where electricity is produced; (ii) *Transmission Networks* — an infrastructure which allows high voltage electricity to be transported over long distances; (iii) *Substations* — where the electricity voltage is changed by utilising a transformer mechanism; and (iv) *Distribution Networks* — an infrastructure, similar to the transmission lines, to transmit lower voltage electricity. High voltage electricity from the generating plants is transmitted along the transmission lines from where it reaches the substations in the grid network. Following a reduction in the voltage of the electricity by the transformer at the substation, it travels along the distribution line to various types of consumers (including industrial consumers, commercial consumers, and resident consumers). Specifically, transmission and distribution networks are set up to cover the whole geographic area in the country. The transmission network carries overhead electricity lines on pylons, a steel lattice tower, while the distribution network transmits electricity either through an overhead line or underground. The overhead line in the distribution network is carried on small steel towers, concrete poles or wooden poles [134].

Reliability is the most crucial element in the operation of electrical grid systems. A failure of the grid system infrastructure can lead to disruption of electricity supply, leading to major economic chaos in the country as well as impacting upon the safety and well being of people in the affected area. In particular, the transmission and distribution network within the grid system infrastructure is frequently affected by climate change and natural hazards such as landslides, earthquakes, and flooding. Grid systems are often the most frequently affected by landslides as can be noted from reports in several countries. In 2018, a landslide occurred near Invergarry in Scotland and damaged overhead power lines[135]. As a consequence, 23,000 people in Skye and the Western Isles had their electricity supply cut off for several hours after the event. The same year, in Thailand[136], a landslide in a waste dump zone at the Electricity Generating Authority of Thailand (EGAT) Mae Moh mine in Lampang province, damaged some electricity poles and led to road closures in the area. The disruption was so serious that an evacuation plan for the people had to be put in place. The Wenchuan Earthquake in 2008 [137] caused serious damage to the Sichuan electrical grid system. A number of electrical equipment, transmission and distribution networks were broken and buried due to landslides that occurred during the earthquake. Due to minimal protection of transmission and distribution networks, they are very vulnerable to such natural hazards. EWS can therefore play a vital role in monitoring the grid network to predict such failures and minimise the consequent disruptions.

Nowadays, EWS for natural hazards utilise strong technical underpinning and sophisticated knowledge of natural hazards such as the hazard context and risk factors to enable dynamic and timely decision-making. Landslides, the natural hazard this chapter focuses on, have global significance given their frequency of occurrence as well as potential to cause disruption. Where electrical grid systems are commonly affected by landslides, it is because of parts of the grid infrastructure being located in landslide-prone areas. Moreover, landslides are also closely linked with a variety of other natural hazards such as storms, earthquakes, floods, and volcanic eruptions. These hazards can also affect electrical grid systems. Therefore, the prediction of individual landslide occurrence is beneficial to the monitoring of electrical grid infrastructure. However, such prediction is complex as it depends on many local factors, variables, and anthropogenic activities (caused or produced by human beings). Current EWS for landslides rely on scientific methods such as hyperlocal rainfall monitoring, slope stability models, and analysis of remotely sensed images. With the emergence of Internet of Things (IoT), decision makers are also analysing observation and measurement data produced by sensors (e.g., soil moisture, soil movement, rainfall, humidity, wind speed) which are deployed in landslide prone areas. [138, 139] Furthermore, other observation and measurement data such as wind speed, wind direction, soil temperature, tilt, and vibration are also being used for the monitoring of electrical grid system. [134, 140]

Before EWS for landslide can optimally utilise information from multiple, heterogeneous time series of data sources (IoT sensors), it is essential to realise a common knowledge base for capturing the core conceptual information and the cross co-relationship between events (that could be potentially discovered by analysing those data sources). Moreover, cross co-analysis of time series data sources is not only useful for the discovery of event correlation but also allows for the interaction of electrical grid system with EWS. For example, a landslide detected by processing of time series data from IoT sensors (e.g. using accumulative rainfall threshold) provides information about location and time of the landslide occurrence. This information can be used to identify parts of the electrical grid infrastructure that are potentially vulnerable to the detected landslide and need to be monitored intensively. However, discovering such cross correlation of events from heterogeneous time series data sources has many challenges including a lack of common terminology that make analysis particularly difficult.

The main *contributions* of this chapter are as follows.

1. A formal knowledge base of landslide domain concepts to enable the integration of time series data from multiple and heterogeneous data sources for the early prediction of landslide events.

2. A process for the harmonisation of the knowledge base and electrical grid information services for monitoring of electricity grid network.

The results of landslide prediction are utilised to suggest the monitoring of electrical grid infrastructure in order to minimise the loss of electric energy during natural hazard events. Underpinning this knowledge base is the Landslip Ontology that captures the concepts of and relationships between landslide, landslide-related hazards, warning signs, sensor data and other time series data sources. The purpose of the ontology is to facilitate data discovery, which will be used to find potential data sources for landslide prediction and electrical grid infrastructure monitoring. The proposed Landslip Ontology is evaluated using competency questions for electrical grid systems in landslide prone areas. The experimental results show the accuracy of the data discovery mechanism and indicate the benefits of using Landslip Ontology in electrical grid management applications.

The rest of the chapter is organised as follows: related work is discussed next followed by a discussion on the landslip scenario in Section 4.3. Landslip Ontology for electrical grid is described in Section 4.4 and the design of data sources discovery system in Section 4.5. Evaluation of the Landslip Ontology is discussed in Section 4.6. Finally, summary of this chapter is presented in Section 4.7.

4.2 Related Works

4.2.1 Data Utilisation in Multi-Hazard Early Warning System

The term, multi-hazard, refers to a collection of multiple major hazards that a country faces [6]. There is a possibility that several hazardous events occur simultaneously and are interrelated. Tropical storms, for example, are one of the most common environmental hazards (in the tropics), which can trigger multiple hazards such as heavy rainfall that in turn can induce flash flooding. Furthermore, heavy rain and flooding can increase the moisture content of soil in mountainous areas inducing landslides. To minimise the loss of life and property damage from these inter-related hazards, a comprehensive strategy for hazard management is required. In general, a strategy for hazard management comprises of four phases [108]: (i) *mitigation* — actions to minimise the cause and impact of hazards and prevent them from developing into full-blown disasters; (ii) *preparedness* — action plans and educational activities for communities to confront unpreventable hazard events; (iii) *response* — actions for emergency situations to protect peoples' lives and properties during hazard or disaster events; and (iv) *recovery* — the actions to restore damaged properties and community's infrastructures and to provide medical care to the affected population. These four phases require supporting tools and technologies to improve the effectiveness of hazard management.

Several modern multi-hazard early warning systems take advantage of the data explosion on social media. The authors in [64] propose using a Twitter data analysis framework for identifying Tweets that are relevant to a particular type of disaster (e.g. earthquake, flood, and wildfire). Several classification techniques, including keywords and hashtags matching and classification machine learning, are also evaluated to identify tweets which are relevant to a particular hazard. The work in [109] studies the potential of using social media data to identify peatland fires and haze events in Sumatra Island, Indonesia. A data classification algorithm is used to analyse the Tweets and the results are verified by using hotspot and air quality data from NASA satellite imagery. A data classification algorithm is also used in [57] to automatically classify

Tweets and text messages (from the Ushahidi crowdsourcing application) generated during the Haiti earthquake in 2010. The goal of their work is to provide an information infrastructure for timely delivery of appropriately classified messages to the appropriate responsible departments. Work in [110] proposed a decision support system that integrates crowd sourcing information with Wireless Sensor Networks (WSN) to improve the coverage of monitoring area in flood risk management in Brazil. The research introduces the Open Geospatial Consortium (OGC) standards to facilitate the data integration among crowd sourcing information and WSN.

4.2.2 IoT Resource Management

The emergence of Internet of Things allows decision makers to analyse observation and measurement data produced by IoT devices. These IoT devices have the ability to sense, process, communicate and store the data observed or measured from the physical world [141]. Moreover, the number of IoT devices has increased dramatically and they are heterogeneous in nature. Based on this, efficient techniques for IoT resources management have been investigated to address the challenging problems of IoT (e.g. IoT management framework, data processing, and security) [142–146]. Here, several frameworks for IoT resource management have been proposed. Authors in [142] proposed a paradigm of Everything-as-a-Resource (*aaR) to enable efficient resource allocation of collaborative applications on the Web. The framework has been applied to IoT applications in the healthcare domain. [143] proposes a resource preservation net (RPN) framework for IoT resource management in edge computing. The framework has been applied to smart healthcare applications where real-time systems with complex and dynamic behaviour are essential parts of the systems but suffer from resource shortage and resource management efficiency challenges. In the RPN framework, a smart healthcare workflow and non-consumable resource pools are defined to enable process execution and resource assignment in the cloud and solve the problem of resource management efficiency. Besides, a number of approaches on IoT data analysis have been proposed to address issues in multi-hazard and electrical grid applications. The work in [144] presents a novel technique of machine learning and neural network to predict the severity of floods. Essentially, a machine learning technique is utilised to analyse new datasets of flood events to predict the severity of flood events and classify outcomes into normal, abnormal, and high-risk flood. The prediction of the flood severity aims to address issues of flood mitigation. Security and privacy are crucial issues in IoT resource management due to the sensitivity of IoT data in many application domains. The research in [145] focuses on securing IoT-enabled applications at the Fog layer to secure a massive amount of sensitive data produced by IoT devices and enable efficient resource consumption (i.e., memory, storage and processing) of the IoT devices. The work in [146] proposes a secure fog-based platform for SCADA-based IoT critical infrastructure. The platform is designed to address the performance and security issue of Supervisory Control and Data Acquisition (SCADA) systems and enhance security of data generated from IoT devices and deploy edge data centers in fog architecture.

4.2.3 Semantic Web Technologies and High Variety Data Management for Multi-hazards

Earth Observation (EO) and ancillary data provided by multiple data sources are accessible by different methods ranging from direct download to various standard Web Services APIs (e.g. Web Map Services, Web Feature Services, Sensor Observation Services, RESTful API, SOAP-based API, etc.). In addition, there is heterogeneity among EO and ancillary data provided by different data sources [111] including: (i) syntactic heterogeneity — the difference in data format or data model for presenting datasets (e.g., plain text, CSV, Excel, XML, JSON, O&M, SensorML, etc.); (ii) structural heterogeneity — the difference in data schema for describing the same types of datasets (e.g., describing soil moisture using different XML Schemas); and (iii) semantic heterogeneity — difference in meaning or context of the content in datasets. This heterogeneity reveals the challenging problems brought forth by the high variety of data in multi-hazard applications. Semantic Web Technologies play a significant role by providing languages and tools for modelling domains including consistent and formal descriptions of concepts and relationships among the data and hazardous events. According to the W3C definition [112, 113], the Semantic Web is a web of data that provides a common framework for data sharing and reuse across applications, enterprises, and communities.

Ontology, a key element of the Semantic Web, is a specification of a conceptual model for describing knowledge about a domain of interest. A basic concept in a form of ontology can be described by an Resource Description Framework (RDF) triple [114] which is comprised of a subject, a predicate and an object. Concepts described by RDF can be extended by Web Ontology Language (OWL) [115] to construct an ontology for representing rich and complex knowledge about things. In the case of multi-hazard applications, an ontology can be used to: (i) represent domain knowledge through concepts, their attributes and relationships between data sources, data and hazards; and (ii) facilitate data integration across multiple data sources that represent variety, velocity and volume characteristics of big data.

Ontologies are widely used in hazard management to model knowledge about hazards and to manage actual data derived from EO and ancillary sources. Ontologies also promote the associative retrieval in spatial big data [147]. Hazard assessment and risk analysis are two of the common application areas where ontologies are used. The Semantic Sensor Network Ontology (SSN) [28] and the Semantic Web for Earth and Environmental Terminology (SWEET) [30] are two of the commonly applied ontologies in hazard management applications. The SWEET ontology is reused to conceptualise the knowledge from several areas, such as buried assets (e.g. pipes and cables), soil, roads, the natural environment and human activities. Additionally, the Ontology of Soil Properties and Process (OSP) is proposed in their work to describe the concept of soil properties (e.g., soil strength) and processes (e.g., soil compaction). The ontology is used to express how asset maintenance activities affect each other. Furthermore, [28] and [116] present the application of SSN for wind monitoring. The former uses SSN with Ontology for Quantity Kinds and Units (QU) [117] to conceptualise wind properties (e.g. wind speed and direction) while the latter uses SSN and SWEET to model the concepts of wind sensors and data streams of wind observations. The Landslides ontology [29] extends SSN to organise knowledge for the landslides domain such as the concepts of landslides, earthquake, geographical units, soil, precipitation and wind. Even though these ontologies provide comprehensive concepts for sensor data and hazard event, and provide a reusable, widely used semantic underpinning, they do not cover conceptual aspects of human sensors (e.g. social media data). Hence, currently additional processes are required when applying these ontologies to EWS for multi-hazard applications.

The related literature in the context of multi-hazard management can be classified based on the following three perspectives: *data sources*, *hazardous event analytics*, and *EO and ancillary time series data management*. It can be seen that effective multi-hazard management demands high quality and rich data from a vast amount of data sources that are related to the hazard of interest. Data sources utilised by multi-hazard management applications can be sensors and/or data services that provide EO and ancillary data. Such data sources include *physical sensors* (e.g., remote sensors, in-situ sensors, wireless sensor network) and *human sensors* (e.g., social media, blogs and crowd sourcing). Recent data analytics research for multi-hazard management is focused on hazardous event analysis, which has three main directions: *event identification*, *event verification*, and *event prediction*. The research in this area reveals the challenging problems in EO and ancillary time series data management, especially the discovery of potential time series data sources given the complexity and high variety of such data sources in multi-hazard management applications.Several ontologies [148–150] have been proposed for not only modelling knowledge about hazards but also managing EO and ancillary data. They have shown that current standard ontologies for data sources discovery do not exist. In addition, existing applications of ontology in this domain mostly investigate specific problems, in other words these approaches are not generalised. They fail to model the relationship between data sources and the domain knowledge, which is an important factor for efficient data integration and data sources discovery.

4.3 Landslide Scenario for electrical grid Early Warning System

Efficient EWS for landslide multi-hazards is essential for the prevention and mitigation of electrical grid failure in hazard-prone areas. Generally, the development of EWS for natural hazards can be accomplished through several approaches [118], depending on:

(i) the rules stakeholders engage in hazard risk reduction, (ii) geographical conditions of the hazard-prone area, and (iii) EO and ancillary data provided by responsible parties. The approaches have shown the significance of the synchronisation of EWS for landslide multi-hazards and EWS for electrical grid systems. Based on this, a scenario-based approach [119] is applied in this work to specify the scope of the EWS for an electrical grid system and its synchronisation. The scenario-based approach describes a story that represents the ordinary uses of a system in the domain of interest from both, domain experts' and ontology developers' viewpoints. The scenario thus helps to identify the scope of the domain ontology.

4.3.1 Scenario

The Landslip scenario for electrical grid EWS focuses on the *preparedness phase* of disaster management where the prediction of individual landslides' occurrence using time-series data sources is used to predict possible failure of electrical grid infrastructure. Several techniques for landslide prediction rely on the analysis of time-series data from rain-gauge sensors [151, 152]. An example of the technique is to calculate local rainfall thresholds for the occurrence of landslides [153]. In the local areas of interest, the rainfall threshold is determined by the extraction of local rainfall events from daily accumulative rainfall to reconstruct triggering rainfall conditions for landslide occurrences in particular areas. Here, the calculated thresholds are used for analysing real-time accumulative rainfall for predicting the occurrence of landslides and their location. Parts of transmission networks or distribution networks that are vulnerable to landslides in the predicted areas are identified and monitored.

The interaction between landslide and electrical grid system is considered when designing the scenario. Figure 4.1 illustrates a situation before the occurrence of a landslide event in a remote location. This area is a high slope and encompasses both natural environment (e.g. rivers and mountains) and built environment (e.g. schools, hospitals, road, water supply and electricity). The area is prone to landslides and is monitored by the National Disaster Management Authority (NDMA). An expert from NDMA explores potential data sources from the *Data Sources Discovery Service (DS)* and gathers EO and ancillary data from the discovered data sources. The expert then utilises the *Early Warning System (EWS) for landslide* to detect warning signs by analysing daily rainfall, soil moisture, and water level and informs decision makers of the potential landslide hazard. The EWS also sends event notifications to other systems to inform of the potential landslide hazard. The event notification is accompanied with additional information including prediction time, geo-location of landslide occurrence, and geographical boundary of the place where the landslide is likely to occur. Meanwhile, the *EWS for electrical grid system* monitors the overall operation for the delivery of electricity to consumers. On receiving an event notification from the landslide EWS, the electrical grid EWS uses the geographical boundary to identify distribution networks and list of distribution poles that are located there is a potential landslide hazard that needs to be monitored. This process is achieved by invoking third party services provided by electrical grid system providers. Here, the geo-locations of the distribution poles are identified. These geo-locations are used to discover data sources that provide observed properties for distribution pole monitoring (e.g. wind speed, wind direction, soil temperature, tilt, and vibration). Gathered from data sources, the observed properties are analysed in real-time to monitor the failure of each individual distribution pole. To summarise, the distribution poles that are highly vulnerable to landslides are identified and the EWS informs the decision maker about the possibility of a potential failure of the distribution poles.

4.3.2 Concepts

The scenario reveals the essential role of data-driven EWS for landslide risk prediction and electrical grid system monitoring, which comprises of five main components.

- *Exposure* — refers to people and the environment in landslide hazard-prone areas. The environment comprises natural and built environments. The natural environment encompasses of living and non-living things (e.g. animals, river, forest, mountain, etc.). The built environment [120] is a core foundation of the community, which is constructed by people. It is comprised of infrastructure and facilities (e.g. house, school, road, bridge, electricity, water supply, etc).
- *Stakeholder* — refers to people or organisations who have a stake in the landslide

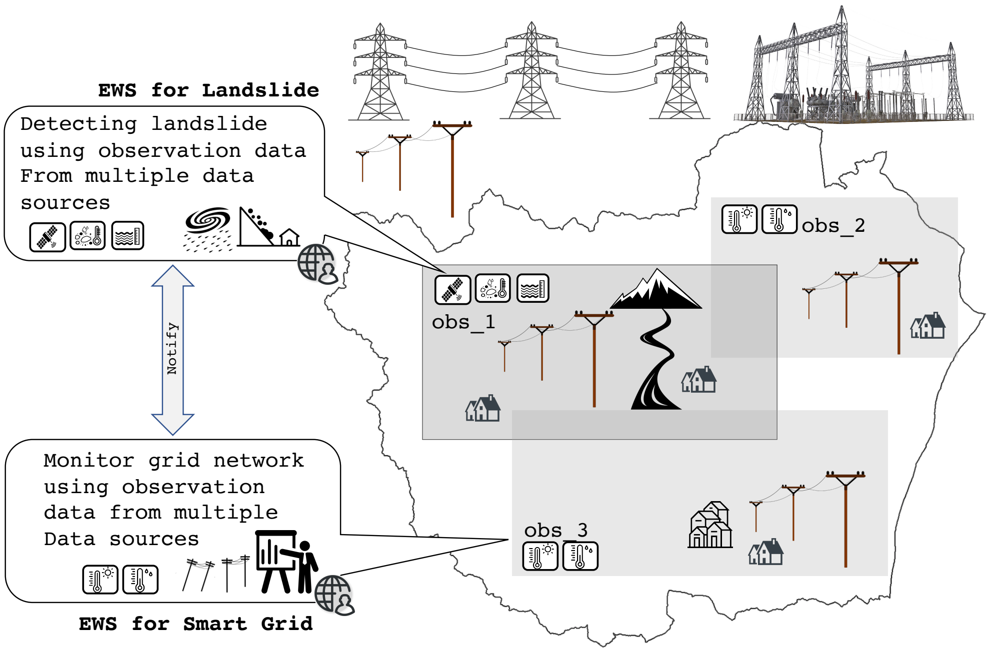

Figure 4.1: Landslip Scenario for Electrical Grid Early Warning system.

or the electricity grid failure event.The stakeolders could include: (i) *data collectors and providers* who deploy sensor devices in a landslide hazard-prone area or electricity grid components and provide EO and ancillary data collected from the sensor devices to EWS for analysis. The third parties who provide EO and ancillary data collected from others are also considered as data providers and (ii) *Decision makers* who have the responsibility for conducting landslide hazard risk assessment using EO and ancillary data. They make a decision based on the result from the Decision Support System and hazard risk management plan in order to inform people in a risk area and other organisations before the occurrence of landslides.

- *Event* — refers to an occurrence which is related to a hazard and electrical grid system. The hazard itself is also considered as an event based on the context. Hazard-related events can be classified as pre-hazard events, post-hazard events and events during a hazard. Since the aim here is to monitor the failure of the electrical grid system, landslide events are the majority of events in this scenario. Landslide events can indicate the distribution networks and list of distribution poles to be monitored by the EWS for an electrical grid system.

- *Data Sources* — refers to any sensor devices and data services that provide EO and ancillary data to data consumers. These data sources have different capacities to provide data. Sensor devices are components that observe and measure physical phenomena and transform the observation and measurement into a human-readable form. A data service is an application software that stores and provides data collected from multiple sensor devices. Nowadays, EO and ancillary data for multi-hazard and electrical grid management applications are available from several types of data sources.

- *Decision Support Applications* — refers to an integrated system that provides functionalities for stakeholders to monitor, forecast and predict, validate and assess hazardous events. In this scenario, EO and ancillary data collection systems, data sources discovery services, and EWS for landslide and electrical grid are significant components of Decision Support Applications. These applications enable stakeholders to take timely actions to reduce the impact of landslide hazard and electric power shortage in advance. For example, once is it established that the failure of distribution poles is likely to happen, a decision-maker can co-operate with an electrical grid provider to prepare for the maintenance of the poles or prepare for mobile power generation in the landslide occurrence area.

The data-driven EWS realises dynamic and timely decision making for landslide prediction and monitoring of the electrical grid system by analysing EO and ancillary data. Such data includes historical landslide events, electrical grid components, and historical and real-time data produced by sensor devices. Additionally, several sensor devices have been deployed in the landslide hazard-prone area by organisations who are responsible for landslide hazard management. Also, electrical grid components are equipped with sensors devices to observe the status of the components. These sensor devices produce EO and ancillary data and send to EWS to monitor the landslide hazard and electrical grid system in real-time.

Furthermore, the organisations store the data in their local repositories and provide the data repositories as data sources for further analysis. Here, metadata of these data sources are published to a *Data Sources Discovery Service (DS)*, which is a part of Decision Support Systems. The DS allows data publishers to advertise their data sources by registering data sources metadata via a data sources registry service. Furthermore, it allows data source consumers to explore potential data sources from the service to be used in their applications.

4.4 Landslip Ontology for electrical grid Network Monitoring

The monitoring of electrical grid network failure, the focus of this chapter, relies on the prediction of landslides. This prediction requires rich information from multiple data sources to provide more accurate predictions. For this purpose, the Landslip Ontology as presented in Chapter 3 is reused and utilised to provide an efficient data sources discovery mechanism in landslide prediction and electrical grid network monitoring. Basically, the landslip Ontology was designed and developed to conceptualise the knowledge of landslide hazard and its warning signs. Moreover, knowledge of data sources is also provided to facilitate data sources discovery and landslide precursor verification. The ontology has been used to support data integration and analysis in landslide early warning application using social media. The Landslip Ontology comprises of two main modules, Landslip Common and Landslip Data Sources. According to the scenario mentioned in Section 4.3, the Landslip Data Sources is reused for efficient data integration in the EWS for landslide and electrical grid systems. To account for the lack of electrical grid knowledge representation, the Landslip Ontology is reused by tying together with external services provided by the electrical grid providers in order to indicate failure of the electrical grid network.

Scope and Purpose — The goal of the Landslip Ontology application for electrical grid system monitoring is to indicate the distribution poles in the electrical grid system that are vulnerable to landslide hazard and are likely to fail. Thus, the application of the Landslip Ontology focuses on the preparedness phase of disaster management where landslide events play an important role to enhance the efficiency of the monitoring. The Landslip Ontology conceptual knowledge of landslide hazard, multi-hazard interaction, and landslide-related incidents is utilised to support data integration and analysis in landslide hazard and electrical grid systems. The concepts of landslide hazards are linked to EO and ancillary data, which constitute a set of properties for landslide observation. Even though the ontology focuses on the landslide multi-hazard domain, the concept of data source in the ontology can also be applied for the electrical grid system application. The level of granularity is determined based on the competency questions and the terms identified thereof. However, external services from the electrical grid system are also required in order to answer the competency questions.

Knowledge Sources — Built for landslide EWS, the ontology is designed based on knowledge and experiences from scientists and experts who have the domain knowledge of landslide hazard management. Research [122, 123] and standard specifications [28, 30, 124–126] involving multi-hazards and geo-spatial data models are also used as additional knowledge sources to design the ontology. Further, related research works [133, 134, 154, 155] were reviewed as knowledge sources of electrical grid systems and the assessment of electrical grid networks for the design the harmonisation of the ontology and the electrical grid information services. Moreover, this knowledge is also used to conceptualise data sources for an electrical grid network assessment.

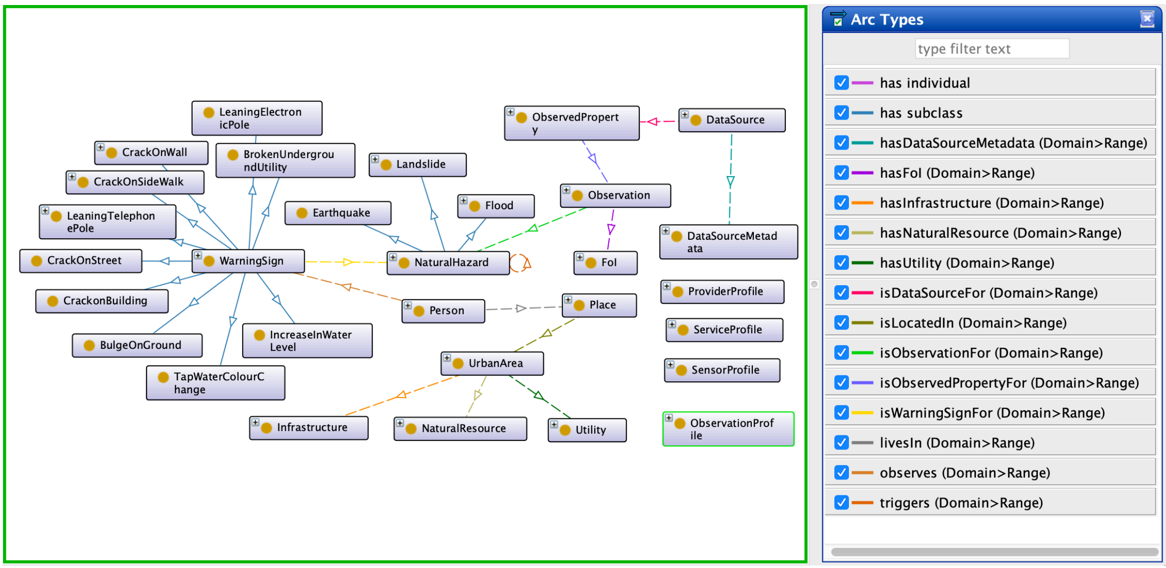

Figure 4.2: Snapshot of Landslip Ontology.

Figure 4.2 is a snapshot of the Landslip Ontology, which is comprised of two modules, *Landslip Common Ontology* and *Landslip Data Sources Ontology*. The Landslip Common Ontology defines concepts about landslide hazard and its interaction with other hazards and anthropogenic processes. The Landslip Data Sources Ontology defines concepts about observation and data sources for landslide hazard and electrical grid systems. The Landslip Ontology reuses SSN ontology and terminology defined in OGC standards (e.g. Observation and Measurement [124], SensorML [125] and SOS [126]).

4.4.1 Landslip Common Ontology

The Landslip Common Ontology conceptualises the knowledge of landslides hazard. The ontology model combines theoretical knowledge and human experiences to identify warning signs before the occurrence of landslides. Landslides are one of the most significant multi-hazards found in many places around the globe [127]. Landslides not only interact with but are also triggered by other hazards [122]. Therefore, the Landslip Common Ontology conceptualises knowledge of landslides and the interaction with other multi-hazards [122, 123]. It also conceptualises knowledge of warning signs, observed by humans, which are used to indicate possible landslide events before the occurrence of a landslide. Thus, the ontology represents warning signs of a landslide, observed and reported by people in a social network, that can be used to facilitate social media-based early warnings.

The Landslip Common Ontology comprises of four main concepts:

- *RemoteArea* — defines concepts about a remote area that is prone to landslide. The remote area encompasses both natural environment (e.g. river and mountain) and built environment which includes infrastructure (e.g. road and railway), utility (e.g. electricity and tap water) and place (e.g. school, health care unit, and house). Located in a landslide-prone area, these elements can be affected by landslides and other multi-hazards.

- *NaturalHazard* — defines a set of multi-hazards that can trigger landslides. This concept mainly captures knowledge about the interactions between landslide hazard and other multi-hazards (e.g. flood, earthquake, tsunami, and drought). In addition, it also captures the interactions between other multi-hazards that can, in turn, indicate the (potential) occurrence of landslides.

- *AnthropogenicProcess* — defines a set of human activities that are contributing factors in causing landslides [122].The knowledge of interactions within the processes is also captured to conceptualise direct and indirect indication of landslide hazards. Direct indications refer to the processes that trigger landslides while indirect indications refer to the processes that trigger other processes, which in turn, trigger landslides.

- *WarningSign* — defines a set of incidents for landslide hazard indication, other multi-hazards and anthropogenic processes. The concept of warning signs is mainly focused on incidents that are observed by a person or EWS.

4.4.2 Landslip Data Sources Ontology

EO and ancillary data observed by sensor devices indicate events or changed pattern of landslide phenomena. Such data (e.g. rain, soil moisture, electrical grid components) from a variety of sensor devices is collected by data providers and provided as data sources for stakeholders to be used in their landslide hazard applications [128]. Due to the wide variety and geographically distributed nature of EO and ancillary data sources, it is essential to investigate efficient data source discovery [129] to provide sufficient amount and quality of data sources for landslide hazard risk assessment and electrical grid system monitoring. the Landslip Data Sources Ontology is thus designed to enable discovery of data sources semantically. This ontology represents concepts and relationships of EO and ancillary data, data sources, sensor devices, and data providers. When combined with the Landslip Common Ontology, the knowledge of landslide hazards can enhance data source discovery mechanisms to efficiently discover data sources that are related to the hazard of interest. Specifically, the knowledge of landslide warning sign can identify appropriate observed properties and data sources for the verification of a landslide precursor. This capability enables EWS to provide dynamic and timely decision-making for landslide hazards.

The Landslip Data Sources Ontology is comprised of three main concepts. The concepts of observation and sensors reuse existing ontologies, SSN Ontology [28] and OGC standard [124–126].

- *DataSource* — is the central concept of the Landslip Data Sources Ontology. A data source is any sensor (e.g. physical sensor, human sensor) or data service that provides EO and ancillary data. DataSource defines a set of comprehensive information about observation and data sources metadata.

- *Observation* — defines a set of observed properties (EO and ancillary data) that are used to observe features of interest related to landslide hazard. Examples of observed properties include rain, earthquake magnitude, soil moisture, soil movement, temperature, humidity, and wind speed. These observed properties are accessible to EWS via data sources.

- *DataSourceMetadata* — defines a set of information, which is essential for the data acquisition mechanism. This concept is comprised of four groups of profiles namely, ObservationProfile, SensorProfile, ServiceProfile, and ProviderProfile. The ObservationProfile represents a set of observed properties provided by a data source. SensorProfile provides information about sensor type, a feature of interest, and a list of events to be observed. ServiceProfile provides information which can be used to access a service (e.g. service type, endpoint, provider). Finally, ProviderProfile provides information about a data provider (e.g. provider name, contact address).

4.4.3 Ontology Metrics

An ontology comprises of a finite list of concepts and the relationships among them to represent the domain of interest [156]. The ontology metrics illustrate the number of classes, properties, individuals, and Description Logic (DL) expressivity of the ontology. *Classes* describe concepts of the domain of interest at an abstract level. *Properties* describe features and attributes of the classes and relationship among classes. *Individuals* are instances that represent concrete objects of the classes. For example, the classes `Landslide`, `Earthquake` represent landslide and earthquake events respectively. A property `triggers` represents the relationship between `Landslide` and `Earthquake` concepts where a specific `Earthquake` event triggers a specific `Landslide` event. A specific landslide event (e.g. landslides triggered by the Hokkaido

Table 4.1: Landslip Ontology features

Feature	Value
Number of classes	98
Number of properties	26
Number of individuals	30
DL expressivity	$\mathcal{ALCH}(\mathcal{D})$

earthquake in Japan, 2018) is an individual or instance of `Landslide`. Table 4.1 shows a summary of the ontological features of Landslip Ontology in terms of size (number of classes, properties, and individuals), expressivity, and complexity of the core knowledge of the Landslip Ontology.

The DL expressivity represents the complexity of the logic underlying a particular ontology [157]. For Landslip Ontology, $\mathcal{AL}$ (Attributive Language) [158] is used to represent its complexity. Here the DL expressivity of Landslip Ontology is represented by $\mathcal{ALCH}(\mathcal{D})$ which comprises of (i) $\mathcal{AL}$ — a Description Logic used to describe the ontology, (ii) $\mathcal{C}$ — an extension for representing Concept Negation; (iii) $\mathcal{H}$ — an extension for representing Role hierarchy; and $(\mathcal{D})$ — an extension for representing data type.

4.5 Electricity Grid Network Monitoring using Landslip Ontology

To enable efficient electrical grid network monitoring under the condition of landslide hazard, EWS for electrical grid systems need to harmonise the Landslip Ontology with information services provided by electrical grid providers. For example, the EWS utilises knowledge base provided by the ontology to indicate the potential occurrence of landslide and its location. In addition, both the knowledge base provided by Landslip Ontology and the electricity grid provider have the coordination information which the grid network monitoring system can use to integrate both systems based on geolocation of detected landslide. Such information, particularly the landslide location, can then be utilised by invoking the information services to retrieve a list of distribution poles that it is necessary to monitor due to the landslide. To realise this, interaction between the electrical grid system and the landslide hazard needs to be investigated and processes for the harmonisation between the Landslip Ontology and the electrical grid information services need to be defined.

In this research, we focus on the monitoring of electrical grid network where the transmission network and the distribution network are deployed across large regions in the country including remote areas that are prone to hazards. A significant correlation between the electrical grid network and landslide occurrence is *geographical information* (e.g. geo-location and geographical coverage). With a prediction of landslide, geographical coverage of the landslide affected area is identified. In addition, the geographical coverage is represented as a boundary or bounding box coordination. Meanwhile, geo-locations of electric pylons and poles within an electrical grid network are identified by their individual geographical coordination (e.g. latitude, longitude). Based on this, potential vulnerable electric pylons and poles are indicated by searching for pylons and poles where their geo-locations are inside the geographical coverage of the predicted landslide.

The process of harmonising the Landslip Ontology and electrical grid information services in electrical grid EWS is divided into two sub-processes: landslide EWS process and electrical grid EWS process. These sub-processes interact with each other and require EO and ancillary data for their analysis. EWS for landslide collects EO and ancillary data from multiple sensors deployed in the landslide prone area to analyse and predict the occurrence of landslide. The EO and ancillary data produced by data sources (e.g. IoT sensors) is a representation of observations. The *observation* is a collection of measurement of phenomena for observing the changing pattern of the area of interest. The measurement of phenomena is represented as an *observed property*, which is observed or measured by sensors deployed in the area of interest. Landslide observation comprises of sensors that observe or measure properties of landslides. Examples of the observed properties for landslide are precipitation, soil movement, soil potential, temperature, and humidity.

The sequence diagram in Figure 4.3 illustrates the interaction among the components of landslide EWS in order to monitor and predict the occurrence of landslides. Initially, multiple data sources provided by different providers are registered to the data sources registry. In addition, the actual knowledge of landslides is constructed based on the Landslip Ontology. Both data sources, metadata and landslide knowledge, are stored in a triplestore, which is a semantic database. A hazard application utilises the system by querying the knowledge base to retrieve processing rules that can be used for the monitoring of landslides in the area of interest. Next, the system submits a query to the knowledge base again to perform data source discovery to search for potential data sources that correspond to the processing rules. Thereafter, the EWS collects data sources based on suggested information from the knowledge base and starts processing based on the suggested rules. Subsequently, decision makers are notified of landslide events detected by the EWS. Such information includes event types, time, geo-location, affected area, and other processing results. This information is also used by other systems including the electricity grid monitoring system.

Figure 4.4 illustrates the utilisation of knowledge base in an electricity grid monitoring systems (EGMS). Once a landslide is predicted, the landslide EWS sends a notification with the landslide information to the EGMS. The information includes geo-locations and the areas likely to be affected by the predicted landslide. The areas are are represented by the bounding box from the observation. This extracted information is used to identify electricity poles, which are at risk of failure caused by landslide. Next, the EGMS calls external services provided by electricity grid providers to retrieve a list of electricity poles located in the affected areas including their metadata. Thereafter, the EGMS submits a query to the knowledge base to get potential processing rules and observed properties, which are used to monitor the failure of the potential electricity poles. Next, data sources are discovered by invoking the data source discovery service to facilitate the monitoring of electricity pole failure. Collecting data from potential data sources, EGMS is able to process the EO and ancillary data to monitor the failure of the electricity grid network in the landslide prone area.

4.6 Evaluation

An evaluation was conducted to verify the coverage of the Landslip Ontology and its application in electrical grid network monitoring. This includes the harmonisation

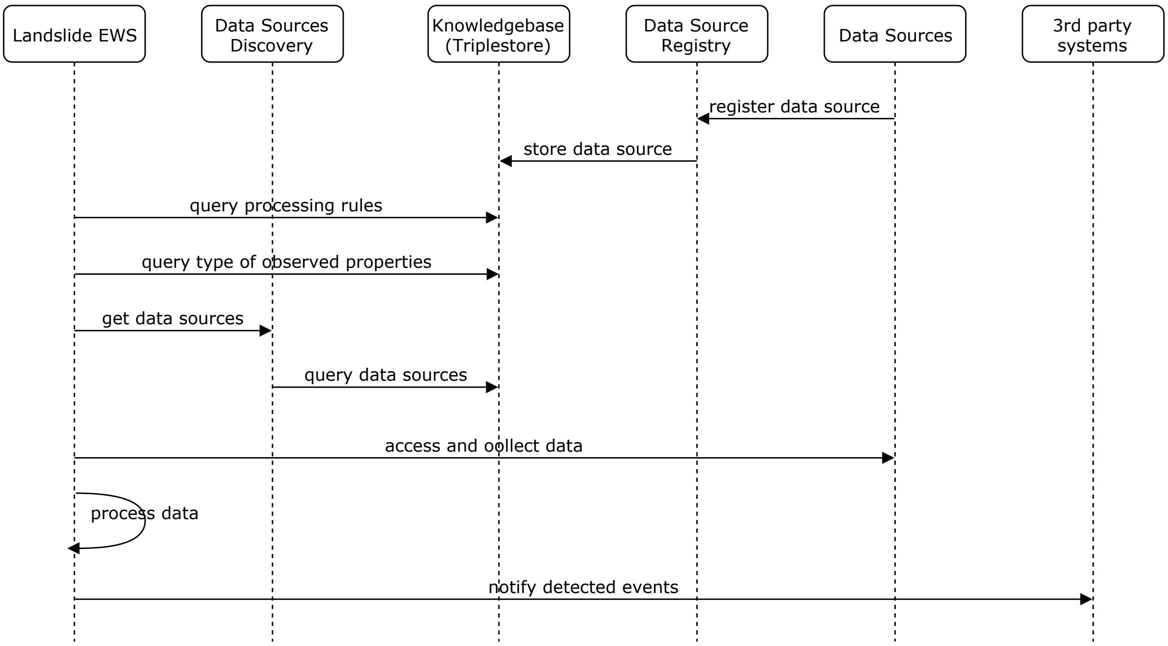

Figure 4.3: An interaction among the components of EWS for landslide.

Table 4.2: An example of Competency Questions

	Competency Questions
Q1	Which distribution networks are affected by landslide L?
Q2	Which distribution poles are affected by landslide L?
Q3	Which distribution networks are located in a landslide prone area?
Q4	Which distribution networks or substations need to be monitored because of the potential of the occurrence of a landslide hazard?
Q5	What observed properties O can be used to monitor a distribution network D?
Q6	What data sources are providing observed property O to monitor a distribution network D?

between the Landlsip ontology and electrical grid information services.Whilst various approaches for evaluating an ontology exist, competency questions remain the most common approach [130, 131]. This approach stipulates that an ontology must be able to represent the competency questions using its terminology and answer these questions using the axioms [132]. According to the use case mentioned in Section 4.3, competency questions were developed as shown in Table 4.2. As the Landslip Ontology provides only knowledge of landslide and data sources, the evaluation requires the the harmonisation between the Landslip ontology and electrical grid information services. Here, the result from the Landslip Ontology query will be used as an input to invoke the electrical grid information services.

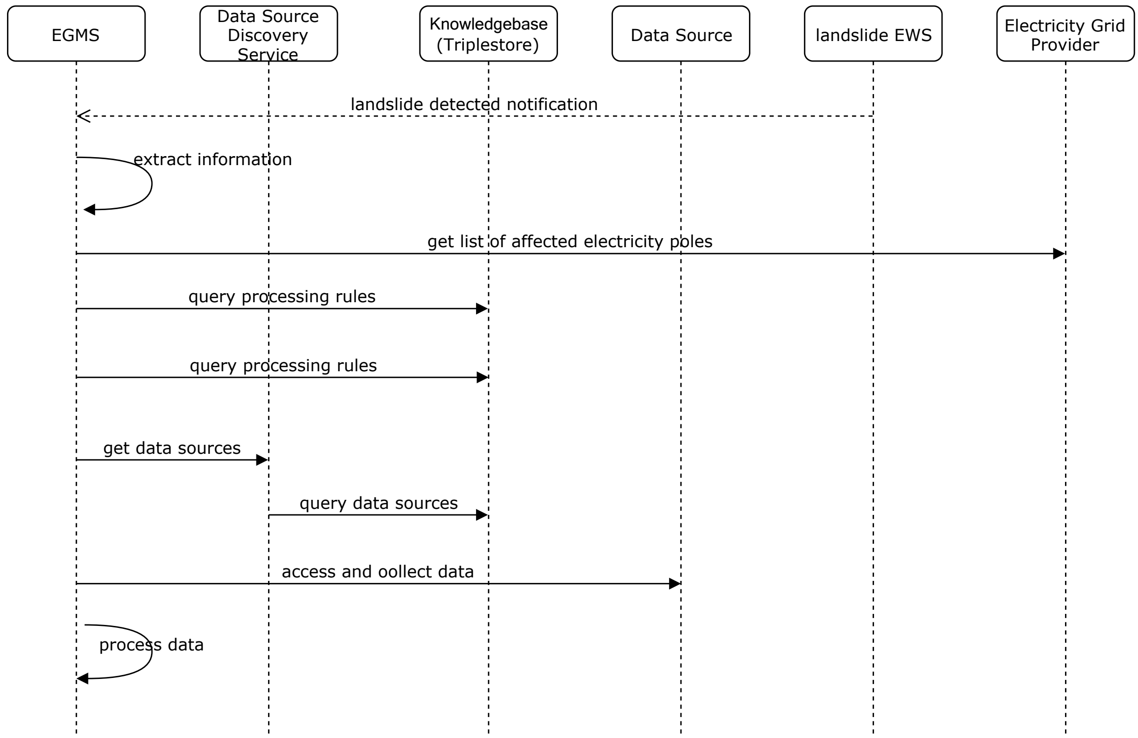

Figure 4.4: A utilisation of knowledge base in electricity grid monitoring system (EGMS).

The evaluation was conducted using a set of synthesised data that represents the use case of landslide hazard mentioned in Section 4.3. We manually added information of natural hazards and EO and ancillary data to our knowledge base. The information includes landslide hazard, hazard triggers, warning signs, EO and ancillary data, and data sources. The data sources include both data sources for landslide hazard and electrical grid system. We performed validation over the dataset using Pellet to check for ontology consistency, concept satisfiability, classification, and realisation. Based on the competency questions, we performed preliminary experiments by querying over the knowledge base.

Q1 Q2

```
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX :
<http://www.semanticweb.org/ncl/ontologies/2018/6/landslip#>

SELECT  ?geo ?p ?value
      WHERE { :landslide_1 :hasGeoLocation ?geo .
                   ?geo ?p ?value
                   FILTER (?p = :bbox)}
```

Q3

```
SELECT  ?hazard ?geo ?p ?value
      WHERE { ?hazard rdf:type :Landslide .
                   ?hazard :hasGeoLocation ?geo .
                   ?geo ?p ?value
                   FILTER (?p = :bbox) }
```

Q4

```
SELECT ?hazard ?geo ?p ?value
      WHERE { :flood_1 :triggers ?hazard .
               ?hazard rdf:type :Landslide .
                   ?hazard :hasGeoLocation ?geo .
                   ?geo ?p ?value
                   FILTER (?p = :bbox)}
```

Q5

```
SELECT    ?observedProperty
WHERE {
      ?observation :isObservationFor :gridnetwork_1 .
      ?observedProperty :isObservedPropertyFor ?observation .}
```

Q6

```
SELECT  ?gridnetwork ?observation ?observedProperty
?dataSource ?metadata ?profile ?p ?value
WHERE {
      ?observation :isObservationFor ?gridnetwork .
      ?observedProperty :isObservedPropertyFor ?observation .
      ?dataSource :isDataSourceFor ?observedProperty .
      ?dataSource :hasDataSourceMetadata ?metadata .
      ?metadata :hasProfile ?profile .
      ?profile ?p ?value .
      VALUES (?gridnetwork) { (:gridnetwork_1) } .
      FILTER (?p != rdf:type)
}
```

Figure 4.5: SPARQL Queries for Competency Questions Q1 to Q6.

gridnetwork	observation	observedProperty	dataSource	metadata	profile	p	value
gridnetwork_1	obs_3	soil_moisture_1	dataSource_1	ds_metadata_1	service_profile_11	where	""@
gridnetwork_1	obs_3	soil_moisture_1	dataSource_1	ds_metadata_1	service_profile_11	dbname	"mysql"@
gridnetwork_1	obs_3	soil_moisture_1	dataSource_1	ds_metadata_1	service_profile_11	table	"NorthSlopHigherRawData"@
gridnetwork_1	obs_3	soil_moisture_1	dataSource_1	ds_metadata_1	service_profile_11	serviceURL	"http://127.0.0.1/rest/moisture"@
gridnetwork_1	obs_3	soil_moisture_1	dataSource_1	ds_metadata_1	service_profile_11	serviceProvider	"Amrita"@
gridnetwork_1	obs_3	soil_moisture_1	dataSource_1	ds_metadata_1	service_profile_11	serviceAdapter	"Rest_adaptor_11"@
gridnetwork_1	obs_3	soil_moisture_1	dataSource_1	ds_metadata_1	service_profile_11	serviceType	"REST"@
gridnetwork_1	obs_3	soil_moisture_1	dataSource_1	ds_metadata_1	service_profile_11	column	"*"@
gridnetwork_1	obs_3	soil_moisture_1	dataSource_1	ds_metadata_1	sensor_profile_1	featureList	"foi_karela_bbox_1"@
gridnetwork_1	obs_3	soil_moisture_1	dataSource_1	ds_metadata_1	sensor_profile_1	eventList	"Landslide"@
gridnetwork_1	obs_3	soil_moisture_1	dataSource_1	ds_metadata_1	sensor_profile_1	sensorType	"in-situ"@
gridnetwork_1	obs_3	soil_moisture_1	dataSource_1	ds_metadata_1	provider_profile_1	providerName	"Amrita"@
gridnetwork_1	obs_3	soil_moisture_1	dataSource_1	ds_metadata_1	obs_profile_11	phenomenonBeginTime	"2004-01-01T00:00:00"^^<http://www.w3
gridnetwork_1	obs_3	soil_moisture_1	dataSource_1	ds_metadata_1	obs_profile_11	phenomenonEndTime	"2018-07-21T00:00:00"^^<http://www.w3
gridnetwork_1	obs_3	soil_moisture_1	dataSource_1	ds_metadata_1	obs_profile_11	featureOfInterest	"foi_karela_bbox_1"@

Figure 4.6: SPARQL Query Output for Competency Question Q6.

In order to write competency questions and to demonstrate that the Landslip Ontology can be applied for electrical grid network monitoring and answer these questions, we used the SPARQL Protocol and RDF Query Language (SPARQL).Using SPARQL, we defined a formal query for each natural language competency question to get answers from the knowledge base. Figures 4.5 and 4.6 show snapshots of the SPARQL queries for *Q2* and *Q6* and the output for the competency question *Q6* on running the query in Protégé [159]. By executing the query based on the competency questions *Q1* -

Q6, we could verify the coverage of the Landslip Ontology. From the results, it can be seen that the ontology is able to identify a list of pylons and poles in an electrical grid network. Furthermore, the ontology can suggest potential data sources and their metadata, which can be used by domain experts to perform timely decision making against the failure of electrical grid network.

4.7 Summary

This chapter demonstrated the application of the Landslip Ontology, an Ontology in the electricity grid domain. Specifically, the *LO* is utilised with external services for electrical grid information services to monitor a failure of the electrical grid network due to a landslide occurrence. The process for harmonising the *LO* and electrical grid information services for an efficient Early Warning System in the electrical grid domain is demonstrated in this chapter. The *LO* supports the EWS for the electrical grid system by enhancing the efficiency of electrical grid network monitoring under landslide hazard. Moreover, the *LO* enables decision-makers to find potential data sources for monitoring. This chapter's contribution is a process for the harmonisation of the knowledge base and electrical grid information services for monitoring of the electrical grid network. This contribution lays at the top layer of the thesis's scope (the EWS and Decision support tools) as an EWS for the electrical grid domain. The next chapter demonstrates an EWS and decision support tool for bi-directional information exchange with social media users for enhanced hazard observation.

5

Social Media Analytics System for Active Hazard Observation

This chapter demonstrates a Social Media Analytics System for Active Hazard Observation (AHOM) that proposed an active way to include the human in the loop of hazard information acquisition from social media. Unlike the state of the art, it supports bi-directional interaction between social media data processing systems and social media users, which leads to the establishment of deeper and more accurate situational awareness of hazard events. The utilisation of Twitter streams and bi-directional information exchange with social media users for enhanced hazard observation is demonstrated in this chapter.

5.1 Introduction

The application of Early warning Systems (EWS) to predict natural hazards and vulnerabilities plays a vital role in preventing loss of life and damage to property. For effective and timely decision-making, EWS requires a strong technical underpinning and sophisticated knowledge of the natural hazards and risk management. Landslides are a commonly occurring natural hazard with global impact and is closely linked to many other natural hazards such as earthquakes, storms, flooding and volcanic eruptions. Predicting individual occurrences of landslide events is complex as it depends upon many local factors, variables and anthropogenic input. For predicting and monitoring landslides, decision makers use scientific models to analyse Earth Observation (EO) data from satellites and ancillary data produced by Internet-of-Things (IoT) sensors deployed in landslide-prone areas. In addition, the ancillary data includes sensor data, social media data and data from other sources which are essential for the prediction and monitoring of hazard events. Such EO and sensor data used for the analysis are usually obtained from multiple and heterogeneous data sources. Furthermore, through social media channels (Facebook, Twitter, Instagram etc.) the general public can also contribute to landslide monitoring by reporting observations that could be warning signs for landslides. However, decision-makers need to verify the detected events reported from social media by analysing sensor or other corroborating data from the area of interest. Hence, EO and data representing the event in the area of interest is essential for the verification.

Effective response [160–162] to crises and hazard events such as landslides, floods, fires, hurricanes, tsunamis, and man-made hazards is dependent on the availability of historical data as well as on the effective real-time integration and utilization of data streaming from social media feeds (such as Facebook, Twitter, and Weibo). However, the existing social media data processing and/or acquisition methods are solely based on Machine Learning (ML) and Natural Language Processing (NLP) classifiers, while lacking the capability to include the human experts who can contribute to the data collection and processing loop in the real-time (i.e., while hazard event is unfolding).

This leads to following drawback: the information extracted by pre-defined ML and NLP classifiers may miss the information about antecedent hazards that leads to full-fledged disaster. To illustrate this shortcoming, let us consider the following real-world example from our research project (*http://www.landslip.org/*) where a twitter user posts a message to report hazards such as leaning utility poles, trees cracking, or collapsed road beds in their village. Given these events, the social media data analysis system can potentially predict the likelihood of the occurrence of a more serious hazard, such as landslide, if some further contextual information can be collected from the Twitter user (human in the loop) such as the location of the village and whether there have been rainfall and flooding events in the past few hours/days.

In order to study how the social media users and human experts can more *actively* contribute to data analysis (collection and processing) loop for natural hazard response and planning, we develop the AHOM system, which has the following unique features that differentiate it from previous social media data analysis systems.

Human in the loop AI system To enable bi-directional interaction with the social media users, AHOM uses a novel ontology, namely the Landslip Ontology (*LO*) as presented in Chapter 3, that abstracts the landslide experts' knowledge showing the relations among landslip, landslip warning signs, and other potential occurrence of hazards. This enables a new generation of interaction between data processing systems and social media users, based on various "what if" scenarios modeled by the ontology-based data integration and querying engine. This way the exhaust of social media is used to develop more deep situational awareness of disaster events.

Integrative data management pipeline As a proof of concept, we develop a social media data processing pipeline (systems) which comprises of a Stream processing engine (Kafka), NoSQL database (Elastic search), Natural Language Processing (NLP) engine (spaCy[1]), and a novel Landslip Ontology for data integration and querying (an Ontology-based data integration and querying (Triplestore) engine). While Kafka and Elastic search are capable of accommodating real-time and historical social-media feeds respectively. spaCy is used and interacted with Kafka stream processing APIs, passively extracting information from social media platforms in real-time. *LO*, which is hosted in an ontology database, Triplestore, enables the generation of automatic and interactive follow-up questions based on various "what if" scenarios modeled by the ontology. A running version of our system is available at GitHub[2], and the live demo is available at here[3].

The rest of the chapter is organised as follows: related work is discussed next in Section 5.2, followed by discussing the development of context-based knowledge discovery and querying for social media data which is a knowledge-based for AHOM in Section 5.3. The overall design of the AHOM and its implementation are discussed in Section 5.4 and 5.5, respectively. Finally, the summary of this chapter is presented in Section 5.6.

5.2 Related Work

This section discusses related work on data integration and analytics, and natural language processing in multi-hazard EWS.

[1] μ`https://spacy.io/`
[2] μ`https://github.com/ncl-iot-team/active-hazard-monitoring`
[3] μ`https://bit.ly/2V9MkG4`

5.2.1 *Data Integration and Analytics In multi-hazard Early Warning Systems*

EWS have played a significant role in natural hazard management to minimise loss of life and damage to property. Additionally, several modern multi-hazard EWS take advantage of various type of data sources, including remote sensing satellites, IoT sensors, and social media. Due to the heterogeneity of the data sources, data integration becomes a vital part of EWS to provide high-quality data for the effective prediction of hazard events. Several works on data integration and analysis for multi-hazard have been proposed in the literature. A data analytic framework for Twitter data was proposed in [64] to identify twitter messages that are related to a particular type of disaster (e.g. earthquake, flood, and wildfire). Several methods, including matching-based and learning-based, to identify relevant tweets, are also evaluated. In [109], the authors describe a study on the identification of peatland fires and haze events in Sumatra Island, Indonesia, by using social media data. A data classification algorithm is applied to analyse the tweets, and the outcomes are verified by using hot spot and air quality data from NASA satellite imagery. The authors in [57] propose an information infrastructure for timely delivery of social media and crowd-sourcing message (from Ushahidi platform) to potentially responsible departments during the Haiti earthquake in 2010. A data classification algorithm is used to provide an automatic classification mechanism over the messages. A decision support system that integrates crowd-sourcing data with Wireless Sensor Networks (WSN) to widen the coverage of the monitoring area for flood risk management in Brazil is proposed in [110]. The Open Geospatial Consortium (OGC) standards are used in the research to aid in the integration of the crowd-sourced data.

5.2.2 *Natural Language Processing*

Natural Language Processing (NLP) is a set of information engineering techniques which enables computers to process and make sense of human (natural) languages. NLP technique has evolved from complex handwritten rules to models trained using machine learning. Earlier machine learning techniques like decision trees[163, 164] generated rules similar to handwritten ones, using machine learning. The application of NLP in this work is to extract useful information from the natural language and to classify the content into different topics of interest. Language modelling techniques apply probability distribution over a sequence of words. Unigram, n-gram [165, 166], Exponential and Neural networks [167–169] are the main types of language models in use. Recent studies promise high accuracy in classifying natural language using a neural network. A unified architecture for NLP using deep learning technique has been introduced in work [170] by NEC Labs. In this work, the input sentence can be processed to perform part-of-speech tagging, chunking, named entity tags, semantic roles etc. using a language model and CNN. A study [171] at New York University reveals a series of experiments using Convolutional Neural Network (CNN) which is trained on a pre-trained model of word vectors for sentence classification in which the model showed significant improvement in performance in several NLP tasks. Over the years several open-source NLP projects like NLTK[172], CoreNLP[173], Spacy[174], GATE[83] etc. gained interest of both academia and industry. While these methods and tools support natural language processing, building knowledge from natural language pose several challenges.

5.3 Development of Context-based Knowledge Discovery and Querying for Social Media Data

5.3.1 *Knowledge-base Development*

Earth Observation (EO) and ancillary data provided by multiple data sources contribute to the high variety of characteristics of data sources in EWS. In addition, such data sources differ in terms of: (i) data type — the different types of EO and ancillary data used in a particular analysis, e.g. soil properties, temperature, humidity; (ii) data storage — the difference in methods for collecting and organising data, e.g. RDBMS, NoSQL database, and distributed file system; and (iii) data access — data sources are accessible by different methods, ranging from direct access through data stores (e.g. JDBC) to standard Web Services (e.g. OGC, SOAP, Restful). These differences make the discovery, access and integration of data in EWS quite challenging. A formal semantic representation of the data sources, domain knowledge about natural hazards and the relationship between them can help address the challenges arising from such differences and enable data integration and analysis.

The knowledge base for data integration and analysis is designed and developed based on the principles of Semantic Web Technology [112]. A core component of the knowledge base is an ontology, which can be can be defined as "a formal, explicit specification of a shared conceptualisation" [175]. That is, an ontology models the agreed knowledge about the real world through explicitly defined concepts and constraints on them that are machine readable. In this work, we have developed an ontology, namely the Landslip Ontology, which is based on OWL 2 [176]. The Landslip Ontology contains knowledge on the relationships between landslides and data sources for EO and other data. Additionally, it also contains the domain knowledge for describing the interaction of landslide event to other hazards and the warning signs, which can be precursors to landslides. Figure 5.1 shows the steps of the Landslip ontology development. The primary knowledge sources for designing the ontology came from interviews with four scientists and experts in landslide hazard management with an average of 10 years of experience between them. Besides, publications [122], [123] and standard specifications [28, 30, 124–126] involving multi-hazards and geospatial data models were also used as knowledge sources to design the ontology. A scenario-based approach [119] was used to define a narrative representing expected uses of an EWS in the domain of interest from the viewpoint of both domain experts as well as ontology developers. The scenario helps to define the scope of the domain ontology to be designed and frame competency questions to model the domain knowledge as well as for evaluating the ontology [130, 131]. The Landslip ontology is developed and imported to a triple store as a model for building the knowledge base for Landslide multi-hazard EWS. Finally, a knowledge API is developed, which provides the consumer with an access point to the knowledge base.

The Landslip Ontology which has been developed in Chapter 1 is used to construct the knowledge base. The ontology is comprised of two modules: (i) *Landslide Hazard Ontology* – defines concepts about landslides and the interaction to other hazards and corresponding warning signs; and (ii) *Data Sources Ontology* – defines concepts about observation and data sources for landslide hazard risk assessment. The Landslip

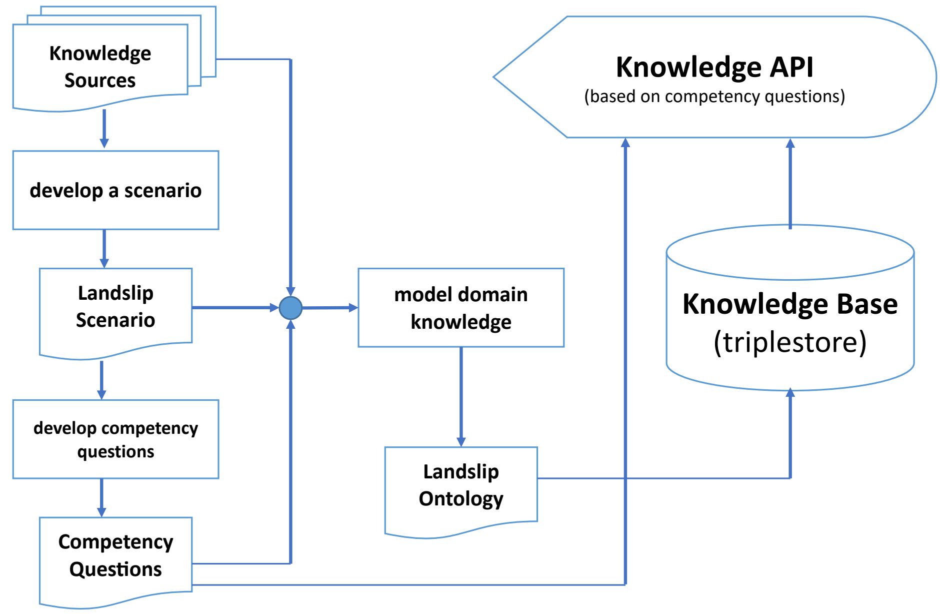

Figure 5.1: Landslip Knowledge base development process

ontology reuses the SSN ontology as well as terminologies defined in OGC standards (e.g. Observation and Measurement [124], SensorML [125] and SOS [126]).

The ontology was evaluated for consistency and correctness through competency questions and a set of synthesised data that represent the use case(s) of landslide hazard. The competency questions were defined in SPARQL [177] and the knowledge base queried for answers. An example competency questions is *"What are data sources and their metadata to observe a set of hazards H1, H2, ... Hn"* and the corresponding SPARQL query to answer this question is defined as follows:

Listing 5.1: SPARQL example for data sources selection

```
SELECT  ?hazard ?observation
    ?observedProperty
    ?dataSource ?metadata
    ?profile ?predicate ?_value
WHERE {
    ?observation :isObservationFor
    ?hazard .
    ?observedProperty
    :isObservedPropertyFor ?observation .
    ?dataSource :isDataSourceFor
    ?observedProperty .
    ?dataSource :hasDataSourceMetadata
    ?metadata . ?metadata :hasProfile
    ?profile . ?profile ?predicate ?_value .
```

```
    VALUES (?hazard)
        { (:flood_1) (:landslide_1) } .
    FILTER (?p != rdf:type)
}
```

5.3.2 Populating the Knowledge Base from Social Media Content

The Landslip Ontology is a conceptual model that formally represents domain knowledge about landslides captured from domain experts of natural hazards management. Th ontology consists of concepts and relationships but does not model concrete objects or *named individuals*, that represent actual events of landslides. With the emergence of social media as a potential resource to build the domain knowledge, social media contents to represent actual events of landslides are dynamically instantiated within the ontology. In order to do this however, sophisticated techniques are required to understand the context of the social media content and extract information from the content to create individuals based on the conceptual model.

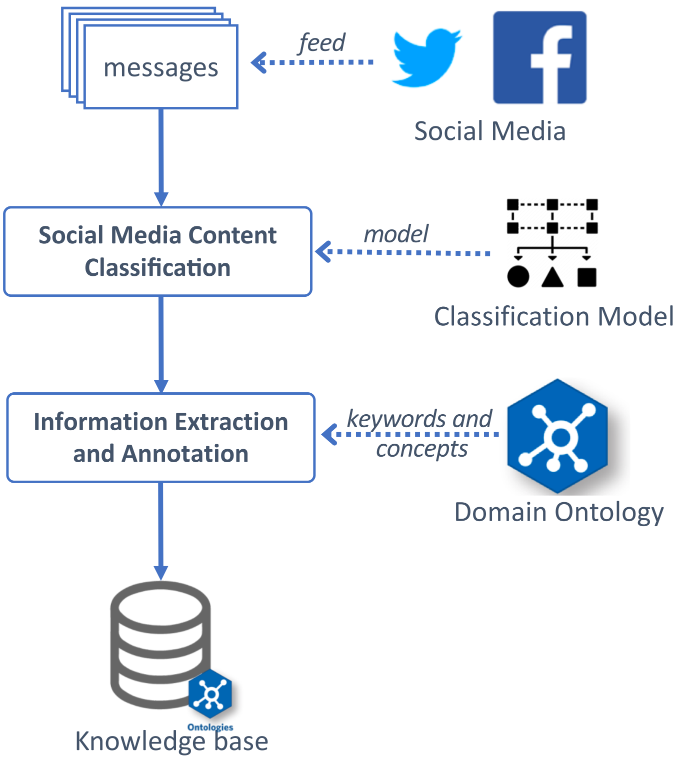

Figure 5.2: Process of populating the Knowledge Base from social media content

Figure 5.2 shows the process of populating the knowledge base from social media content to facilitate landslide-related knowledge discovery in an EWS. Historical data related to past events of landslides collected from social media platforms are also added to the knowledge base. The knowledge base thus consists of a set of synthesised data for hazards (landslides, flood), warning signs (e.g. leaning light pole, blocked road), and EO and ancillary data (e.g. water potential, moisture and temperature). Due to the wide variety of information provided in each text from the collected social media data, classification of the text into the topics of interest and extraction of useful information from the content are required. One of the most critical challenges while dealing with user-generated content is to capture the semantics of the content using Natural Language Processing (NLP)/ Natural Language Understanding (NLU) techniques. Our prototype system is designed with two modules to achieve this task: A data classification/topic detection module for social media content classification and a data extraction module to extract useful information, which can be used for instantiating objects in the ontology. The modules are described next.

1. **Social Media Content Classification** — The first step to process the user text is to identify the hazard which the user is referring to on social media. Recurrent Neural Network (RNN) [178] such as Long Short Term Memory networks (LSTM) and Convolutional Neural Network (CNN) are widely used in text classification. In this work, we have used CNN, a deep learning technique, [171] to perform text classification. A model, which is custom trained using weather related text is used. The model training and inference processes are explained in the following sections.

    - **Model** — A classification hierarchy, as shown in Figure 5.3, has been defined for our prototype system. A model is trained using 1000 user-generated texts related to hazards, which are marked for different hazard events and warning signs (for example, Events: Flood, Heavy Rainfall, Snow etc.; Warning Signs: Leaning Light Pole, Water Discolouration etc.). Each class has around 200 records. TensorFlow [179], an open source library, was used for data preparation, training and inference. The model is similar to the one proposed by Kim Yoon in his work Convolutional Neural Networks for Sentence Classification [171], which achieved good classification performance for different text classification tasks like sentiment analysis and is a standard baseline for new text classification methods. The model consists of a word embedding layer, which maps vocabulary word indices to lower dimensional vector spaces. The convolutional layer calculates convolutions over the embedded word vectors using different filter sizes as each convolution produces tensors of different shapes followed by max-pooling, which is a sampling-based discretization process. These vectors are later merged to form a large feature vector. Full details of the CNN layers and training process are beyond the scope of this work.

- **Inference** — Every text message received by the Landslip agent is passed to the classifier, which outputs the hazard event and any warning sign mentioned in the text. This step enables the system to understand the topic from the user-generated content. Data classification for this system is a two-step process involving the classification of hazards and warning signs. In the first step, the classifier tags the message whether a hazard or warning sign is present in the text. The second step involves two classifiers, one for classifying the type of hazard and the second for categorising the kind of warning sign as per the classification hierarchy.In some cases, a message may contain information about both hazard and warning sign. In such a scenario, the system passes this message to both classifiers. In the inference step, the result is attached as metadata to the input text.

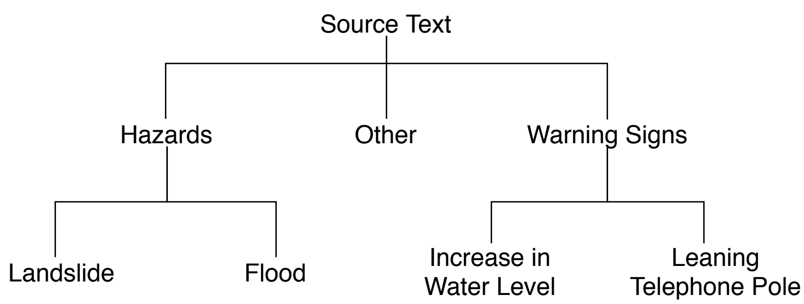

Figure 5.3: Data Classification Hierarchy

2. **Information Extraction and Annotation** — In this step, a scenario is de-

veloped using information about the situation from user-provided text. NLP techniques, namely Part of Speech (PoS) tagging and Named Entity Recognition (NER) are used to extract useful information from the text. Pre-trained NLP models for the English language recognise Geo-location and affected entities (for example, Road, Building, Electric Pole, etc.). The English language model,a multi-task CNN trained on OntoNotes[180], with GloVe[181] vectors trained on Common Crawl[182] is used in this work. This model is built for assigning word vectors, context-specific token vectors, POS tags, dependency parse and named entity recognition.

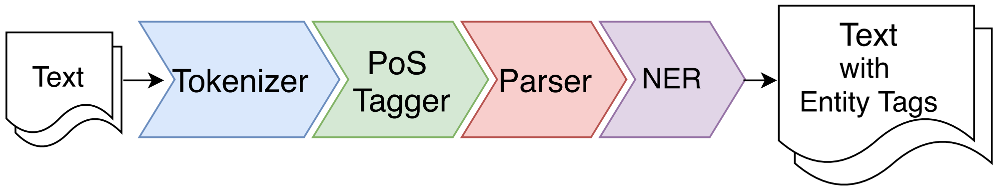

Figure 5.4: NLP Pipeline for Named Entity Recognition

As mentioned in Section 5.2, named entities are extracted from the user-generated content using NER. An NLP tool called Spacy [174] was used to perform the series of tasks required to perform NER. The processing pipeline consists of a tokenizer, PoS tagger, Parser, and NER. The tokenizer tokenizes sentences into words for which a PoS tag is attached based on the sentence structure. Then the parser performs a dependency parsing of the sentence, which represents its grammatical structure and defines the relationship between words. This step is followed by the NER phase, which identifies the type of entity such as geo-location, person, organisation, physical object, date, time, building/infrastructure etc. This model and pipeline gave a largely accurate prediction of the type of the entity from the noun words tagged from the PoS tagging phase. Figure 5.5 shows an example of a sentence being processed and labelled.

The entity tags are attached to the original data as metadata for storage and indexing. This extracted information is instantiated as objects based on the concepts defined in the ontology and stored in the knowledge base.

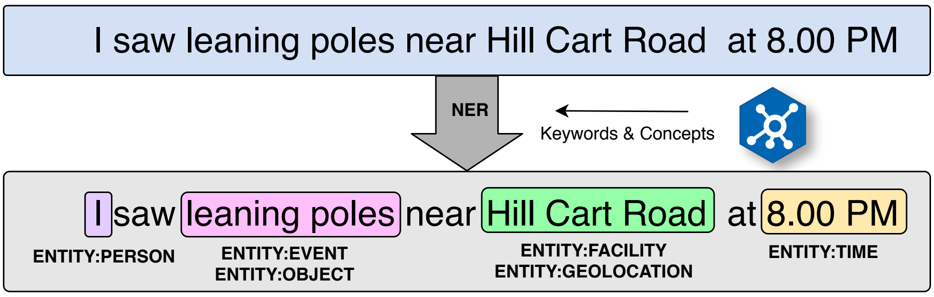

Figure 5.5: Named Entity Recognition Example

5.4 System Overview

5.4.1 System architecture

Essentially, AHOM is a loosely-coupled run-time system that allows the human experts to participate in the information acquisition. As shown in Figure 5.6, there are four main components in AHOM: **Storage system**, **Stream processing**, **Human machine interaction** and **AHOM API**. First, social media data streams from various sources are injected into the **Storage system** where Apache Kafka is used to consume the input streams. Next, the **Stream processing** component applies a set of machine learning models (mainly NLP models) to process the injected streams and the outputs of the models are stored in an Elastic Search Database for further queries. The machine learning models are containerized as microservices that are easy to plug-in and plug-off to AHOM via the Kafka publish/subscribe message system. The details of each machine learning model are illustrated in §5.4.2. When the pre-defined antecedent hazard events are detected and the number of the detected events exceeds a threshold, the **Human machine interaction** framework utilizes the *LO* that we developed in Chapter 3 to strengthen the awareness of the situation that domain experts and/or the hazard managers may seek further information from the social media users based on automated follow-up questions generated by our system (see §5.4.3). Thus, the higher resolution information particularly interested by human experts can be collected from users' replies with a few iterations. Finally, **AHOM API** provides a dashboard to visualize the information of the detected events, extracted from massive social media data.

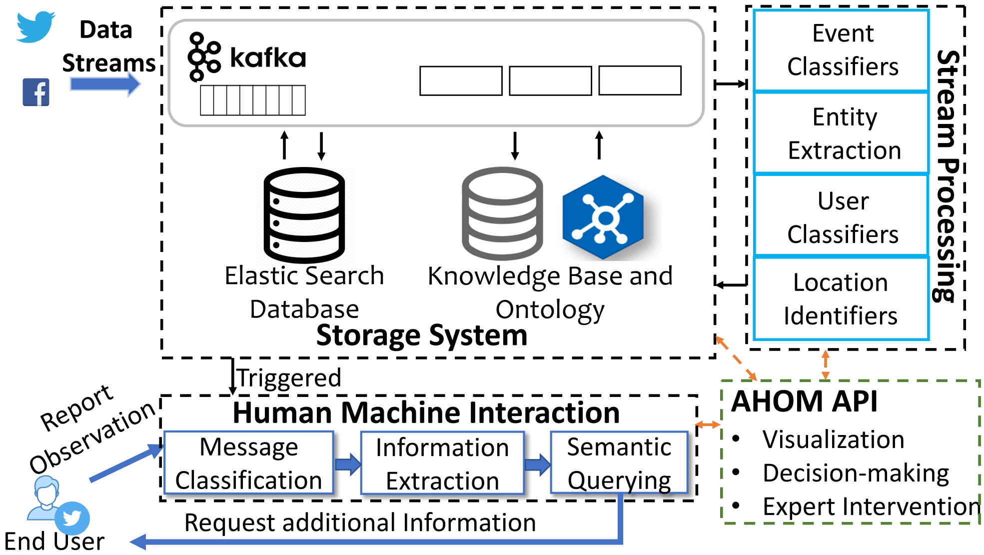

Figure 5.6: AHOM architecture

5.4.2 Data processing pipeline

In this section, we discuss the details of each machine learning model and its execution workflow. First, the data streams are fed to *User Classifiers* that identify who is posting this message. In this demo, we only consider the classification of two type of users: one is an official account such as Meteorological Office; the other is for normal users. In general, the information provided by official accounts is more reliable compared with normal users, but the normal users can provide richer information. Since Twitter and Facebook already provide the user information, we use the Twitter handle to classify the users based on a dictionary. Next, the *Event Classifiers* are used to identify whether a message relates to landslide hazard and antecedent hazard events (i.e., warning signs). These classifiers are developed by using spaCy a NLP framework. To this end, we collect a small amount of labeled data and then use the dataset to retrain spaCy's convolutional neural network (CNN) models [183, 184] to improve the accuracy of detecting events (these models will also be used to support the *Human machine interaction* in §5.4.3). Obtaining the geolocation information from the mentions in the message is also essential for analyzing landslides. This task consists of two steps: *Information Extraction* and *Location Identification*. The *Information Extraction* component aims to extract the named entities of a message via `en_core_web_lg` model (available on spaCy), a CNN model trained on OntoNotes[180]. These extracted named entities are the inputs of the *Location Identification* that classifies the named entities as including geolocation information or not. If yes, these geo-names are converted to geo-coordinates by OpenStreetMap (OSM) datasets [4] using geocoding method[5]. A single geo-name may have multiple entries in the OSM dataset. For this demo purpose we take the first entry from the OSM dataset. As future work we will seed the module with the location of interest. Finally, all the outputs are published to Kafka and stored in Elastic Search Database for further analysis.

- **Scalability** — Data processing pipeline runs on Apache Kafka[185] which provides parallelism using data partitioning and consumer groups. In the demonstration system, for the Twitter topic, we used 3 partitions with the Twitter handle as the partition key. The system can be scaled easily by increasing the number of partitions. The number of workers for each processing step in the data processing pipeline can be scaled up to the number of partitions for the given topic. Kafka cluster consists of multiple brokers and a degree of replication ensuring scalability and resilience.

- **NLP model accuracy evaluation** — In this demo, we trained our models with a small dataset of 5000 tweets. The data is collected using Twitter streaming API using keywords landslide and flood. We manually labeled the dataset in two regards: 1) we labeled tweets related to landslide and its antecedent hazard events; 2) for each tweet, we extracted the geolocation named entities e.g., country, state, street etc. The trained models have very good accuracy, where the *Event detection* model only for landslide hazard achieves 92% accuracy and the *Geolocation extraction* model extracts geolocation named entities with 87% accuracy.

5.4.3 Human machine interaction

Based on input obtained from domain experts, *LO* is a knowledge model representing information about natural hazards, including hazards such as landslide, flood, rockfall, and earthquake; hazard warning signs such as rainfall and leaning utility poles; and

[4] μhttps://wiki.openstreetmap.org/wiki/Downloading_data
[5] μhttps://nominatim.org/

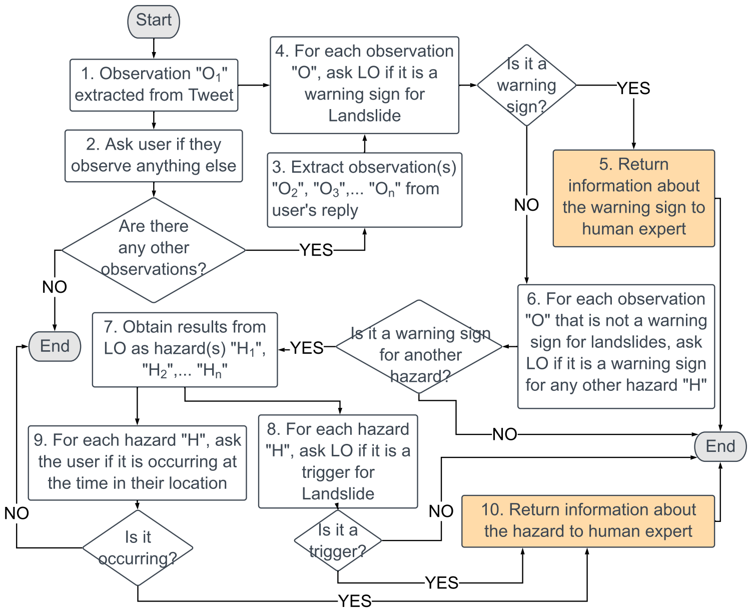

Figure 5.7: Human machine interaction framework and its execution flow

observers such as social media users. The actual events detected or collected from social media can be mapped to the *LO*. *LO*, together with the actual events, acts as a knowledge base that can be queried and used for decision making. Based on the mapping, the "hot zone" of the knowledge base can be identified, which represents information that is of special interested to the decision makers and human experts. Next, the missing knowledge in this "hot zone" can be used to generate the follow-up questions for social media users to obtain further contextual information vital for decision making.

To this end, we develop a framework of human-machine interaction based on *LO* and the knowledge base discussed above and shown in Figure 5.6. This framework utilizes machine learning techniques implemented in the stream processing component to understand and extract landslide-related information from a large number of social media users and then proactively acquire the missing landslide related information from specific users. The framework consists of three main steps.

- **Message Classification** — event classifiers developed in the stream processing component are utilized to collect the tweets related to the context of landslide hazard and landslide warning signs. In addition, the classification models are trained based on the taxonomy of natural hazards and warning signs defined in the *LO*.

- **Information Extraction** — allows the machine to extract actual contents from a tweet using the NLP techniques. For example, we use the location identifier to extract the geolocation information. The essential information for landslide events extracted from the social media includes location, time and warning signs. This information will be used in the next process to investigate missing pieces of landslide related information.

- **Semantic Querying** — provides more comprehensive information about events for decision making. In AHOM, we present three types of information using NLP techniques and ontology based querying: i) essential information about an event (such as location, time, and observation of events), ii) warning signs to indicate potential occurrence of landslides, and iii) potential occurrence of other hazards that could act as triggers for landslide. The aim is to provide as much relevant information as possible to support human experts in making informed decisions. AHOM uses *LO*, to provide information about warning signs and other related hazards to enable human experts to predict the likelihood of the occurrence of landslides with increasing certainty. As can be seen from Figure 5.7, vital information about warning signs and related hazards is obtained from *LO*.

We use the following example to explain the process depicted in Figure 5.7. Using NLP techniques, it can be determined that a user has tweeted about a *leaning pole* (step 1 in Figure 5.7). The system then seeks further clarification from the user by asking if they have noticed or observed anything else such as *rainfall* (step 2). The observations from the user's replies are extracted (step 3) and fed into *LO* to determine the relation of these observations (*leaning pole* and *rainfall*) with landslide and also to extract further relevant information. For each observation extracted from the tweet and the user's subsequent replies, SPARQL query [177] is used to ask the *LO* whether it is an indicator (warning sign) for landslide (natural hazard) (step 4). The shortened query is presented below:

Listing 5.2: SPARQL example for warning sign indicator

```
ASK {?observation:isWarningSignFor?landslide}
```

If the answer to the query is 'true' (yes), then the information about the warning sign is returned to the human expert (step 5). For example, in this case, the *leaning pole* is a warning sign for landslide as modeled in the *LO* and hence the answer would be true. If, however, the answer obtained is 'false' (no) then for each observation in question, *LO* is further queried to determine if the observation is an indicator for any other hazard(s). For example, since *rainfall* is not by itself an indicator of landslides according to the relationship modeling in *LO*, the answer obtained for the above query for *rainfall* would be 'false'. In this case, the following question is asked of *LO* to determine if rainfall is an indicator of any other hazard (step 6):

Listing 5.3: SPARQL example for hazard indicator using a warning sign

```
SELECT ?hazard
WHERE {:Rainfall:isWarningSignFor?hazard}
```

If the query yields an answer (step 7), which in this case would be *flood* and *rockfall*, then follow-on questions are asked of the *LO* about the hazards identified. In this example, the questions would be whether *flood* and/or *rockfall* can trigger landslides (step 8):

Listing 5.4: SPARQL example for examining a landslide trigger by other hazards

```
ASK {:Flood:triggers?landslide}
#and
ASK {:Rockfall:triggers?landslide}
```

If the answer is 'true', then the information about the type of hazard is returned to the human expert (step 10). In parallel with step 8, the user is also asked whether the hazard(s) (obtained as answer to the query in step 7, which in this example are *flood* and *rockfall*) are occurring or have occurred in their location (step 9). If the answer is 'true', then this information is presented to the human expert (step 10). Thus, in this example where the hazard of interest is landslide, in addition to the location and time of each observation, the following critical information is presented to the human expert with the help of *LO*: i) *Leaning pole* has been observed, which is a warning sign for the occurrence of landslides and ii) *rainfall* has been observed, which is a warning sign for *flood*, which in turn can trigger landslides. Further, the user has confirmed there is *flooding* in their location (see §5.5.2).

The human expert thus receives information about possible indicators or warning signs for a natural hazard as well as information about other hazards that may eventually lead to the hazard in question. This example demonstrates how rich semantic querying with *LO* can help to identify further relevant information, which may not have been otherwise directly available, thereby providing more comprehensive knowledge that is essential for informed decision making.

5.5 Implementation and Deployment

This section gives a demonstration of AHOM, with experiments conducted with Twitter via its commercial APIs.

5.5.1 System deployment

The core parts of AHOM (i.e., *Stream processing* and *Semantic query*) were implemented in the Python language, and deployed on an Ubuntu server with 20 cores (Intel(R) Xeon(R) Silver 4114 CPU @ 2.20GHz) and 64 GB memory. The *Storage System* was deployed on another server with the same configuration. The Kafka cluster was set up using Apache Kafka 2.12, the *Elastic Search Database* and *Knowledge Base* were developed by using Elastic search 7.6 and Stardog respectively.

5.5.2 Execution sequence

This subsection illustrates the execution sequence of our AHOM to demonstrate its ease-of-use and support for disaster management (see Figure 5.8).

1. **Stream monitoring** — The Twitter stream APIs are used to pull tweets, using a set of subscribed keywords generated by landslide experts, e.g., landslide, landslip and land movement etc [186] (see Step 1). Then, the user accounts are identified in Step 2. Step 3 leverages the trained NLP models to classify landslide related events. If the tweet is related, in Step 4 and 5, the geolocation name entities are extracted (if they exist), via the models discussed in §5.4.2. Finally, all processing results are stored in *Elastic Search Database* (Step 6) and visualized

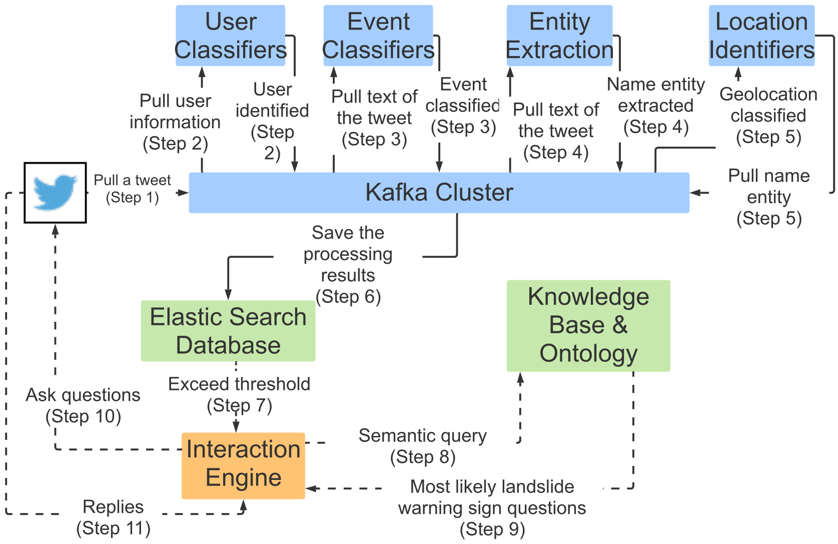

Figure 5.8: Execution pipeline

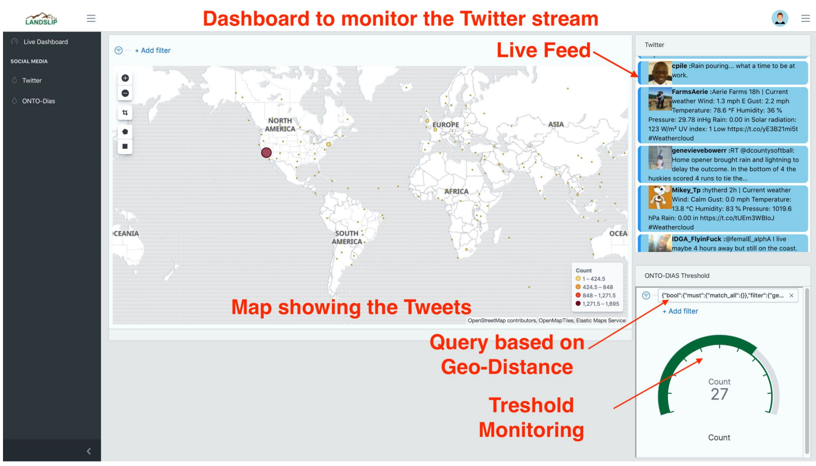

Figure 5.9: Screenshot of stream monitoring

in real-time as shown in Figure 5.9, where the size of cycle represents the number of tweets i.e., the bigger size the higher the number of tweets. Moreover, our system can monitor and visualize over a defined geographical area. For example, the right side of Figure 5.9) shows the number of tweets related to landslide within a 200KM radius of Newcastle upon Tyne. If the number of tweets exceeds the predefined threshold (i.e., 50 in this demo), the system will randomly select some tweets and ask some questions (see Step 7).

2. **Semantic query** — Figure 5.10 is a snapshot of the human-machine interaction system in action. The series of questions posed here for the user and the reasoning

Figure 5.10: Screenshot of Twitter interaction

behind them is explained in *Semantic query* in §5.4.3 with reference to Figure 5.7. The process described in Figure 5.7 corresponds to steps 8 - 11 in Figure 5.8, which gives the big picture view of the execution pipeline. These steps in Figure 5.8 represent the bidirectional interaction of the interaction engine with *LO* (steps 8 and 9) and of the engine with the social media user (steps 10 and 11).

5.6 Summary

A comprehensive set of EO and ancillary data from multiple and heterogeneous data sources is essential for an effective early warning system (EWS). Social media and other unstructured data are increasingly important for EWSs in natural hazard management by augmenting traditional data sources used by, for example, landslide scientists. This chapter shows that the proposed prototype, AHOM, is able to process massive amounts of data from social media to provide meaningful content to emergency responders, planners and local and national decision makers. Additional benefits accrue by enhancing the completeness of a dataset through automated question-based information gathering which in turn improves perceived trust in and reliability of the data collected.

6 Conclusion

6.1 Thesis Summary

This thesis's research works were motivated by modern Early Warning System (EWS) requirements for analysing high volume-velocity-variety Earth Observation (EO) and ancillary data to monitor and predict the occurrence of natural hazards for efficient urban area management. Such time-series data are produced by a variety of IoT sensors deployed in a natural hazard-prone area. Moreover, social media's emergence allows people to act as human sensors to report warning signs or any incidents related to natural hazards. This thesis mainly addresses the problems of big data integration and analysis for urban hazard management, especially the discovery of a wide variety of time-series data sources, and the organisation and construction of a knowledge base of hazard-related social media data. Here, the landslide hazard domain was selected as a case study of a natural hazard to investigate the techniques for time-series data integration. In particular, this is because landslides are a common form of natural hazard with global impact and are closely linked with various other natural hazards such as storms, earthquakes, flooding and volcanic eruptions.

The research problems were addressed by utilising knowledge representation and reasoning techniques to capture knowledge from landslide experts and develop machine-understandable ontology and knowledge base to facilitate data integration for landslide EWS. Here, the Landslip Ontology was developed to represent the knowledge of the landslide domain and provide a knowledge base that establishes relationships between landslide hazard, social media, and time-series data sources. The major knowledge sources for the *LO* development are from interviews of landslide experts who are members of the Landslip project. The evaluation of the *LO* was conducted based on the domain experts' competency questions to verify the coverage of the Landslip Ontology and its application in a landslide early warning. Besides, the *LO* is a key contribution of the thesis to facilitate data sources discovery and the organisation and construction of a social media knowledge base to support landslide decision-making.

The *LO* knowledge base was adopted to design an ontology-based process to discover landslide-induced emergencies in the electrical grid. The process harmonises the knowledge base and electrical grid information services to monitor electricity grid networks and identify parts of the electrical grid infrastructure potentially vulnerable to a detected landslide. The evaluation was performed by verifying the coverage of the ontology in harmonising with the electrical grid information services based on competency questions.

Finally, a data integration and analysis system for active hazard observation using social media was developed to enable bi-directional information exchange with social media users for enhanced hazard observation. The system utilises the *LO* and knowledge base populated from social media data to detect the potential occurrence of landslide events and generate questions for communicating with social media users. As a result, the system helps decision-makers to obtain missing pieces of information from potential social media users to enable more accurate decision-making.

6.1.1 Limitations

The Landslip Ontology provides a formal knowledge base of landslide domain concepts to enable time-series data integration in urban hazard management. The *LO* and knowledge base can answer decision-makers questions. For example, is a leaning telephone pole an indicator for a landslide? However, the questions to be answered by the *LO* are limited by the competency questions, which are defined by landslide experts from the Landslip project. In addition, domain experts working in another landslide-prone area may have different experiences, expertise, and applications in landslide hazard management. Based on this, the *LO* could be improved by collaborating with other domain experts to answer more questions.

The current version of the active hazard observation system using social media is a research-level prototype that contributes to the Landslip project. The automated communication between the system and social media users was implemented based on the Twitter API with many Automation rules[1]. Thus, the system is only tested and demonstrated within a closed group of the Landslip project. It needs to be compiled with the Twitter rules, especially the sufficient consent to take automated actions through Twitter users, to be deployed at the production level and provided to the public.

6.1.2 Contributions

The research work in this thesis contributes to big data integration and analysis for urban hazard management at several levels, ranging from conceptual model to concrete applications. The main contribution of this research is the Landslip Ontology (*LO*), a formal model that represents the knowledge of the landslide domain and provides a knowledge base that establishes relationships between landslide hazard, social media, and time-series data sources. Moreover, the discovery of time-series data sources and the detection of social media events for the landslide Early Warning System are the other contributions of this research to support decision-making in landslide hazard management. The active hazard observation system using social media contributes to enabling information enrichment from social media to enhance decision-making accuracy. Finally, this research has also contributed to the Landslip project, a Natural Environment Research Council (NERC) funded project that aims to reduce the impacts of landslide multi-hazards in India.

6.2 Future Research Directions

Directions for further research in big data integration and analysis techniques as outlined in this thesis can be seen in the following three fields.

6.2.1 Ontology-based data integration

As mentioned in the discussion on the limitation of the research work in this thesis, there are many research opportunities to enhance the efficiency and capability of the

[1]https://help.twitter.com/en/rules-and-policies/twitter-automation

ontology-based data integration techniques for urban hazard management. The collaboration with domain experts from other institutions to develop more competency questions is a significant key to improve the capability of the Landslip Ontology to answer more questions for urban hazard management.

Current urban hazard knowledge proposed in this thesis is driven by the Landslip Ontology, where the warning signs and hazard interactions can indicate the potential of landslide occurrence. Thus, the Landslip Ontology can be expanded to form an urban hazard ontology which is the combination of supplementary knowledge of other multi-hazards (e.g. the warning sign of other hazards). However, it is necessary to collaborate with domain experts of the hazard of interest to develop the urban hazard ontology.

The research work on knowledge representation and reasoning techniques, which is a field of Artificial Intelligence, for emergency management can be a long term research direction achieved from this thesis. Additionally, cross-domain ontology integration is an opportunity to adopt the Landslip Ontology in other emergency management applications. For example, the Landslip Ontology integrated with Electric Vehicle (EV) related ontology can help the electric energy planing in an urban area that has experienced an urban hazard. Additionally, the EVs with power batteries are considered as a mobile power supply that can contribute their electric power to hospitals, emergency shelters, and public transport etc., during an emergency situation. A cross-domain ontology that captures knowledge from domain experts could help decision-makers in planning and communicating with EV owners.

6.2.2 Cloud-based Risk Analytics Framework for Emergency Management

The following research activities within the context of a cloud-based risk analytics framework for emergency management can be the future direction of the work described in the thesis.

- Algorithmic techniques for urban risk analytics that support storage, classification, and event detection over data obtained from multiple sources, both in real-time (such as data emitted by wireless sensor networks) and via historical repositories (for example, Twitter Firehose);

- Scalable data integration (meta-data management) techniques that can enable multi-dimensional querying over heterogeneous, real-time, and historical data in multiple contexts (such as spatial, temporal, semantics, source types, event types); and
- Cloud resource management methodologies that can seamlessly deal with heterogeneity in data analytic tasks, computational models, big data programming models, and cloud resource types (datacenter versus network edge, for example).

6.2.3 Data Management for Electric Vehicles Energy Management in an emergency

The emergence of Electric Vehicles (EVs) raised new research opportunities for energy management during emergency situations. In an emergency situation such as disaster, the EVs providing battery units can be considered a mobile service provider of energy stores to charge and discharge the energy to/from the grid and supply to critical infrastructures (CIs). For example, landslides can cause power supply disruption so that EVs, as a mobile energy provider, can charge electric energy to the grid to provide power supply for hospitals and emergency shelters in disaster relief services. Here, data management has thus played a significant role in managing IoT and EV sensor streams during the emergency. The potential research direction in data management for EV energy management in an emergency is as follows.

- **Analysis of core data and access types** – The data collection of IoT and EV sensors (such as smart meters, battery/photovoltaics (PV) panels for monitoring CIs energy supply and demand) is a challenge in EV energy management for emergencies. The variety of access protocols to the IoT and EV sensor systems requires technical and socio-technical processes to enable access to the data stream. Additionally, the data stream can be accessed directly through a device with internet connectivity or a third-party cloud data platform. Data provided retrospectively (such as city planning data from City Information Modeling

(CIM)/Building Information Modeling (BIM) and road networks) may require the social and process-based access system to data incorporating suitable trust, security, provenance, governance, Quality Assurance (QA) and management requirements.

- **Urban Information Modeling of cities** — This research involves the development of a model enabling representation of critical infrastructure identified in the city (such as building, hospitals, schools) and identify the interactions between the microscale model of Building Information Model (BIM) with the large-scale model at the City level provided by the Urban Information Model (UIM). The occurrence of micro-scale phenomena in the city (such as an electric power disruption) may affect the service at a local scale, for example, people in a railway station or hospital are prevented from using an elevator due to an electric blackout. This research will investigate the modelling of the city services provided by critical infrastructures (e.g. Transportation system, power network, Telecom Mobile service) into the BIM.

- **Development of technique for holistic (IoT-Edge-Cloud) data collection, filtering, and monitoring** — Data collection, filtering and monitoring of EV data during emergencies may experience unexpected events that can degrade the abilities of edge devices, for example, the sudden unavailability of sensing devices due to the failure of internet communications. Several techniques, such as statistical data sampling, online caching (such as storing previous computation results), have to be investigated to deal with the uncertainties at the data collection level. The research focuses on characterising high latencies and software/hardware failures during data offloading between the edge and cloud due to undesirable dynamics in communication and computation processes. The expected result will be an intelligent data collection and filtering technique that yields high data availability to provide energy restoration to CIs in disaster-affected urban areas.

- **Edge-Cloud data storage and querying** — This research investigates techniques for collecting real-time data from IoT monitoring devices, smart sensors, and EVs, and the storing and sharing of such data using edge-cloud infrastructure over longer time frames. Here, the development of hybrid big data management platforms for heterogeneous database architectures (such as stream processing, batch processing, SQL database, and NoSQL database) for disparate IoT data sensors (such as smart meters, EV sensors) to support seamless access to data distributed across different architectures (edge-cloud) is a major challenge of this research.

6.2.4 Time-series data analytics for event detection in hazard management

Time-series data analytics for event detection in hazard management is a significant future research direction of the thesis. The real-time response and future event prediction are essential topics in this research. In addition, the real-time response process analyses massive time-series data stream in real-time to identify hazards and hazard-related events to provide effective response to hazard and support timely decision making. The future event prediction is a process for analysing historical time-series data to understand the changing pattern of data and predict the potential of future hazard events.

- **Social media data analysis for emergency management** — Social media has played essential roles as valuable data sources to support timely decision making in hazard management. The techniques such as Natural Language Processing (NLP) and Convolutional Neural Network (CNN) has been used to extract and detect events in natural hazard contexts. However, there have been several open issues in social media data analysis that need to be investigated. The examples are (i) spatial and temporal analysis of social media data to identify/predict location and occurrence time of the hazard of interest; (ii) sentiment analysis from social media in active hazard warning and emergency response; (iii) real-time disaster damage assessment on social media; (iv) the analysis of social media pattern during hazard events; and (v) Hazard data quality assurance for social media data.

- **Knowledge graph data integration analysis** — The knowledge graph can be used to solve the challenge in the wide variety and heterogeneity of time-series data. Data extracted from social media is a potential source of time-series data that can be used to construct a domain-specific knowledge graph in real-time to enable social-driven decision support in emergency management. Based on this, several techniques for real-time construction of knowledge graph (e.g. data acquisition, information extraction) and the knowledge graph analysis (real-time reasoning, Machine Learning over knowledge graph) to support emergency response during hazard need to be further investigated.

- **Time-series data analysis for event prediction in hazard management** — Nowadays, there is a great effort to develop techniques that analyse a massive amount of time-series data to forecast hazards. Event prediction is an essential process to identify the potential occurrence of a future hazard. The process analyses historical time-series data using sophisticated techniques such as Data mining, Machine Learning to predict the future occurrence of hazards. The orchestration of time-series data from physical sensors and social media data is one of the challenging research in this topic to enhance the effectiveness of hazard event prediction. Such research includes (i) pattern recognition to forecast hazard occurrence using social media and physical sensors; (ii) development of geospatial techniques for hazards risk assessment using a time series of satellite images and social media; and (ii) Machine learning techniques (e.g. Bayesian network learning) to analyse time-series data and deal with uncertainty factors in natural hazard assessments.

Appendix I: Landslip Ontology

Appendix I: Landslip Ontology

```
<?xml version="1.0"?>
<Ontology xmlns="http://www.w3.org/2002/07/owl#"
     xml:base="http://www.semanticweb.org/ncl/ontologies/2018/6/landslip#"
     xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
     xmlns:xml="http://www.w3.org/XML/1998/namespace"
     xmlns:xsd="http://www.w3.org/2001/XMLSchema#"
     xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"
     ontologyIRI="http://www.semanticweb.org/ncl/ontologies/2018/6/landslip#">
    <Prefix name="" IRI="http://www.semanticweb.org/ncl/ontologies/2018/6/landslip#"/>
    <Prefix name="owl" IRI="http://www.w3.org/2002/07/owl#"/>
    <Prefix name="rdf" IRI="http://www.w3.org/1999/02/22-rdf-syntax-ns#"/>
    <Prefix name="xml" IRI="http://www.w3.org/XML/1998/namespace"/>
    <Prefix name="xsd" IRI="http://www.w3.org/2001/XMLSchema#"/>
    <Prefix name="rdfs" IRI="http://www.w3.org/2000/01/rdf-schema#"/>
    <Declaration>
        <ObjectProperty IRI="hasNaturalResource"/>
    </Declaration>
    <Declaration>
        <Class IRI="DistributionLine"/>
    </Declaration>
    <Declaration>
        <NamedIndividual IRI="soil_moisture_1"/>
    </Declaration>
    <Declaration>
        <DataProperty IRI="column"/>
    </Declaration>
    <Declaration>
        <Class IRI="WindDirection"/>
    </Declaration>
    <Declaration>
        <ObjectProperty IRI="isLocatedIn"/>
    </Declaration>
    <Declaration>
        <Class IRI="BulkDensity"/>
    </Declaration>
    <Declaration>
        <ObjectProperty IRI="hasIndicator"/>
    </Declaration>
    <Declaration>
        <Class IRI="SnowStorm"/>
    </Declaration>
    <Declaration>
        <Class IRI="Wildfire"/>
    </Declaration>
    <Declaration>
        <NamedIndividual IRI="flood_2"/>
    </Declaration>
    <Declaration>
        <Class IRI="VegetationRemoval"/>
    </Declaration>
    <Declaration>
        <DataProperty IRI="featureList"/>
    </Declaration>
    <Declaration>
        <NamedIndividual IRI="ds_metadata_2"/>
    </Declaration>
    <Declaration>
        <Class IRI="PhysicalSensor"/>
    </Declaration>
    <Declaration>
        <ObjectProperty IRI="livesIn"/>
    </Declaration>
    <Declaration>
        <NamedIndividual IRI="provider_profile_1"/>
    </Declaration>
    <Declaration>
        <DataProperty IRI="account"/>
    </Declaration>
    <Declaration>
        <Class IRI="AnthropogenicProcess"/>
    </Declaration>
    <Declaration>
        <Class IRI="NaturalResource"/>
    </Declaration>
    <Declaration>
        <Class IRI="RegionalSubsidence"/>
    </Declaration>
    <Declaration>
        <ObjectProperty IRI="isConnectedTo"/>
    </Declaration>
    <Declaration>
        <Class IRI="QuarryingSerfaceMining"/>
    </Declaration>
    <Declaration>
```

```xml
        <Class IRI="AgriculturalPracticeChange"/>
    </Declaration>
    <Declaration>
        <ObjectProperty IRI="hasProfile"/>
    </Declaration>
    <Declaration>
        <Class IRI="ObservedPropertyProfile"/>
    </Declaration>
    <Declaration>
        <DataProperty IRI="collection"/>
    </Declaration>
    <Declaration>
        <NamedIndividual IRI="crack_on_sideWalk_2"/>
    </Declaration>
    <Declaration>
        <ObjectProperty IRI="measures"/>
    </Declaration>
    <Declaration>
        <Class IRI="TreeFall"/>
    </Declaration>
    <Declaration>
        <Class IRI="WaterAddition"/>
    </Declaration>
    <Declaration>
        <Class IRI="HeavyRainFall"/>
    </Declaration>
    <Declaration>
        <Class IRI="TransmissionNetwork"/>
    </Declaration>
    <Declaration>
        <NamedIndividual IRI="ds_metadata_3"/>
    </Declaration>
    <Declaration>
        <Class IRI="DistributionPoleFall"/>
    </Declaration>
    <Declaration>
        <Class IRI="WarningSign"/>
    </Declaration>
    <Declaration>
        <Class IRI="IncreaseInWaterLevel"/>
    </Declaration>
    <Declaration>
        <Class IRI="ObservedProperty"/>
    </Declaration>
    <Declaration>
        <Class IRI="Railway"/>
    </Declaration>
    <Declaration>
        <Class IRI="InfrastructureConstruction"/>
    </Declaration>
    <Declaration>
        <DataProperty IRI="eventList"/>
    </Declaration>
    <Declaration>
        <NamedIndividual IRI="sensor_profile_1"/>
    </Declaration>
    <Declaration>
        <ObjectProperty IRI="isObservedPropertyFor"/>
    </Declaration>
    <Declaration>
        <Class IRI="Hospital"/>
    </Declaration>
    <Declaration>
        <Class IRI="ServiceProfile"/>
    </Declaration>
    <Declaration>
        <ObjectProperty IRI="triggers"/>
    </Declaration>
    <Declaration>
        <Class IRI="TransmissionLine"/>
    </Declaration>
    <Declaration>
        <ObjectProperty IRI="hasSocialMedia"/>
    </Declaration>
    <Declaration>
        <Class IRI="Resident"/>
    </Declaration>
    <Declaration>
        <NamedIndividual IRI="soil_movement_1"/>
    </Declaration>
    <Declaration>
        <Class IRI="Residence"/>
    </Declaration>
    <Declaration>
        <Class IRI="Flood"/>
    </Declaration>
    <Declaration>
        <Class IRI="Utility"/>
    </Declaration>
    <Declaration>
        <ObjectProperty IRI="hasMetadataProperty"/>
    </Declaration>
    <Declaration>
        <Class IRI="ChemicalExplosion"/>
    </Declaration>
    <Declaration>
```

```
        <NamedIndividual IRI="sensor_profile_2"/>
    </Declaration>
    <Declaration>
        <Class IRI="NewclearExplostion"/>
    </Declaration>
    <Declaration>
        <DataProperty IRI="table"/>
    </Declaration>
    <Declaration>
        <Class IRI="ExtremeTemperatureHot"/>
    </Declaration>
    <Declaration>
        <Class IRI="FlashOver"/>
    </Declaration>
    <Declaration>
        <ObjectProperty IRI="hasObservedProperty"/>
    </Declaration>
    <Declaration>
        <Class IRI="InfilledGround"/>
    </Declaration>
    <Declaration>
        <Class IRI="Infrastructure"/>
    </Declaration>
    <Declaration>
        <Class IRI="LeaningTelephonePole"/>
    </Declaration>
    <Declaration>
        <NamedIndividual IRI="obs_profile_21"/>
    </Declaration>
    <Declaration>
        <Class IRI="UrbanArea"/>
    </Declaration>
    <Declaration>
        <NamedIndividual IRI="obs_3"/>
    </Declaration>
    <Declaration>
        <DataProperty IRI="longitude"/>
    </Declaration>
    <Declaration>
        <Class IRI="Mountain"/>
    </Declaration>
    <Declaration>
        <NamedIndividual IRI="service_profile_12"/>
    </Declaration>
    <Declaration>
        <Class IRI="Slope"/>
    </Declaration>
    <Declaration>
        <Class IRI="ProviderProfile"/>
    </Declaration>
    <Declaration>
        <Class IRI="OverheadTransmissionPoleBroken"/>
    </Declaration>
    <Declaration>
        <ObjectProperty IRI="hasPlace"/>
    </Declaration>
    <Declaration>
        <ObjectProperty IRI="hasGeoLocation"/>
    </Declaration>
    <Declaration>
        <Class IRI="ConcretePole"/>
    </Declaration>
    <Declaration>
        <Class IRI="TreeLeaning"/>
    </Declaration>
    <Declaration>
        <ObjectProperty IRI="hasObservation"/>
    </Declaration>
    <Declaration>
        <Class IRI="Urbanisation"/>
    </Declaration>
    <Declaration>
        <Class IRI="GridNetwork"/>
    </Declaration>
    <Declaration>
        <NamedIndividual IRI="obs_2"/>
    </Declaration>
    <Declaration>
        <Class IRI="Tide"/>
    </Declaration>
    <Declaration>
        <NamedIndividual IRI="service_profile_11"/>
    </Declaration>
    <Declaration>
        <Class IRI="EarthQuakeMagnitude"/>
    </Declaration>
    <Declaration>
        <Class IRI="TransmissionPoleLeaning"/>
    </Declaration>
    <Declaration>
        <Class IRI="GroundWater"/>
    </Declaration>
    <Declaration>
        <DataProperty IRI="phenomenonEndTime"/>
    </Declaration>
    <Declaration>
```

```
        <Class IRI="Substation"/>
    </Declaration>
    <Declaration>
        <Class IRI="AirPressure"/>
    </Declaration>
    <Declaration>
        <Class IRI="Person"/>
    </Declaration>
    <Declaration>
        <Class IRI="CrackOnSideWalk"/>
    </Declaration>
    <Declaration>
        <Class IRI="Lightning"/>
    </Declaration>
    <Declaration>
        <ObjectProperty IRI="represent"/>
    </Declaration>
    <Declaration>
        <Class IRI="TransmissionPoleFall"/>
    </Declaration>
    <Declaration>
        <DataProperty IRI="observedPropertyType"/>
    </Declaration>
    <Declaration>
        <Class IRI="VoltageSurge"/>
    </Declaration>
    <Declaration>
        <NamedIndividual IRI="obs_1"/>
    </Declaration>
    <Declaration>
        <Class IRI="tapwater"/>
    </Declaration>
    <Declaration>
        <Class IRI="LeaningElectronicPole"/>
    </Declaration>
    <Declaration>
        <Class IRI="GroundCollapse"/>
    </Declaration>
    <Declaration>
        <Class IRI="Drought"/>
    </Declaration>
    <Declaration>
        <Class IRI="SeismicWave"/>
    </Declaration>
    <Declaration>
        <Class IRI="electricity"/>
    </Declaration>
    <Declaration>
        <NamedIndividual IRI="foi_Karela_bbox_2"/>
    </Declaration>
    <Declaration>
        <DataProperty IRI="procedure"/>
    </Declaration>
    <Declaration>
        <ObjectProperty IRI="isWarningSignFor"/>
    </Declaration>
    <Declaration>
        <ObjectProperty IRI="isObservationFor"/>
    </Declaration>
    <Declaration>
        <NamedIndividual IRI="flood_1"/>
    </Declaration>
    <Declaration>
        <NamedIndividual IRI="rain_1"/>
    </Declaration>
    <Declaration>
        <ObjectProperty IRI="hasUtility"/>
    </Declaration>
    <Declaration>
        <NamedIndividual IRI="service_profile_21"/>
    </Declaration>
    <Declaration>
        <Class IRI="ReservoirAndDamConstruction"/>
    </Declaration>
    <Declaration>
        <DataProperty IRI="serviceAdapter"/>
    </Declaration>
    <Declaration>
        <Class IRI="CrackOnConcreteFloor"/>
    </Declaration>
    <Declaration>
        <NamedIndividual IRI="foi_Karela_bbox_1"/>
    </Declaration>
    <Declaration>
        <DataProperty IRI="postdate"/>
    </Declaration>
    <Declaration>
        <Class IRI="BulgeOnGround"/>
    </Declaration>
    <Declaration>
        <Class IRI="GroundHeave"/>
    </Declaration>
    <Declaration>
        <Class IRI="BiologicalAttack"/>
    </Declaration>
    <Declaration>
```

```
    <DataProperty IRI="latitude"/>
</Declaration>
<Declaration>
    <ObjectProperty IRI="isDataSourceFor"/>
</Declaration>
<Declaration>
    <Class IRI="PublicPlace"/>
</Declaration>
<Declaration>
    <Class IRI="SoilMoisture"/>
</Declaration>
<Declaration>
    <Class IRI="SocialMedia"/>
</Declaration>
<Declaration>
    <Class IRI="Earthquake"/>
</Declaration>
<Declaration>
    <Class IRI="ImpactEvent"/>
</Declaration>
<Declaration>
    <NamedIndividual IRI="landslide_2"/>
</Declaration>
<Declaration>
    <ObjectProperty IRI="isGeoLocationFor"/>
</Declaration>
<Declaration>
    <DataProperty IRI="serviceURL"/>
</Declaration>
<Declaration>
    <ObjectProperty IRI="hasDataSourceMetadata"/>
</Declaration>
<Declaration>
    <Class IRI="ExtremeTemperatureCold"/>
</Declaration>
<Declaration>
    <ObjectProperty IRI="isIndicationOf"/>
</Declaration>
<Declaration>
    <Class IRI="Expert"/>
</Declaration>
<Declaration>
    <Class IRI="OverheadDistributionLineBroken"/>
</Declaration>
<Declaration>
    <NamedIndividual IRI="obs_profile_12"/>
</Declaration>
<Declaration>
    <Class IRI="SoilLocalSubsidence"/>
</Declaration>
<Declaration>
    <Class IRI="House"/>
</Declaration>
<Declaration>
    <Class IRI="Observation"/>
</Declaration>
<Declaration>
    <Class IRI="GeneratingPlant"/>
</Declaration>
<Declaration>
    <Class IRI="DataSource"/>
</Declaration>
<Declaration>
    <Class IRI="MobileApp"/>
</Declaration>
<Declaration>
    <NamedIndividual IRI="landslide_1"/>
</Declaration>
<Declaration>
    <NamedIndividual IRI="gridnetwork_1"/>
</Declaration>
<Declaration>
    <ObjectProperty IRI="hasWarningSign"/>
</Declaration>
<Declaration>
    <Class IRI="GridSystem"/>
</Declaration>
<Declaration>
    <Class IRI="CrackOnWall"/>
</Declaration>
<Declaration>
    <Class IRI="InfrastructureLoading"/>
</Declaration>
<Declaration>
    <Class IRI="Tilt"/>
</Declaration>
<Declaration>
    <NamedIndividual IRI="obs_profile_11"/>
</Declaration>
<Declaration>
    <Class IRI="School"/>
</Declaration>
<Declaration>
    <NamedIndividual IRI="geolocation_2"/>
</Declaration>
<Declaration>
```

```
        <Class IRI="Flat"/>
    </Declaration>
    <Declaration>
        <DataProperty IRI="mediaType"/>
    </Declaration>
    <Declaration>
        <Class IRI="CumulativeRainfall"/>
    </Declaration>
    <Declaration>
        <Class IRI="FoI"/>
    </Declaration>
    <Declaration>
        <Class IRI="Humidity"/>
    </Declaration>
    <Declaration>
        <Class IRI="Place"/>
    </Declaration>
    <Declaration>
        <DataProperty IRI="serviceProvider"/>
    </Declaration>
    <Declaration>
        <Class IRI="HailStorm"/>
    </Declaration>
    <Declaration>
        <DataProperty IRI="dbname"/>
    </Declaration>
    <Declaration>
        <Class IRI="Road"/>
    </Declaration>
    <Declaration>
        <Class IRI="OilGasExtraction"/>
    </Declaration>
    <Declaration>
        <Class IRI="SensorProfile"/>
    </Declaration>
    <Declaration>
        <ObjectProperty IRI="observes"/>
    </Declaration>
    <Declaration>
        <Class IRI="HumanSensor"/>
    </Declaration>
    <Declaration>
        <Class IRI="Lake"/>
    </Declaration>
    <Declaration>
        <Class IRI="Rain"/>
    </Declaration>
    <Declaration>
        <Class IRI="VolcanicEruption"/>
    </Declaration>
    <Declaration>
        <Class IRI="SoilMovement"/>
    </Declaration>
    <Declaration>
        <ObjectProperty IRI="contributes"/>
    </Declaration>
    <Declaration>
        <Class IRI="BrokenUndergroundUtility"/>
    </Declaration>
    <Declaration>
        <Class IRI="River"/>
    </Declaration>
    <Declaration>
        <DataProperty IRI="observedPropertyName"/>
    </Declaration>
    <Declaration>
        <Class IRI="DistributionPole"/>
    </Declaration>
    <Declaration>
        <Class IRI="WindSpeed"/>
    </Declaration>
    <Declaration>
        <Class IRI="Storm"/>
    </Declaration>
    <Declaration>
        <Class IRI="Tsunami"/>
    </Declaration>
    <Declaration>
        <Class IRI="SubsurfaceMining"/>
    </Declaration>
    <Declaration>
        <Class IRI="DataSourceMetadata"/>
    </Declaration>
    <Declaration>
        <Class IRI="ObservationProfile"/>
    </Declaration>
    <Declaration>
        <Class IRI="AirTemperature"/>
    </Declaration>
    <Declaration>
        <Class IRI="MaterialFluidInjection"/>
    </Declaration>
    <Declaration>
        <Class IRI="Waterfall"/>
    </Declaration>
    <Declaration>
```

```
        <Class IRI="DrainageAndDewatering"/>
    </Declaration>
    <Declaration>
        <DataProperty IRI="bbox"/>
    </Declaration>
    <Declaration>
        <DataProperty IRI="featureOfInterest"/>
    </Declaration>
    <Declaration>
        <Class IRI="DistributionPoleLeaning"/>
    </Declaration>
    <Declaration>
        <Class IRI="InSitu"/>
    </Declaration>
    <Declaration>
        <Class IRI="SnowAvalanche"/>
    </Declaration>
    <Declaration>
        <Class IRI="CrackonBuilding"/>
    </Declaration>
    <Declaration>
        <Class IRI="Vibration"/>
    </Declaration>
    <Declaration>
        <Class IRI="GeoLocation"/>
    </Declaration>
    <Declaration>
        <Class IRI="SubsurfaceInfrastructureConstruction"/>
    </Declaration>
    <Declaration>
        <ObjectProperty IRI="hasDataSource"/>
    </Declaration>
    <Declaration>
        <Class IRI="TapWaterColourChange"/>
    </Declaration>
    <Declaration>
        <NamedIndividual IRI="geolocation_1"/>
    </Declaration>
    <Declaration>
        <Class IRI="SteelPole"/>
    </Declaration>
    <Declaration>
        <Class IRI="CrackOnStreet"/>
    </Declaration>
    <Declaration>
        <ObjectProperty IRI="hasObservedPropertyProfile"/>
    </Declaration>
    <Declaration>
        <NamedIndividual IRI="dataSource_1"/>
    </Declaration>
    <Declaration>
        <DataProperty IRI="serviceType"/>
    </Declaration>
    <Declaration>
        <NamedIndividual IRI="leaning_telephone_pole_1"/>
    </Declaration>
    <Declaration>
        <DataProperty IRI="providerName"/>
    </Declaration>
    <Declaration>
        <ObjectProperty IRI="hasInfrastructure"/>
    </Declaration>
    <Declaration>
        <Class IRI="WaterLevel"/>
    </Declaration>
    <Declaration>
        <Class IRI="Landslide"/>
    </Declaration>
    <Declaration>
        <DataProperty IRI="timeInstanceOrPeriod"/>
    </Declaration>
    <Declaration>
        <Class IRI="SoilTemperature"/>
    </Declaration>
    <Declaration>
        <DataProperty IRI="path"/>
    </Declaration>
    <Declaration>
        <NamedIndividual IRI="dataSource_2"/>
    </Declaration>
    <Declaration>
        <Class IRI="Pylon"/>
    </Declaration>
    <Declaration>
        <Class IRI="SocialMediaUser"/>
    </Declaration>
    <Declaration>
        <NamedIndividual IRI="crack_on_wall_2"/>
    </Declaration>
    <Declaration>
        <DataProperty IRI="sensorType"/>
    </Declaration>
    <Declaration>
        <NamedIndividual IRI="provider_profile_2"/>
    </Declaration>
    <Declaration>
```

```
        <Class IRI="LandslideProneArea"/>
    </Declaration>
    <Declaration>
        <Class IRI="NaturalHazard"/>
    </Declaration>
    <Declaration>
        <Class IRI="DataService"/>
    </Declaration>
    <Declaration>
        <DataProperty IRI="where"/>
    </Declaration>
    <Declaration>
        <DataProperty IRI="phenomenonBeginTime"/>
    </Declaration>
    <Declaration>
        <Class IRI="WoodenPole"/>
    </Declaration>
    <Declaration>
        <NamedIndividual IRI="dataSource_3"/>
    </Declaration>
    <Declaration>
        <ObjectProperty IRI="hasFoI"/>
    </Declaration>
    <Declaration>
        <Class IRI="SocialMediaAccount"/>
    </Declaration>
    <Declaration>
        <NamedIndividual IRI="ds_metadata_1"/>
    </Declaration>
    <SubClassOf>
        <Class IRI="AgriculturalPracticeChange"/>
        <Class IRI="AnthropogenicProcess"/>
    </SubClassOf>
    <SubClassOf>
        <Class IRI="AirPressure"/>
        <Class IRI="ObservedProperty"/>
    </SubClassOf>
    <SubClassOf>
        <Class IRI="AirTemperature"/>
        <Class IRI="ObservedProperty"/>
    </SubClassOf>
    <SubClassOf>
        <Class IRI="BiologicalAttack"/>
        <Class IRI="WarningSign"/>
    </SubClassOf>
    <SubClassOf>
        <Class IRI="BrokenUndergroundUtility"/>
        <Class IRI="WarningSign"/>
    </SubClassOf>
    <SubClassOf>
        <Class IRI="BulgeOnGround"/>
        <Class IRI="WarningSign"/>
    </SubClassOf>
    <SubClassOf>
        <Class IRI="BulkDensity"/>
        <Class IRI="ObservedProperty"/>
    </SubClassOf>
    <SubClassOf>
        <Class IRI="ChemicalExplosion"/>
        <Class IRI="AnthropogenicProcess"/>
    </SubClassOf>
    <SubClassOf>
        <Class IRI="ConcretePole"/>
        <Class IRI="DistributionPole"/>
    </SubClassOf>
    <SubClassOf>
        <Class IRI="CrackOnConcreteFloor"/>
        <Class IRI="WarningSign"/>
    </SubClassOf>
    <SubClassOf>
        <Class IRI="CrackOnSideWalk"/>
        <Class IRI="WarningSign"/>
    </SubClassOf>
    <SubClassOf>
        <Class IRI="CrackOnStreet"/>
        <Class IRI="WarningSign"/>
    </SubClassOf>
    <SubClassOf>
        <Class IRI="CrackOnWall"/>
        <Class IRI="WarningSign"/>
    </SubClassOf>
    <SubClassOf>
        <Class IRI="CrackonBuilding"/>
        <Class IRI="WarningSign"/>
    </SubClassOf>
    <SubClassOf>
        <Class IRI="CumulativeRainfall"/>
        <Class IRI="ObservedProperty"/>
    </SubClassOf>
    <SubClassOf>
        <Class IRI="DataService"/>
        <Class IRI="DataSource"/>
    </SubClassOf>
    <SubClassOf>
        <Class IRI="DistributionLine"/>
        <Class IRI="GridNetwork"/>
```

```xml
</SubClassOf>
<SubClassOf>
    <Class IRI="DistributionPole"/>
    <Class IRI="GridNetwork"/>
</SubClassOf>
<SubClassOf>
    <Class IRI="DistributionPoleFall"/>
    <Class IRI="WarningSign"/>
</SubClassOf>
<SubClassOf>
    <Class IRI="DistributionPoleLeaning"/>
    <Class IRI="WarningSign"/>
</SubClassOf>
<SubClassOf>
    <Class IRI="DrainageAndDewatering"/>
    <Class IRI="AnthropogenicProcess"/>
</SubClassOf>
<SubClassOf>
    <Class IRI="Drought"/>
    <Class IRI="NaturalHazard"/>
</SubClassOf>
<SubClassOf>
    <Class IRI="EarthQuakeMagnitude"/>
    <Class IRI="ObservedProperty"/>
</SubClassOf>
<SubClassOf>
    <Class IRI="Earthquake"/>
    <Class IRI="NaturalHazard"/>
</SubClassOf>
<SubClassOf>
    <Class IRI="Expert"/>
    <Class IRI="Person"/>
</SubClassOf>
<SubClassOf>
    <Class IRI="ExtremeTemperatureCold"/>
    <Class IRI="NaturalHazard"/>
</SubClassOf>
<SubClassOf>
    <Class IRI="ExtremeTemperatureHot"/>
    <Class IRI="NaturalHazard"/>
</SubClassOf>
<SubClassOf>
    <Class IRI="FlashOver"/>
    <Class IRI="WarningSign"/>
</SubClassOf>
<SubClassOf>
    <Class IRI="Flat"/>
    <Class IRI="Place"/>
</SubClassOf>
<SubClassOf>
    <Class IRI="Flood"/>
    <Class IRI="NaturalHazard"/>
</SubClassOf>
<SubClassOf>
    <Class IRI="GeneratingPlant"/>
    <Class IRI="GridSystem"/>
</SubClassOf>
<SubClassOf>
    <Class IRI="GridNetwork"/>
    <Class IRI="GridSystem"/>
</SubClassOf>
<SubClassOf>
    <Class IRI="GroundCollapse"/>
    <Class IRI="NaturalHazard"/>
</SubClassOf>
<SubClassOf>
    <Class IRI="GroundHeave"/>
    <Class IRI="NaturalHazard"/>
</SubClassOf>
<SubClassOf>
    <Class IRI="GroundWater"/>
    <Class IRI="ObservedProperty"/>
</SubClassOf>
<SubClassOf>
    <Class IRI="HailStorm"/>
    <Class IRI="NaturalHazard"/>
</SubClassOf>
<SubClassOf>
    <Class IRI="HeavyRainFall"/>
    <Class IRI="WarningSign"/>
</SubClassOf>
<SubClassOf>
    <Class IRI="Hospital"/>
    <Class IRI="PublicPlace"/>
</SubClassOf>
<SubClassOf>
    <Class IRI="House"/>
    <Class IRI="Place"/>
</SubClassOf>
<SubClassOf>
    <Class IRI="HumanSensor"/>
    <Class IRI="DataSource"/>
</SubClassOf>
<SubClassOf>
    <Class IRI="Humidity"/>
    <Class IRI="ObservedProperty"/>
```

```
</SubClassOf>
<SubClassOf>
    <Class IRI="ImpactEvent"/>
    <Class IRI="NaturalHazard"/>
</SubClassOf>
<SubClassOf>
    <Class IRI="InSitu"/>
    <Class IRI="PhysicalSensor"/>
</SubClassOf>
<SubClassOf>
    <Class IRI="IncreaseInWaterLevel"/>
    <Class IRI="WarningSign"/>
</SubClassOf>
<SubClassOf>
    <Class IRI="InfilledGround"/>
    <Class IRI="AnthropogenicProcess"/>
</SubClassOf>
<SubClassOf>
    <Class IRI="InfrastructureConstruction"/>
    <Class IRI="AnthropogenicProcess"/>
</SubClassOf>
<SubClassOf>
    <Class IRI="InfrastructureLoading"/>
    <Class IRI="AnthropogenicProcess"/>
</SubClassOf>
<SubClassOf>
    <Class IRI="Lake"/>
    <Class IRI="NaturalResource"/>
</SubClassOf>
<SubClassOf>
    <Class IRI="Landslide"/>
    <Class IRI="NaturalHazard"/>
</SubClassOf>
<SubClassOf>
    <Class IRI="LandslideProneArea"/>
    <Class IRI="UrbanArea"/>
</SubClassOf>
<SubClassOf>
    <Class IRI="LeaningElectronicPole"/>
    <Class IRI="WarningSign"/>
</SubClassOf>
<SubClassOf>
    <Class IRI="LeaningTelephonePole"/>
    <Class IRI="WarningSign"/>
</SubClassOf>
<SubClassOf>
    <Class IRI="Lightning"/>
    <Class IRI="NaturalHazard"/>
</SubClassOf>
<SubClassOf>
    <Class IRI="MaterialFluidInjection"/>
    <Class IRI="AnthropogenicProcess"/>
</SubClassOf>
<SubClassOf>
    <Class IRI="MobileApp"/>
    <Class IRI="HumanSensor"/>
</SubClassOf>
<SubClassOf>
    <Class IRI="Mountain"/>
    <Class IRI="NaturalResource"/>
</SubClassOf>
<SubClassOf>
    <Class IRI="NewclearExplostion"/>
    <Class IRI="AnthropogenicProcess"/>
</SubClassOf>
<SubClassOf>
    <Class IRI="OilGasExtraction"/>
    <Class IRI="AnthropogenicProcess"/>
</SubClassOf>
<SubClassOf>
    <Class IRI="OverheadDistributionLineBroken"/>
    <Class IRI="WarningSign"/>
</SubClassOf>
<SubClassOf>
    <Class IRI="OverheadTransmissionPoleBroken"/>
    <Class IRI="WarningSign"/>
</SubClassOf>
<SubClassOf>
    <Class IRI="PhysicalSensor"/>
    <Class IRI="DataSource"/>
</SubClassOf>
<SubClassOf>
    <Class IRI="PublicPlace"/>
    <Class IRI="Place"/>
</SubClassOf>
<SubClassOf>
    <Class IRI="Pylon"/>
    <Class IRI="TransmissionNetwork"/>
</SubClassOf>
<SubClassOf>
    <Class IRI="QuarryingSerfaceMining"/>
    <Class IRI="AnthropogenicProcess"/>
</SubClassOf>
<SubClassOf>
    <Class IRI="Railway"/>
    <Class IRI="Infrastructure"/>
```

```
</SubClassOf>
<SubClassOf>
    <Class IRI="Rain"/>
    <Class IRI="ObservedProperty"/>
</SubClassOf>
<SubClassOf>
    <Class IRI="RegionalSubsidence"/>
    <Class IRI="NaturalHazard"/>
</SubClassOf>
<SubClassOf>
    <Class IRI="ReservoirAndDamConstruction"/>
    <Class IRI="AnthropogenicProcess"/>
</SubClassOf>
<SubClassOf>
    <Class IRI="Residence"/>
    <Class IRI="Place"/>
</SubClassOf>
<SubClassOf>
    <Class IRI="Resident"/>
    <Class IRI="Person"/>
</SubClassOf>
<SubClassOf>
    <Class IRI="River"/>
    <Class IRI="NaturalResource"/>
</SubClassOf>
<SubClassOf>
    <Class IRI="Road"/>
    <Class IRI="Infrastructure"/>
</SubClassOf>
<SubClassOf>
    <Class IRI="School"/>
    <Class IRI="PublicPlace"/>
</SubClassOf>
<SubClassOf>
    <Class IRI="SeismicWave"/>
    <Class IRI="ObservedProperty"/>
</SubClassOf>
<SubClassOf>
    <Class IRI="Slope"/>
    <Class IRI="ObservedProperty"/>
</SubClassOf>
<SubClassOf>
    <Class IRI="SnowAvalanche"/>
    <Class IRI="NaturalHazard"/>
</SubClassOf>
<SubClassOf>
    <Class IRI="SnowStorm"/>
    <Class IRI="NaturalHazard"/>
</SubClassOf>
<SubClassOf>
    <Class IRI="SocialMedia"/>
    <Class IRI="HumanSensor"/>
</SubClassOf>
<SubClassOf>
    <Class IRI="SocialMediaUser"/>
    <Class IRI="Person"/>
</SubClassOf>
<SubClassOf>
    <Class IRI="SoilLocalSubsidence"/>
    <Class IRI="NaturalHazard"/>
</SubClassOf>
<SubClassOf>
    <Class IRI="SoilMoisture"/>
    <Class IRI="ObservedProperty"/>
</SubClassOf>
<SubClassOf>
    <Class IRI="SoilMovement"/>
    <Class IRI="ObservedProperty"/>
</SubClassOf>
<SubClassOf>
    <Class IRI="SoilTemperature"/>
    <Class IRI="ObservedProperty"/>
</SubClassOf>
<SubClassOf>
    <Class IRI="SteelPole"/>
    <Class IRI="DistributionPole"/>
</SubClassOf>
<SubClassOf>
    <Class IRI="Storm"/>
    <Class IRI="NaturalHazard"/>
</SubClassOf>
<SubClassOf>
    <Class IRI="Substation"/>
    <Class IRI="GridSystem"/>
</SubClassOf>
<SubClassOf>
    <Class IRI="SubsurfaceInfrastructureConstruction"/>
    <Class IRI="AnthropogenicProcess"/>
</SubClassOf>
<SubClassOf>
    <Class IRI="SubsurfaceMining"/>
    <Class IRI="AnthropogenicProcess"/>
</SubClassOf>
<SubClassOf>
    <Class IRI="TapWaterColourChange"/>
    <Class IRI="WarningSign"/>
```

```
    </SubClassOf>
    <SubClassOf>
        <Class IRI="Tide"/>
        <Class IRI="ObservedProperty"/>
    </SubClassOf>
    <SubClassOf>
        <Class IRI="Tilt"/>
        <Class IRI="ObservedProperty"/>
    </SubClassOf>
    <SubClassOf>
        <Class IRI="TransmissionLine"/>
        <Class IRI="TransmissionNetwork"/>
    </SubClassOf>
    <SubClassOf>
        <Class IRI="TransmissionNetwork"/>
        <Class IRI="GridSystem"/>
    </SubClassOf>
    <SubClassOf>
        <Class IRI="TransmissionPoleFall"/>
        <Class IRI="WarningSign"/>
    </SubClassOf>
    <SubClassOf>
        <Class IRI="TransmissionPoleLeaning"/>
        <Class IRI="WarningSign"/>
    </SubClassOf>
    <SubClassOf>
        <Class IRI="TreeFall"/>
        <Class IRI="WarningSign"/>
    </SubClassOf>
    <SubClassOf>
        <Class IRI="TreeLeaning"/>
        <Class IRI="WarningSign"/>
    </SubClassOf>
    <SubClassOf>
        <Class IRI="Tsunami"/>
        <Class IRI="NaturalHazard"/>
    </SubClassOf>
    <SubClassOf>
        <Class IRI="Urbanisation"/>
        <Class IRI="AnthropogenicProcess"/>
    </SubClassOf>
    <SubClassOf>
        <Class IRI="VegetationRemoval"/>
        <Class IRI="AnthropogenicProcess"/>
    </SubClassOf>
    <SubClassOf>
        <Class IRI="Vibration"/>
        <Class IRI="ObservedProperty"/>
    </SubClassOf>
    <SubClassOf>
        <Class IRI="VolcanicEruption"/>
        <Class IRI="NaturalHazard"/>
    </SubClassOf>
    <SubClassOf>
        <Class IRI="VoltageSurge"/>
        <Class IRI="WarningSign"/>
    </SubClassOf>
    <SubClassOf>
        <Class IRI="WaterAddition"/>
        <Class IRI="AnthropogenicProcess"/>
    </SubClassOf>
    <SubClassOf>
        <Class IRI="WaterLevel"/>
        <Class IRI="ObservedProperty"/>
    </SubClassOf>
    <SubClassOf>
        <Class IRI="Waterfall"/>
        <Class IRI="NaturalResource"/>
    </SubClassOf>
    <SubClassOf>
        <Class IRI="Wildfire"/>
        <Class IRI="NaturalHazard"/>
    </SubClassOf>
    <SubClassOf>
        <Class IRI="WindDirection"/>
        <Class IRI="ObservedProperty"/>
    </SubClassOf>
    <SubClassOf>
        <Class IRI="WindSpeed"/>
        <Class IRI="ObservedProperty"/>
    </SubClassOf>
    <SubClassOf>
        <Class IRI="WoodenPole"/>
        <Class IRI="DistributionPole"/>
    </SubClassOf>
    <SubClassOf>
        <Class IRI="electricity"/>
        <Class IRI="Utility"/>
    </SubClassOf>
    <SubClassOf>
        <Class IRI="tapwater"/>
        <Class IRI="Utility"/>
    </SubClassOf>
    <DisjointClasses>
        <Class IRI="AgriculturalPracticeChange"/>
        <Class IRI="ChemicalExplosion"/>
```

```
        <Class IRI="DrainageAndDewatering"/>
        <Class IRI="InfilledGround"/>
        <Class IRI="InfrastructureConstruction"/>
        <Class IRI="InfrastructureLoading"/>
        <Class IRI="MaterialFluidInjection"/>
        <Class IRI="NewclearExplostion"/>
        <Class IRI="OilGasExtraction"/>
        <Class IRI="QuarryingSerfaceMining"/>
        <Class IRI="ReservoirAndDamConstruction"/>
        <Class IRI="SubsurfaceInfrastructureConstruction"/>
        <Class IRI="SubsurfaceMining"/>
        <Class IRI="Urbanisation"/>
        <Class IRI="VegetationRemoval"/>
        <Class IRI="WaterAddition"/>
    </DisjointClasses>
    <DisjointClasses>
        <Class IRI="AirPressure"/>
        <Class IRI="AirTemperature"/>
        <Class IRI="BulkDensity"/>
        <Class IRI="CumulativeRainfall"/>
        <Class IRI="EarthQuakeMagnitude"/>
        <Class IRI="GroundWater"/>
        <Class IRI="Humidity"/>
        <Class IRI="Rain"/>
        <Class IRI="SeismicWave"/>
        <Class IRI="Slope"/>
        <Class IRI="SoilMoisture"/>
        <Class IRI="SoilMovement"/>
        <Class IRI="SoilTemperature"/>
        <Class IRI="Tide"/>
        <Class IRI="Tilt"/>
        <Class IRI="Vibration"/>
        <Class IRI="WaterLevel"/>
        <Class IRI="WindDirection"/>
        <Class IRI="WindSpeed"/>
    </DisjointClasses>
    <DisjointClasses>
        <Class IRI="AirPressure"/>
        <Class IRI="AirTemperature"/>
        <Class IRI="EarthQuakeMagnitude"/>
        <Class IRI="Humidity"/>
        <Class IRI="Rain"/>
        <Class IRI="SeismicWave"/>
        <Class IRI="SoilMoisture"/>
        <Class IRI="SoilMovement"/>
        <Class IRI="Tide"/>
        <Class IRI="WindSpeed"/>
    </DisjointClasses>
    <DisjointClasses>
        <Class IRI="AnthropogenicProcess"/>
        <Class IRI="DataSource"/>
        <Class IRI="DataSourceMetadata"/>
        <Class IRI="FoI"/>
        <Class IRI="GeoLocation"/>
        <Class IRI="Infrastructure"/>
        <Class IRI="NaturalHazard"/>
        <Class IRI="NaturalResource"/>
        <Class IRI="Observation"/>
        <Class IRI="ObservationProfile"/>
        <Class IRI="ObservedProperty"/>
        <Class IRI="Person"/>
        <Class IRI="Place"/>
        <Class IRI="SensorProfile"/>
        <Class IRI="ServiceProfile"/>
        <Class IRI="UrbanArea"/>
        <Class IRI="Utility"/>
        <Class IRI="WarningSign"/>
    </DisjointClasses>
    <DisjointClasses>
        <Class IRI="AnthropogenicProcess"/>
        <Class IRI="DataSource"/>
        <Class IRI="DataSourceMetadata"/>
        <Class IRI="FoI"/>
        <Class IRI="GeoLocation"/>
        <Class IRI="Infrastructure"/>
        <Class IRI="NaturalHazard"/>
        <Class IRI="NaturalResource"/>
        <Class IRI="Observation"/>
        <Class IRI="ObservedProperty"/>
        <Class IRI="Person"/>
        <Class IRI="Place"/>
        <Class IRI="UrbanArea"/>
        <Class IRI="Utility"/>
    </DisjointClasses>
    <DisjointClasses>
        <Class IRI="AnthropogenicProcess"/>
        <Class IRI="DataSource"/>
        <Class IRI="DataSourceMetadata"/>
        <Class IRI="FoI"/>
        <Class IRI="Infrastructure"/>
        <Class IRI="NaturalHazard"/>
        <Class IRI="NaturalResource"/>
        <Class IRI="Observation"/>
        <Class IRI="ObservedProperty"/>
        <Class IRI="Person"/>
        <Class IRI="Place"/>
```

```
        <Class IRI="UrbanArea"/>
        <Class IRI="Utility"/>
    </DisjointClasses>
    <DisjointClasses>
        <Class IRI="AnthropogenicProcess"/>
        <Class IRI="DataSource"/>
        <Class IRI="DataSourceMetadata"/>
        <Class IRI="Infrastructure"/>
        <Class IRI="NaturalHazard"/>
        <Class IRI="NaturalResource"/>
        <Class IRI="Observation"/>
        <Class IRI="ObservedProperty"/>
        <Class IRI="Person"/>
        <Class IRI="Place"/>
        <Class IRI="UrbanArea"/>
        <Class IRI="Utility"/>
    </DisjointClasses>
    <DisjointClasses>
        <Class IRI="AnthropogenicProcess"/>
        <Class IRI="DataSource"/>
        <Class IRI="Infrastructure"/>
        <Class IRI="NaturalHazard"/>
        <Class IRI="NaturalResource"/>
        <Class IRI="Observation"/>
        <Class IRI="ObservedProperty"/>
        <Class IRI="Person"/>
        <Class IRI="Place"/>
        <Class IRI="UrbanArea"/>
        <Class IRI="Utility"/>
    </DisjointClasses>
    <DisjointClasses>
        <Class IRI="AnthropogenicProcess"/>
        <Class IRI="Infrastructure"/>
        <Class IRI="NaturalHazard"/>
        <Class IRI="NaturalResource"/>
        <Class IRI="Observation"/>
        <Class IRI="ObservedProperty"/>
        <Class IRI="Person"/>
        <Class IRI="Place"/>
        <Class IRI="UrbanArea"/>
        <Class IRI="Utility"/>
    </DisjointClasses>
    <DisjointClasses>
        <Class IRI="AnthropogenicProcess"/>
        <Class IRI="Infrastructure"/>
        <Class IRI="NaturalHazard"/>
        <Class IRI="NaturalResource"/>
        <Class IRI="Observation"/>
        <Class IRI="Person"/>
        <Class IRI="Place"/>
        <Class IRI="UrbanArea"/>
        <Class IRI="Utility"/>
    </DisjointClasses>
    <DisjointClasses>
        <Class IRI="AnthropogenicProcess"/>
        <Class IRI="Infrastructure"/>
        <Class IRI="NaturalHazard"/>
        <Class IRI="NaturalResource"/>
        <Class IRI="Person"/>
        <Class IRI="Place"/>
        <Class IRI="UrbanArea"/>
        <Class IRI="Utility"/>
    </DisjointClasses>
    <DisjointClasses>
        <Class IRI="AnthropogenicProcess"/>
        <Class IRI="Infrastructure"/>
        <Class IRI="NaturalHazard"/>
        <Class IRI="NaturalResource"/>
        <Class IRI="Place"/>
        <Class IRI="UrbanArea"/>
        <Class IRI="Utility"/>
    </DisjointClasses>
    <DisjointClasses>
        <Class IRI="AnthropogenicProcess"/>
        <Class IRI="Infrastructure"/>
        <Class IRI="NaturalHazard"/>
        <Class IRI="NaturalResource"/>
        <Class IRI="Place"/>
        <Class IRI="Utility"/>
    </DisjointClasses>
    <DisjointClasses>
        <Class IRI="AnthropogenicProcess"/>
        <Class IRI="Infrastructure"/>
        <Class IRI="NaturalHazard"/>
        <Class IRI="NaturalResource"/>
        <Class IRI="Utility"/>
    </DisjointClasses>
    <DisjointClasses>
        <Class IRI="AnthropogenicProcess"/>
        <Class IRI="NaturalHazard"/>
    </DisjointClasses>
    <DisjointClasses>
        <Class IRI="AnthropogenicProcess"/>
        <Class IRI="NaturalHazard"/>
        <Class IRI="NaturalResource"/>
    </DisjointClasses>
```

```xml
<DisjointClasses>
    <Class IRI="AnthropogenicProcess"/>
    <Class IRI="NaturalHazard"/>
    <Class IRI="NaturalResource"/>
    <Class IRI="Utility"/>
</DisjointClasses>
<DisjointClasses>
    <Class IRI="BiologicalAttack"/>
    <Class IRI="BrokenUndergroundUtility"/>
    <Class IRI="BulgeOnGround"/>
    <Class IRI="CrackOnConcreteFloor"/>
    <Class IRI="CrackOnSideWalk"/>
    <Class IRI="CrackOnStreet"/>
    <Class IRI="CrackOnWall"/>
    <Class IRI="CrackonBuilding"/>
    <Class IRI="DistributionPoleFall"/>
    <Class IRI="DistributionPoleLeaning"/>
    <Class IRI="FlashOver"/>
    <Class IRI="HeavyRainFall"/>
    <Class IRI="IncreaseInWaterLevel"/>
    <Class IRI="LeaningElectronicPole"/>
    <Class IRI="LeaningTelephonePole"/>
    <Class IRI="OverheadDistributionLineBroken"/>
    <Class IRI="OverheadTransmissionPoleBroken"/>
    <Class IRI="TapWaterColourChange"/>
    <Class IRI="TransmissionPoleFall"/>
    <Class IRI="TransmissionPoleLeaning"/>
    <Class IRI="TreeFall"/>
    <Class IRI="TreeLeaning"/>
    <Class IRI="VoltageSurge"/>
</DisjointClasses>
<DisjointClasses>
    <Class IRI="BiologicalAttack"/>
    <Class IRI="BrokenUndergroundUtility"/>
    <Class IRI="BulgeOnGround"/>
    <Class IRI="CrackOnConcreteFloor"/>
    <Class IRI="CrackOnSideWalk"/>
    <Class IRI="CrackOnStreet"/>
    <Class IRI="CrackOnWall"/>
    <Class IRI="CrackonBuilding"/>
    <Class IRI="DistributionPoleFall"/>
    <Class IRI="DistributionPoleLeaning"/>
    <Class IRI="FlashOver"/>
    <Class IRI="IncreaseInWaterLevel"/>
    <Class IRI="LeaningElectronicPole"/>
    <Class IRI="LeaningTelephonePole"/>
    <Class IRI="OverheadDistributionLineBroken"/>
    <Class IRI="OverheadTransmissionPoleBroken"/>
    <Class IRI="TapWaterColourChange"/>
    <Class IRI="TransmissionPoleFall"/>
    <Class IRI="TransmissionPoleLeaning"/>
    <Class IRI="TreeFall"/>
    <Class IRI="TreeLeaning"/>
    <Class IRI="VoltageSurge"/>
</DisjointClasses>
<DisjointClasses>
    <Class IRI="BrokenUndergroundUtility"/>
    <Class IRI="BulgeOnGround"/>
    <Class IRI="CrackOnConcreteFloor"/>
    <Class IRI="CrackOnSideWalk"/>
    <Class IRI="CrackOnStreet"/>
    <Class IRI="CrackOnWall"/>
    <Class IRI="CrackonBuilding"/>
    <Class IRI="IncreaseInWaterLevel"/>
    <Class IRI="LeaningElectronicPole"/>
    <Class IRI="LeaningTelephonePole"/>
    <Class IRI="TapWaterColourChange"/>
</DisjointClasses>
<DisjointClasses>
    <Class IRI="ConcretePole"/>
    <Class IRI="SteelPole"/>
    <Class IRI="WoodenPole"/>
</DisjointClasses>
<DisjointClasses>
    <Class IRI="DistributionLine"/>
    <Class IRI="DistributionPole"/>
</DisjointClasses>
<DisjointClasses>
    <Class IRI="Drought"/>
    <Class IRI="Earthquake"/>
    <Class IRI="ExtremeTemperatureCold"/>
    <Class IRI="ExtremeTemperatureHot"/>
    <Class IRI="Flood"/>
    <Class IRI="GroundCollapse"/>
    <Class IRI="GroundHeave"/>
    <Class IRI="HailStorm"/>
    <Class IRI="ImpactEvent"/>
    <Class IRI="Landslide"/>
    <Class IRI="Lightning"/>
    <Class IRI="RegionalSubsidence"/>
    <Class IRI="SnowAvalanche"/>
    <Class IRI="SnowStorm"/>
    <Class IRI="SoilLocalSubsidence"/>
    <Class IRI="Storm"/>
    <Class IRI="Tsunami"/>
    <Class IRI="VolcanicEruption"/>
```

```
        <Class IRI="Wildfire"/>
    </DisjointClasses>
    <DisjointClasses>
        <Class IRI="Expert"/>
        <Class IRI="Resident"/>
        <Class IRI="SocialMediaUser"/>
    </DisjointClasses>
    <DisjointClasses>
        <Class IRI="Flat"/>
        <Class IRI="House"/>
        <Class IRI="PublicPlace"/>
        <Class IRI="Residence"/>
    </DisjointClasses>
    <DisjointClasses>
        <Class IRI="GeneratingPlant"/>
        <Class IRI="GridNetwork"/>
        <Class IRI="Substation"/>
        <Class IRI="TransmissionNetwork"/>
    </DisjointClasses>
    <DisjointClasses>
        <Class IRI="Hospital"/>
        <Class IRI="School"/>
    </DisjointClasses>
    <DisjointClasses>
        <Class IRI="HumanSensor"/>
        <Class IRI="PhysicalSensor"/>
    </DisjointClasses>
    <DisjointClasses>
        <Class IRI="Lake"/>
        <Class IRI="Mountain"/>
        <Class IRI="River"/>
        <Class IRI="Waterfall"/>
    </DisjointClasses>
    <DisjointClasses>
        <Class IRI="MobileApp"/>
        <Class IRI="SocialMedia"/>
    </DisjointClasses>
    <DisjointClasses>
        <Class IRI="Pylon"/>
        <Class IRI="TransmissionLine"/>
    </DisjointClasses>
    <DisjointClasses>
        <Class IRI="Railway"/>
        <Class IRI="Road"/>
    </DisjointClasses>
    <ClassAssertion>
        <Class IRI="CrackOnSideWalk"/>
        <NamedIndividual IRI="crack_on_sideWalk_2"/>
    </ClassAssertion>
    <ClassAssertion>
        <Class IRI="CrackOnWall"/>
        <NamedIndividual IRI="crack_on_wall_2"/>
    </ClassAssertion>
    <ClassAssertion>
        <Class IRI="PhysicalSensor"/>
        <NamedIndividual IRI="dataSource_1"/>
    </ClassAssertion>
    <ClassAssertion>
        <Class IRI="PhysicalSensor"/>
        <NamedIndividual IRI="dataSource_2"/>
    </ClassAssertion>
    <ClassAssertion>
        <Class IRI="PhysicalSensor"/>
        <NamedIndividual IRI="dataSource_3"/>
    </ClassAssertion>
    <ClassAssertion>
        <Class IRI="DataSourceMetadata"/>
        <NamedIndividual IRI="ds_metadata_1"/>
    </ClassAssertion>
    <ClassAssertion>
        <Class IRI="DataSourceMetadata"/>
        <NamedIndividual IRI="ds_metadata_2"/>
    </ClassAssertion>
    <ClassAssertion>
        <Class IRI="DataSourceMetadata"/>
        <NamedIndividual IRI="ds_metadata_3"/>
    </ClassAssertion>
    <ClassAssertion>
        <Class IRI="Flood"/>
        <NamedIndividual IRI="flood_1"/>
    </ClassAssertion>
    <ClassAssertion>
        <Class IRI="Flood"/>
        <NamedIndividual IRI="flood_2"/>
    </ClassAssertion>
    <ClassAssertion>
        <Class IRI="FoI"/>
        <NamedIndividual IRI="foi_Karela_bbox_1"/>
    </ClassAssertion>
    <ClassAssertion>
        <Class IRI="FoI"/>
        <NamedIndividual IRI="foi_Karela_bbox_2"/>
    </ClassAssertion>
    <ClassAssertion>
        <Class IRI="GeoLocation"/>
        <NamedIndividual IRI="geolocation_1"/>
```

```
    </ClassAssertion>
    <ClassAssertion>
        <Class IRI="GridNetwork"/>
        <NamedIndividual IRI="gridnetwork_1"/>
    </ClassAssertion>
    <ClassAssertion>
        <Class IRI="Landslide"/>
        <NamedIndividual IRI="landslide_1"/>
    </ClassAssertion>
    <ClassAssertion>
        <Class IRI="Landslide"/>
        <NamedIndividual IRI="landslide_2"/>
    </ClassAssertion>
    <ClassAssertion>
        <Class IRI="LeaningTelephonePole"/>
        <NamedIndividual IRI="leaning_telephone_pole_1"/>
    </ClassAssertion>
    <ClassAssertion>
        <Class IRI="Observation"/>
        <NamedIndividual IRI="obs_1"/>
    </ClassAssertion>
    <ClassAssertion>
        <Class IRI="Observation"/>
        <NamedIndividual IRI="obs_2"/>
    </ClassAssertion>
    <ClassAssertion>
        <Class IRI="ObservedPropertyProfile"/>
        <NamedIndividual IRI="obs_profile_11"/>
    </ClassAssertion>
    <ClassAssertion>
        <Class IRI="ObservedPropertyProfile"/>
        <NamedIndividual IRI="obs_profile_12"/>
    </ClassAssertion>
    <ClassAssertion>
        <Class IRI="ObservedPropertyProfile"/>
        <NamedIndividual IRI="obs_profile_21"/>
    </ClassAssertion>
    <ClassAssertion>
        <Class IRI="ProviderProfile"/>
        <NamedIndividual IRI="provider_profile_1"/>
    </ClassAssertion>
    <ClassAssertion>
        <Class IRI="ProviderProfile"/>
        <NamedIndividual IRI="provider_profile_2"/>
    </ClassAssertion>
    <ClassAssertion>
        <Class IRI="Rain"/>
        <NamedIndividual IRI="rain_1"/>
    </ClassAssertion>
    <ClassAssertion>
        <Class IRI="SensorProfile"/>
        <NamedIndividual IRI="sensor_profile_1"/>
    </ClassAssertion>
    <ClassAssertion>
        <Class IRI="SensorProfile"/>
        <NamedIndividual IRI="sensor_profile_2"/>
    </ClassAssertion>
    <ClassAssertion>
        <Class IRI="ServiceProfile"/>
        <NamedIndividual IRI="service_profile_11"/>
    </ClassAssertion>
    <ClassAssertion>
        <Class IRI="ServiceProfile"/>
        <NamedIndividual IRI="service_profile_12"/>
    </ClassAssertion>
    <ClassAssertion>
        <Class IRI="ServiceProfile"/>
        <NamedIndividual IRI="service_profile_21"/>
    </ClassAssertion>
    <ClassAssertion>
        <Class IRI="SoilMoisture"/>
        <NamedIndividual IRI="soil_moisture_1"/>
    </ClassAssertion>
    <ClassAssertion>
        <Class IRI="SoilMovement"/>
        <NamedIndividual IRI="soil_movement_1"/>
    </ClassAssertion>
    <ObjectPropertyAssertion>
        <ObjectProperty IRI="isWarningSignFor"/>
        <NamedIndividual IRI="crack_on_sideWalk_2"/>
        <NamedIndividual IRI="landslide_2"/>
    </ObjectPropertyAssertion>
    <ObjectPropertyAssertion>
        <ObjectProperty IRI="isWarningSignFor"/>
        <NamedIndividual IRI="crack_on_wall_2"/>
        <NamedIndividual IRI="landslide_2"/>
    </ObjectPropertyAssertion>
    <ObjectPropertyAssertion>
        <ObjectProperty IRI="hasDataSourceMetadata"/>
        <NamedIndividual IRI="dataSource_1"/>
        <NamedIndividual IRI="ds_metadata_1"/>
    </ObjectPropertyAssertion>
    <ObjectPropertyAssertion>
        <ObjectProperty IRI="isDataSourceFor"/>
        <NamedIndividual IRI="dataSource_1"/>
        <NamedIndividual IRI="soil_moisture_1"/>
```

```
    </ObjectPropertyAssertion>
    <ObjectPropertyAssertion>
        <ObjectProperty IRI="hasDataSourceMetadata"/>
        <NamedIndividual IRI="dataSource_2"/>
        <NamedIndividual IRI="ds_metadata_2"/>
    </ObjectPropertyAssertion>
    <ObjectPropertyAssertion>
        <ObjectProperty IRI="isDataSourceFor"/>
        <NamedIndividual IRI="dataSource_2"/>
        <NamedIndividual IRI="soil_movement_1"/>
    </ObjectPropertyAssertion>
    <ObjectPropertyAssertion>
        <ObjectProperty IRI="hasDataSourceMetadata"/>
        <NamedIndividual IRI="dataSource_3"/>
        <NamedIndividual IRI="ds_metadata_3"/>
    </ObjectPropertyAssertion>
    <ObjectPropertyAssertion>
        <ObjectProperty IRI="isDataSourceFor"/>
        <NamedIndividual IRI="dataSource_3"/>
        <NamedIndividual IRI="rain_1"/>
    </ObjectPropertyAssertion>
    <ObjectPropertyAssertion>
        <ObjectProperty IRI="hasProfile"/>
        <NamedIndividual IRI="ds_metadata_1"/>
        <NamedIndividual IRI="obs_profile_11"/>
    </ObjectPropertyAssertion>
    <ObjectPropertyAssertion>
        <ObjectProperty IRI="hasProfile"/>
        <NamedIndividual IRI="ds_metadata_1"/>
        <NamedIndividual IRI="provider_profile_1"/>
    </ObjectPropertyAssertion>
    <ObjectPropertyAssertion>
        <ObjectProperty IRI="hasProfile"/>
        <NamedIndividual IRI="ds_metadata_1"/>
        <NamedIndividual IRI="sensor_profile_1"/>
    </ObjectPropertyAssertion>
    <ObjectPropertyAssertion>
        <ObjectProperty IRI="hasProfile"/>
        <NamedIndividual IRI="ds_metadata_1"/>
        <NamedIndividual IRI="service_profile_11"/>
    </ObjectPropertyAssertion>
    <ObjectPropertyAssertion>
        <ObjectProperty IRI="hasProfile"/>
        <NamedIndividual IRI="ds_metadata_2"/>
        <NamedIndividual IRI="obs_profile_12"/>
    </ObjectPropertyAssertion>
    <ObjectPropertyAssertion>
        <ObjectProperty IRI="hasProfile"/>
        <NamedIndividual IRI="ds_metadata_2"/>
        <NamedIndividual IRI="provider_profile_1"/>
    </ObjectPropertyAssertion>
    <ObjectPropertyAssertion>
        <ObjectProperty IRI="hasProfile"/>
        <NamedIndividual IRI="ds_metadata_2"/>
        <NamedIndividual IRI="sensor_profile_1"/>
    </ObjectPropertyAssertion>
    <ObjectPropertyAssertion>
        <ObjectProperty IRI="hasProfile"/>
        <NamedIndividual IRI="ds_metadata_2"/>
        <NamedIndividual IRI="service_profile_12"/>
    </ObjectPropertyAssertion>
    <ObjectPropertyAssertion>
        <ObjectProperty IRI="hasProfile"/>
        <NamedIndividual IRI="ds_metadata_3"/>
        <NamedIndividual IRI="obs_profile_21"/>
    </ObjectPropertyAssertion>
    <ObjectPropertyAssertion>
        <ObjectProperty IRI="hasProfile"/>
        <NamedIndividual IRI="ds_metadata_3"/>
        <NamedIndividual IRI="provider_profile_2"/>
    </ObjectPropertyAssertion>
    <ObjectPropertyAssertion>
        <ObjectProperty IRI="hasProfile"/>
        <NamedIndividual IRI="ds_metadata_3"/>
        <NamedIndividual IRI="sensor_profile_2"/>
    </ObjectPropertyAssertion>
    <ObjectPropertyAssertion>
        <ObjectProperty IRI="hasProfile"/>
        <NamedIndividual IRI="ds_metadata_3"/>
        <NamedIndividual IRI="service_profile_21"/>
    </ObjectPropertyAssertion>
    <ObjectPropertyAssertion>
        <ObjectProperty IRI="triggers"/>
        <NamedIndividual IRI="flood_1"/>
        <NamedIndividual IRI="landslide_1"/>
    </ObjectPropertyAssertion>
    <ObjectPropertyAssertion>
        <ObjectProperty IRI="hasGeoLocation"/>
        <NamedIndividual IRI="landslide_1"/>
        <NamedIndividual IRI="geolocation_1"/>
    </ObjectPropertyAssertion>
    <ObjectPropertyAssertion>
        <ObjectProperty IRI="hasGeoLocation"/>
        <NamedIndividual IRI="landslide_2"/>
        <NamedIndividual IRI="geolocation_2"/>
    </ObjectPropertyAssertion>
```

```
<ObjectPropertyAssertion>
    <ObjectProperty IRI="isWarningSignFor"/>
    <NamedIndividual IRI="leaning_telephone_pole_1"/>
    <NamedIndividual IRI="flood_1"/>
</ObjectPropertyAssertion>
<ObjectPropertyAssertion>
    <ObjectProperty IRI="isWarningSignFor"/>
    <NamedIndividual IRI="leaning_telephone_pole_1"/>
    <NamedIndividual IRI="landslide_1"/>
</ObjectPropertyAssertion>
<ObjectPropertyAssertion>
    <ObjectProperty IRI="hasFoI"/>
    <NamedIndividual IRI="obs_1"/>
    <NamedIndividual IRI="foi_Karela_bbox_1"/>
</ObjectPropertyAssertion>
<ObjectPropertyAssertion>
    <ObjectProperty IRI="isObservationFor"/>
    <NamedIndividual IRI="obs_1"/>
    <NamedIndividual IRI="landslide_1"/>
</ObjectPropertyAssertion>
<ObjectPropertyAssertion>
    <ObjectProperty IRI="hasFoI"/>
    <NamedIndividual IRI="obs_2"/>
    <NamedIndividual IRI="foi_Karela_bbox_2"/>
</ObjectPropertyAssertion>
<ObjectPropertyAssertion>
    <ObjectProperty IRI="isObservationFor"/>
    <NamedIndividual IRI="obs_2"/>
    <NamedIndividual IRI="flood_1"/>
</ObjectPropertyAssertion>
<ObjectPropertyAssertion>
    <ObjectProperty IRI="hasFoI"/>
    <NamedIndividual IRI="obs_3"/>
    <NamedIndividual IRI="foi_Karela_bbox_1"/>
</ObjectPropertyAssertion>
<ObjectPropertyAssertion>
    <ObjectProperty IRI="isObservationFor"/>
    <NamedIndividual IRI="obs_3"/>
    <NamedIndividual IRI="gridnetwork_1"/>
</ObjectPropertyAssertion>
<ObjectPropertyAssertion>
    <ObjectProperty IRI="isObservedPropertyFor"/>
    <NamedIndividual IRI="rain_1"/>
    <NamedIndividual IRI="obs_2"/>
</ObjectPropertyAssertion>
<ObjectPropertyAssertion>
    <ObjectProperty IRI="isObservedPropertyFor"/>
    <NamedIndividual IRI="soil_moisture_1"/>
    <NamedIndividual IRI="obs_1"/>
</ObjectPropertyAssertion>
<ObjectPropertyAssertion>
    <ObjectProperty IRI="isObservedPropertyFor"/>
    <NamedIndividual IRI="soil_moisture_1"/>
    <NamedIndividual IRI="obs_3"/>
</ObjectPropertyAssertion>
<ObjectPropertyAssertion>
    <ObjectProperty IRI="isObservedPropertyFor"/>
    <NamedIndividual IRI="soil_movement_1"/>
    <NamedIndividual IRI="obs_1"/>
</ObjectPropertyAssertion>
<DataPropertyAssertion>
    <DataProperty IRI="bbox"/>
    <NamedIndividual IRI="geolocation_1"/>
    <Literal datatypeIRI="http://www.w3.org/1999/02/22-rdf-syntax-ns#PlainLiteral">[2553624.5, 6988200.5
</DataPropertyAssertion>
<DataPropertyAssertion>
    <DataProperty IRI="bbox"/>
    <NamedIndividual IRI="geolocation_2"/>
    <Literal datatypeIRI="http://www.w3.org/1999/02/22-rdf-syntax-ns#PlainLiteral">[1543624.5,7988200.5,
</DataPropertyAssertion>
<DataPropertyAssertion>
    <DataProperty IRI="latitude"/>
    <NamedIndividual IRI="geolocation_2"/>
    <Literal datatypeIRI="http://www.w3.org/2001/XMLSchema#decimal">8.434</Literal>
</DataPropertyAssertion>
<DataPropertyAssertion>
    <DataProperty IRI="longitude"/>
    <NamedIndividual IRI="geolocation_2"/>
    <Literal datatypeIRI="http://www.w3.org/2001/XMLSchema#decimal">44.295</Literal>
</DataPropertyAssertion>
<DataPropertyAssertion>
    <DataProperty IRI="featureOfInterest"/>
    <NamedIndividual IRI="obs_profile_11"/>
    <Literal datatypeIRI="http://www.w3.org/1999/02/22-rdf-syntax-ns#PlainLiteral">foi_karela_bbox_1</Li
</DataPropertyAssertion>
<DataPropertyAssertion>
    <DataProperty IRI="observedPropertyName"/>
    <NamedIndividual IRI="obs_profile_11"/>
    <Literal datatypeIRI="http://www.w3.org/1999/02/22-rdf-syntax-ns#PlainLiteral">soil_moisture_1</Lite
</DataPropertyAssertion>
<DataPropertyAssertion>
    <DataProperty IRI="observedPropertyType"/>
    <NamedIndividual IRI="obs_profile_11"/>
    <Literal datatypeIRI="http://www.w3.org/1999/02/22-rdf-syntax-ns#PlainLiteral">SoilMoisture</Literal>
</DataPropertyAssertion>
<DataPropertyAssertion>
```

```
        <DataProperty IRI="phenomenonBeginTime"/>
        <NamedIndividual IRI="obs_profile_11"/>
        <Literal datatypeIRI="http://www.w3.org/2001/XMLSchema#dateTime">2004-01-01T00:00:00</Literal>
    </DataPropertyAssertion>
    <DataPropertyAssertion>
        <DataProperty IRI="phenomenonEndTime"/>
        <NamedIndividual IRI="obs_profile_11"/>
        <Literal datatypeIRI="http://www.w3.org/2001/XMLSchema#dateTime">2018-07-21T00:00:00</Literal>
    </DataPropertyAssertion>
    <DataPropertyAssertion>
        <DataProperty IRI="featureOfInterest"/>
        <NamedIndividual IRI="obs_profile_12"/>
        <Literal datatypeIRI="http://www.w3.org/1999/02/22-rdf-syntax-ns#PlainLiteral">foi_karela_bbox_1</Li
    </DataPropertyAssertion>
    <DataPropertyAssertion>
        <DataProperty IRI="observedPropertyName"/>
        <NamedIndividual IRI="obs_profile_12"/>
        <Literal datatypeIRI="http://www.w3.org/1999/02/22-rdf-syntax-ns#PlainLiteral">soil_movement_1</Lite
    </DataPropertyAssertion>
    <DataPropertyAssertion>
        <DataProperty IRI="observedPropertyType"/>
        <NamedIndividual IRI="obs_profile_12"/>
        <Literal datatypeIRI="http://www.w3.org/1999/02/22-rdf-syntax-ns#PlainLiteral">SoilMovement</Literal>
    </DataPropertyAssertion>
    <DataPropertyAssertion>
        <DataProperty IRI="phenomenonBeginTime"/>
        <NamedIndividual IRI="obs_profile_12"/>
        <Literal datatypeIRI="http://www.w3.org/2001/XMLSchema#dateTime">2004-01-01T00:00:00</Literal>
    </DataPropertyAssertion>
    <DataPropertyAssertion>
        <DataProperty IRI="phenomenonEndTime"/>
        <NamedIndividual IRI="obs_profile_12"/>
        <Literal datatypeIRI="http://www.w3.org/2001/XMLSchema#dateTime">2018-07-21T00:00:00</Literal>
    </DataPropertyAssertion>
    <DataPropertyAssertion>
        <DataProperty IRI="featureOfInterest"/>
        <NamedIndividual IRI="obs_profile_21"/>
        <Literal datatypeIRI="http://www.w3.org/1999/02/22-rdf-syntax-ns#PlainLiteral">foi_karela_bbox_2</Li
    </DataPropertyAssertion>
    <DataPropertyAssertion>
        <DataProperty IRI="observedPropertyName"/>
        <NamedIndividual IRI="obs_profile_21"/>
        <Literal datatypeIRI="http://www.w3.org/1999/02/22-rdf-syntax-ns#PlainLiteral">rain_1</Literal>
    </DataPropertyAssertion>
    <DataPropertyAssertion>
        <DataProperty IRI="observedPropertyType"/>
        <NamedIndividual IRI="obs_profile_21"/>
        <Literal datatypeIRI="http://www.w3.org/1999/02/22-rdf-syntax-ns#PlainLiteral">Rain</Literal>
    </DataPropertyAssertion>
    <DataPropertyAssertion>
        <DataProperty IRI="phenomenonBeginTime"/>
        <NamedIndividual IRI="obs_profile_21"/>
        <Literal datatypeIRI="http://www.w3.org/2001/XMLSchema#dateTime">2004-01-01T00:00:00</Literal>
    </DataPropertyAssertion>
    <DataPropertyAssertion>
        <DataProperty IRI="phenomenonEndTime"/>
        <NamedIndividual IRI="obs_profile_21"/>
        <Literal datatypeIRI="http://www.w3.org/2001/XMLSchema#dateTime">2018-07-21T00:00:00</Literal>
    </DataPropertyAssertion>
    <DataPropertyAssertion>
        <DataProperty IRI="providerName"/>
        <NamedIndividual IRI="provider_profile_1"/>
        <Literal datatypeIRI="http://www.w3.org/1999/02/22-rdf-syntax-ns#PlainLiteral">Amrita</Literal>
    </DataPropertyAssertion>
    <DataPropertyAssertion>
        <DataProperty IRI="providerName"/>
        <NamedIndividual IRI="provider_profile_2"/>
        <Literal datatypeIRI="http://www.w3.org/1999/02/22-rdf-syntax-ns#PlainLiteral">MetOffice</Literal>
    </DataPropertyAssertion>
    <DataPropertyAssertion>
        <DataProperty IRI="eventList"/>
        <NamedIndividual IRI="sensor_profile_1"/>
        <Literal datatypeIRI="http://www.w3.org/1999/02/22-rdf-syntax-ns#PlainLiteral">Landslide</Literal>
    </DataPropertyAssertion>
    <DataPropertyAssertion>
        <DataProperty IRI="featureList"/>
        <NamedIndividual IRI="sensor_profile_1"/>
        <Literal datatypeIRI="http://www.w3.org/1999/02/22-rdf-syntax-ns#PlainLiteral">foi_karela_bbox_1</Li
    </DataPropertyAssertion>
    <DataPropertyAssertion>
        <DataProperty IRI="sensorType"/>
        <NamedIndividual IRI="sensor_profile_1"/>
        <Literal datatypeIRI="http://www.w3.org/1999/02/22-rdf-syntax-ns#PlainLiteral">in-situ</Literal>
    </DataPropertyAssertion>
    <DataPropertyAssertion>
        <DataProperty IRI="eventList"/>
        <NamedIndividual IRI="sensor_profile_2"/>
        <Literal datatypeIRI="http://www.w3.org/1999/02/22-rdf-syntax-ns#PlainLiteral">Flood</Literal>
    </DataPropertyAssertion>
    <DataPropertyAssertion>
        <DataProperty IRI="featureOfInterest"/>
        <NamedIndividual IRI="sensor_profile_2"/>
        <Literal datatypeIRI="http://www.w3.org/1999/02/22-rdf-syntax-ns#PlainLiteral">foi_karela_bbox_1</Li
    </DataPropertyAssertion>
    <DataPropertyAssertion>
        <DataProperty IRI="sensorType"/>
```

```
        <NamedIndividual IRI="sensor_profile_2"/>
        <Literal datatypeIRI="http://www.w3.org/1999/02/22-rdf-syntax-ns#PlainLiteral">in-situ</Literal>
    </DataPropertyAssertion>
    <DataPropertyAssertion>
        <DataProperty IRI="column"/>
        <NamedIndividual IRI="service_profile_11"/>
        <Literal datatypeIRI="http://www.w3.org/1999/02/22-rdf-syntax-ns#PlainLiteral">*</Literal>
    </DataPropertyAssertion>
    <DataPropertyAssertion>
        <DataProperty IRI="dbname"/>
        <NamedIndividual IRI="service_profile_11"/>
        <Literal datatypeIRI="http://www.w3.org/1999/02/22-rdf-syntax-ns#PlainLiteral">mysql</Literal>
    </DataPropertyAssertion>
    <DataPropertyAssertion>
        <DataProperty IRI="serviceAdapter"/>
        <NamedIndividual IRI="service_profile_11"/>
        <Literal datatypeIRI="http://www.w3.org/1999/02/22-rdf-syntax-ns#PlainLiteral">Rest_adaptor_11</Lite
    </DataPropertyAssertion>
    <DataPropertyAssertion>
        <DataProperty IRI="serviceProvider"/>
        <NamedIndividual IRI="service_profile_11"/>
        <Literal datatypeIRI="http://www.w3.org/1999/02/22-rdf-syntax-ns#PlainLiteral">Amrita</Literal>
    </DataPropertyAssertion>
    <DataPropertyAssertion>
        <DataProperty IRI="serviceType"/>
        <NamedIndividual IRI="service_profile_11"/>
        <Literal datatypeIRI="http://www.w3.org/1999/02/22-rdf-syntax-ns#PlainLiteral">REST</Literal>
    </DataPropertyAssertion>
    <DataPropertyAssertion>
        <DataProperty IRI="serviceURL"/>
        <NamedIndividual IRI="service_profile_11"/>
        <Literal datatypeIRI="http://www.w3.org/1999/02/22-rdf-syntax-ns#PlainLiteral">http://127.0.0.1/rest/
    </DataPropertyAssertion>
    <DataPropertyAssertion>
        <DataProperty IRI="table"/>
        <NamedIndividual IRI="service_profile_11"/>
        <Literal datatypeIRI="http://www.w3.org/1999/02/22-rdf-syntax-ns#PlainLiteral">NorthSlopHigherRawData
    </DataPropertyAssertion>
    <DataPropertyAssertion>
        <DataProperty IRI="where"/>
        <NamedIndividual IRI="service_profile_11"/>
        <Literal datatypeIRI="http://www.w3.org/1999/02/22-rdf-syntax-ns#PlainLiteral"></Literal>
    </DataPropertyAssertion>
    <DataPropertyAssertion>
        <DataProperty IRI="collection"/>
        <NamedIndividual IRI="service_profile_12"/>
        <Literal datatypeIRI="http://www.w3.org/1999/02/22-rdf-syntax-ns#PlainLiteral">BGS-north-slope-higher
    </DataPropertyAssertion>
    <DataPropertyAssertion>
        <DataProperty IRI="column"/>
        <NamedIndividual IRI="service_profile_12"/>
        <Literal datatypeIRI="http://www.w3.org/1999/02/22-rdf-syntax-ns#PlainLiteral">*</Literal>
    </DataPropertyAssertion>
    <DataPropertyAssertion>
        <DataProperty IRI="dbname"/>
        <NamedIndividual IRI="service_profile_12"/>
        <Literal datatypeIRI="http://www.w3.org/1999/02/22-rdf-syntax-ns#PlainLiteral">mogodb</Literal>
    </DataPropertyAssertion>
    <DataPropertyAssertion>
        <DataProperty IRI="serviceAdapter"/>
        <NamedIndividual IRI="service_profile_12"/>
        <Literal datatypeIRI="http://www.w3.org/1999/02/22-rdf-syntax-ns#PlainLiteral">Rest_adaptor_12</Lite
    </DataPropertyAssertion>
    <DataPropertyAssertion>
        <DataProperty IRI="serviceProvider"/>
        <NamedIndividual IRI="service_profile_12"/>
        <Literal datatypeIRI="http://www.w3.org/1999/02/22-rdf-syntax-ns#PlainLiteral">Amrita</Literal>
    </DataPropertyAssertion>
    <DataPropertyAssertion>
        <DataProperty IRI="serviceType"/>
        <NamedIndividual IRI="service_profile_12"/>
        <Literal datatypeIRI="http://www.w3.org/1999/02/22-rdf-syntax-ns#PlainLiteral">REST</Literal>
    </DataPropertyAssertion>
    <DataPropertyAssertion>
        <DataProperty IRI="serviceURL"/>
        <NamedIndividual IRI="service_profile_12"/>
        <Literal datatypeIRI="http://www.w3.org/1999/02/22-rdf-syntax-ns#PlainLiteral">http://127.0.0.1/rest/
    </DataPropertyAssertion>
    <DataPropertyAssertion>
        <DataProperty IRI="where"/>
        <NamedIndividual IRI="service_profile_12"/>
        <Literal datatypeIRI="http://www.w3.org/1999/02/22-rdf-syntax-ns#PlainLiteral"></Literal>
    </DataPropertyAssertion>
    <DataPropertyAssertion>
        <DataProperty IRI="dbname"/>
        <NamedIndividual IRI="service_profile_21"/>
        <Literal datatypeIRI="http://www.w3.org/1999/02/22-rdf-syntax-ns#PlainLiteral">hdfs</Literal>
    </DataPropertyAssertion>
    <DataPropertyAssertion>
        <DataProperty IRI="path"/>
        <NamedIndividual IRI="service_profile_21"/>
        <Literal datatypeIRI="http://www.w3.org/1999/02/22-rdf-syntax-ns#PlainLiteral">img/1.jpg,img/2.jpg,in
    </DataPropertyAssertion>
    <DataPropertyAssertion>
        <DataProperty IRI="serviceAdapter"/>
        <NamedIndividual IRI="service_profile_21"/>
```

```
        <Literal datatypeIRI="http://www.w3.org/1999/02/22-rdf-syntax-ns#PlainLiteral">Rest_adaptor_21</Lite
    </DataPropertyAssertion>
    <DataPropertyAssertion>
        <DataProperty IRI="serviceProvider"/>
        <NamedIndividual IRI="service_profile_21"/>
        <Literal datatypeIRI="http://www.w3.org/1999/02/22-rdf-syntax-ns#PlainLiteral">MetOffice</Literal>
    </DataPropertyAssertion>
    <DataPropertyAssertion>
        <DataProperty IRI="serviceType"/>
        <NamedIndividual IRI="service_profile_21"/>
        <Literal datatypeIRI="http://www.w3.org/1999/02/22-rdf-syntax-ns#PlainLiteral">REST</Literal>
    </DataPropertyAssertion>
    <DataPropertyAssertion>
        <DataProperty IRI="serviceURL"/>
        <NamedIndividual IRI="service_profile_21"/>
        <Literal datatypeIRI="http://www.w3.org/1999/02/22-rdf-syntax-ns#PlainLiteral">http://127.0.0.1/rest/
    </DataPropertyAssertion>
    <SubObjectPropertyOf>
        <ObjectProperty IRI="hasGeoLocation"/>
        <ObjectProperty abbreviatedIRI="owl:topObjectProperty"/>
    </SubObjectPropertyOf>
    <SubObjectPropertyOf>
        <ObjectProperty IRI="isConnectedTo"/>
        <ObjectProperty abbreviatedIRI="owl:topObjectProperty"/>
    </SubObjectPropertyOf>
    <SubObjectPropertyOf>
        <ObjectProperty IRI="isGeoLocationFor"/>
        <ObjectProperty abbreviatedIRI="owl:topObjectProperty"/>
    </SubObjectPropertyOf>
    <SubObjectPropertyOf>
        <ObjectProperty IRI="livesIn"/>
        <ObjectProperty abbreviatedIRI="owl:topObjectProperty"/>
    </SubObjectPropertyOf>
    <InverseObjectProperties>
        <ObjectProperty IRI="hasDataSource"/>
        <ObjectProperty IRI="isDataSourceFor"/>
    </InverseObjectProperties>
    <InverseObjectProperties>
        <ObjectProperty IRI="hasGeoLocation"/>
        <ObjectProperty IRI="isGeoLocationFor"/>
    </InverseObjectProperties>
    <InverseObjectProperties>
        <ObjectProperty IRI="hasIndicator"/>
        <ObjectProperty IRI="isIndicationOf"/>
    </InverseObjectProperties>
    <InverseObjectProperties>
        <ObjectProperty IRI="hasObservation"/>
        <ObjectProperty IRI="isObservationFor"/>
    </InverseObjectProperties>
    <InverseObjectProperties>
        <ObjectProperty IRI="hasObservedProperty"/>
        <ObjectProperty IRI="isObservedPropertyFor"/>
    </InverseObjectProperties>
    <InverseObjectProperties>
        <ObjectProperty IRI="hasPlace"/>
        <ObjectProperty IRI="isLocatedIn"/>
    </InverseObjectProperties>
    <InverseObjectProperties>
        <ObjectProperty IRI="hasWarningSign"/>
        <ObjectProperty IRI="isWarningSignFor"/>
    </InverseObjectProperties>
    <ObjectPropertyDomain>
        <ObjectProperty IRI="hasDataSourceMetadata"/>
        <Class IRI="DataSource"/>
    </ObjectPropertyDomain>
    <ObjectPropertyDomain>
        <ObjectProperty IRI="hasFoI"/>
        <Class IRI="Observation"/>
    </ObjectPropertyDomain>
    <ObjectPropertyDomain>
        <ObjectProperty IRI="hasInfrastructure"/>
        <Class IRI="UrbanArea"/>
    </ObjectPropertyDomain>
    <ObjectPropertyDomain>
        <ObjectProperty IRI="hasNaturalResource"/>
        <Class IRI="UrbanArea"/>
    </ObjectPropertyDomain>
    <ObjectPropertyDomain>
        <ObjectProperty IRI="hasObservedPropertyProfile"/>
        <Class IRI="ObservationProfile"/>
    </ObjectPropertyDomain>
    <ObjectPropertyDomain>
        <ObjectProperty IRI="hasSocialMedia"/>
        <Class IRI="SocialMediaUser"/>
    </ObjectPropertyDomain>
    <ObjectPropertyDomain>
        <ObjectProperty IRI="hasUtility"/>
        <Class IRI="UrbanArea"/>
    </ObjectPropertyDomain>
    <ObjectPropertyDomain>
        <ObjectProperty IRI="isDataSourceFor"/>
        <Class IRI="DataSource"/>
    </ObjectPropertyDomain>
    <ObjectPropertyDomain>
        <ObjectProperty IRI="isIndicationOf"/>
        <Class IRI="WarningSign"/>
```

```
    </ObjectPropertyDomain>
    <ObjectPropertyDomain>
        <ObjectProperty IRI="isLocatedIn"/>
        <Class IRI="Place"/>
    </ObjectPropertyDomain>
    <ObjectPropertyDomain>
        <ObjectProperty IRI="isObservationFor"/>
        <Class IRI="Observation"/>
    </ObjectPropertyDomain>
    <ObjectPropertyDomain>
        <ObjectProperty IRI="isObservedPropertyFor"/>
        <Class IRI="ObservedProperty"/>
    </ObjectPropertyDomain>
    <ObjectPropertyDomain>
        <ObjectProperty IRI="isWarningSignFor"/>
        <Class IRI="WarningSign"/>
    </ObjectPropertyDomain>
    <ObjectPropertyDomain>
        <ObjectProperty IRI="livesIn"/>
        <Class IRI="Person"/>
    </ObjectPropertyDomain>
    <ObjectPropertyDomain>
        <ObjectProperty IRI="observes"/>
        <Class IRI="Person"/>
    </ObjectPropertyDomain>
    <ObjectPropertyDomain>
        <ObjectProperty IRI="triggers"/>
        <Class IRI="NaturalHazard"/>
    </ObjectPropertyDomain>
    <ObjectPropertyRange>
        <ObjectProperty IRI="hasDataSourceMetadata"/>
        <Class IRI="DataSourceMetadata"/>
    </ObjectPropertyRange>
    <ObjectPropertyRange>
        <ObjectProperty IRI="hasFoI"/>
        <Class IRI="FoI"/>
    </ObjectPropertyRange>
    <ObjectPropertyRange>
        <ObjectProperty IRI="hasInfrastructure"/>
        <Class IRI="Infrastructure"/>
    </ObjectPropertyRange>
    <ObjectPropertyRange>
        <ObjectProperty IRI="hasNaturalResource"/>
        <Class IRI="NaturalResource"/>
    </ObjectPropertyRange>
    <ObjectPropertyRange>
        <ObjectProperty IRI="hasObservedPropertyProfile"/>
        <Class IRI="ObservedPropertyProfile"/>
    </ObjectPropertyRange>
    <ObjectPropertyRange>
        <ObjectProperty IRI="hasSocialMedia"/>
        <Class IRI="SocialMediaAccount"/>
    </ObjectPropertyRange>
    <ObjectPropertyRange>
        <ObjectProperty IRI="hasUtility"/>
        <Class IRI="Utility"/>
    </ObjectPropertyRange>
    <ObjectPropertyRange>
        <ObjectProperty IRI="isDataSourceFor"/>
        <Class IRI="ObservedProperty"/>
    </ObjectPropertyRange>
    <ObjectPropertyRange>
        <ObjectProperty IRI="isIndicationOf"/>
        <Class IRI="AnthropogenicProcess"/>
    </ObjectPropertyRange>
    <ObjectPropertyRange>
        <ObjectProperty IRI="isLocatedIn"/>
        <Class IRI="UrbanArea"/>
    </ObjectPropertyRange>
    <ObjectPropertyRange>
        <ObjectProperty IRI="isObservationFor"/>
        <Class IRI="NaturalHazard"/>
    </ObjectPropertyRange>
    <ObjectPropertyRange>
        <ObjectProperty IRI="isObservedPropertyFor"/>
        <Class IRI="Observation"/>
    </ObjectPropertyRange>
    <ObjectPropertyRange>
        <ObjectProperty IRI="isWarningSignFor"/>
        <Class IRI="NaturalHazard"/>
    </ObjectPropertyRange>
    <ObjectPropertyRange>
        <ObjectProperty IRI="livesIn"/>
        <Class IRI="Place"/>
    </ObjectPropertyRange>
    <ObjectPropertyRange>
        <ObjectProperty IRI="observes"/>
        <Class IRI="WarningSign"/>
    </ObjectPropertyRange>
    <ObjectPropertyRange>
        <ObjectProperty IRI="triggers"/>
        <Class IRI="NaturalHazard"/>
    </ObjectPropertyRange>
    <SubDataPropertyOf>
        <DataProperty IRI="account"/>
        <DataProperty abbreviatedIRI="owl:topDataProperty"/>
```

```
    </SubDataPropertyOf>
    <SubDataPropertyOf>
        <DataProperty IRI="mediaType"/>
        <DataProperty abbreviatedIRI="owl:topDataProperty"/>
    </SubDataPropertyOf>
    <SubDataPropertyOf>
        <DataProperty IRI="postdate"/>
        <DataProperty abbreviatedIRI="owl:topDataProperty"/>
    </SubDataPropertyOf>
    <DataPropertyDomain>
        <DataProperty IRI="account"/>
        <Class IRI="SocialMediaAccount"/>
    </DataPropertyDomain>
    <DataPropertyDomain>
        <DataProperty IRI="bbox"/>
        <Class IRI="GeoLocation"/>
    </DataPropertyDomain>
    <DataPropertyDomain>
        <DataProperty IRI="collection"/>
        <Class IRI="ServiceProfile"/>
    </DataPropertyDomain>
    <DataPropertyDomain>
        <DataProperty IRI="column"/>
        <Class IRI="ServiceProfile"/>
    </DataPropertyDomain>
    <DataPropertyDomain>
        <DataProperty IRI="dbname"/>
        <Class IRI="ServiceProfile"/>
    </DataPropertyDomain>
    <DataPropertyDomain>
        <DataProperty IRI="eventList"/>
        <Class IRI="SensorProfile"/>
    </DataPropertyDomain>
    <DataPropertyDomain>
        <DataProperty IRI="featureList"/>
        <Class IRI="SensorProfile"/>
    </DataPropertyDomain>
    <DataPropertyDomain>
        <DataProperty IRI="featureOfInterest"/>
        <Class IRI="ObservedPropertyProfile"/>
    </DataPropertyDomain>
    <DataPropertyDomain>
        <DataProperty IRI="latitude"/>
        <Class IRI="GeoLocation"/>
    </DataPropertyDomain>
    <DataPropertyDomain>
        <DataProperty IRI="longitude"/>
        <Class IRI="GeoLocation"/>
    </DataPropertyDomain>
    <DataPropertyDomain>
        <DataProperty IRI="mediaType"/>
        <Class IRI="SocialMediaAccount"/>
    </DataPropertyDomain>
    <DataPropertyDomain>
        <DataProperty IRI="observedPropertyName"/>
        <Class IRI="ObservedPropertyProfile"/>
    </DataPropertyDomain>
    <DataPropertyDomain>
        <DataProperty IRI="observedPropertyType"/>
        <Class IRI="ObservedPropertyProfile"/>
    </DataPropertyDomain>
    <DataPropertyDomain>
        <DataProperty IRI="path"/>
        <Class IRI="ServiceProfile"/>
    </DataPropertyDomain>
    <DataPropertyDomain>
        <DataProperty IRI="phenomenonBeginTime"/>
        <Class IRI="ObservedPropertyProfile"/>
    </DataPropertyDomain>
    <DataPropertyDomain>
        <DataProperty IRI="phenomenonEndTime"/>
        <Class IRI="ObservedPropertyProfile"/>
    </DataPropertyDomain>
    <DataPropertyDomain>
        <DataProperty IRI="postdate"/>
        <Class IRI="SocialMediaAccount"/>
    </DataPropertyDomain>
    <DataPropertyDomain>
        <DataProperty IRI="procedure"/>
        <Class IRI="SensorProfile"/>
    </DataPropertyDomain>
    <DataPropertyDomain>
        <DataProperty IRI="providerName"/>
        <Class IRI="ProviderProfile"/>
    </DataPropertyDomain>
    <DataPropertyDomain>
        <DataProperty IRI="sensorType"/>
        <Class IRI="SensorProfile"/>
    </DataPropertyDomain>
    <DataPropertyDomain>
        <DataProperty IRI="serviceAdapter"/>
        <Class IRI="ServiceProfile"/>
    </DataPropertyDomain>
    <DataPropertyDomain>
        <DataProperty IRI="serviceProvider"/>
        <Class IRI="ServiceProfile"/>
```

```
    </DataPropertyDomain>
    <DataPropertyDomain>
        <DataProperty IRI="serviceType"/>
        <Class IRI="ServiceProfile"/>
    </DataPropertyDomain>
    <DataPropertyDomain>
        <DataProperty IRI="serviceURL"/>
        <Class IRI="ServiceProfile"/>
    </DataPropertyDomain>
    <DataPropertyDomain>
        <DataProperty IRI="table"/>
        <Class IRI="ServiceProfile"/>
    </DataPropertyDomain>
    <DataPropertyDomain>
        <DataProperty IRI="timeInstanceOrPeriod"/>
        <Class IRI="SensorProfile"/>
    </DataPropertyDomain>
    <DataPropertyDomain>
        <DataProperty IRI="where"/>
        <Class IRI="ServiceProfile"/>
    </DataPropertyDomain>
    <DataPropertyRange>
        <DataProperty IRI="account"/>
        <Datatype abbreviatedIRI="xsd:string"/>
    </DataPropertyRange>
    <DataPropertyRange>
        <DataProperty IRI="bbox"/>
        <Datatype abbreviatedIRI="xsd:string"/>
    </DataPropertyRange>
    <DataPropertyRange>
        <DataProperty IRI="collection"/>
        <Datatype abbreviatedIRI="xsd:string"/>
    </DataPropertyRange>
    <DataPropertyRange>
        <DataProperty IRI="column"/>
        <Datatype abbreviatedIRI="xsd:string"/>
    </DataPropertyRange>
    <DataPropertyRange>
        <DataProperty IRI="dbname"/>
        <Datatype abbreviatedIRI="xsd:string"/>
    </DataPropertyRange>
    <DataPropertyRange>
        <DataProperty IRI="eventList"/>
        <Datatype abbreviatedIRI="xsd:string"/>
    </DataPropertyRange>
    <DataPropertyRange>
        <DataProperty IRI="featureList"/>
        <Datatype abbreviatedIRI="xsd:string"/>
    </DataPropertyRange>
    <DataPropertyRange>
        <DataProperty IRI="featureOfInterest"/>
        <Datatype abbreviatedIRI="xsd:string"/>
    </DataPropertyRange>
    <DataPropertyRange>
        <DataProperty IRI="latitude"/>
        <Datatype abbreviatedIRI="xsd:double"/>
    </DataPropertyRange>
    <DataPropertyRange>
        <DataProperty IRI="longitude"/>
        <Datatype abbreviatedIRI="xsd:double"/>
    </DataPropertyRange>
    <DataPropertyRange>
        <DataProperty IRI="mediaType"/>
        <Datatype abbreviatedIRI="xsd:string"/>
    </DataPropertyRange>
    <DataPropertyRange>
        <DataProperty IRI="observedPropertyName"/>
        <Datatype abbreviatedIRI="xsd:string"/>
    </DataPropertyRange>
    <DataPropertyRange>
        <DataProperty IRI="observedPropertyType"/>
        <Datatype abbreviatedIRI="xsd:string"/>
    </DataPropertyRange>
    <DataPropertyRange>
        <DataProperty IRI="path"/>
        <Datatype abbreviatedIRI="xsd:string"/>
    </DataPropertyRange>
    <DataPropertyRange>
        <DataProperty IRI="phenomenonBeginTime"/>
        <Datatype abbreviatedIRI="xsd:dateTime"/>
    </DataPropertyRange>
    <DataPropertyRange>
        <DataProperty IRI="phenomenonEndTime"/>
        <Datatype abbreviatedIRI="xsd:dateTime"/>
    </DataPropertyRange>
    <DataPropertyRange>
        <DataProperty IRI="postdate"/>
        <Datatype abbreviatedIRI="xsd:string"/>
    </DataPropertyRange>
    <DataPropertyRange>
        <DataProperty IRI="procedure"/>
        <Datatype abbreviatedIRI="xsd:string"/>
    </DataPropertyRange>
    <DataPropertyRange>
        <DataProperty IRI="providerName"/>
        <Datatype abbreviatedIRI="xsd:string"/>
```

```xml
    </DataPropertyRange>
    <DataPropertyRange>
        <DataProperty IRI="sensorType"/>
        <Datatype abbreviatedIRI="xsd:string"/>
    </DataPropertyRange>
    <DataPropertyRange>
        <DataProperty IRI="serviceAdapter"/>
        <Datatype abbreviatedIRI="xsd:string"/>
    </DataPropertyRange>
    <DataPropertyRange>
        <DataProperty IRI="serviceProvider"/>
        <Datatype abbreviatedIRI="xsd:string"/>
    </DataPropertyRange>
    <DataPropertyRange>
        <DataProperty IRI="serviceType"/>
        <Datatype abbreviatedIRI="xsd:string"/>
    </DataPropertyRange>
    <DataPropertyRange>
        <DataProperty IRI="serviceURL"/>
        <Datatype abbreviatedIRI="xsd:string"/>
    </DataPropertyRange>
    <DataPropertyRange>
        <DataProperty IRI="table"/>
        <Datatype abbreviatedIRI="xsd:string"/>
    </DataPropertyRange>
    <DataPropertyRange>
        <DataProperty IRI="timeInstanceOrPeriod"/>
        <Datatype abbreviatedIRI="xsd:dateTime"/>
    </DataPropertyRange>
    <DataPropertyRange>
        <DataProperty IRI="where"/>
        <Datatype abbreviatedIRI="xsd:string"/>
    </DataPropertyRange>
</Ontology>
```

References

[1] N. Bansal, M. Mukherjee, and A. Gairola, "Urban risk management," 02 2013.

[2] UNDP, "Urban risk management," 2010. Available at μhttps://www.undp.org/content/dam/undp/library/crisis prevention/disaster/6Disaster Risk Reduction - Urban Risk Management.pdf.

[3] B. Mihai, I. Savulescu, I. Sandric, and Z. Chitu, "Integration of landslide susceptibility assessment in urban development: a case study in predeal town, romanian carpathians," *Area*, vol. 46, no. 4, pp. 377–388, 2014.

[4] A. S. K. E. D. S. S. D. Duncan, Andrew; Chen, "Rapids: Early warning system for urban flooding and water quality hazards," 2013.

[5] M. Calvello, R. N. d'Orsi, L. Piciullo, N. Paes, M. Magalhaes, and W. A. Lacerda, "The rio de janeiro early warning system for rainfall-induced landslides: Analysis of performance for the years 2010–2013," *International Journal of Disaster Risk Reduction*, vol. 12, pp. 3 – 15, 2015.

[6] UNISDR, "Terminology," 2017. Available at μhttps://www.unisdr.org/we/inform/terminology.

[7] K. Ashton *et al.*, "That 'internet of things' thing," *RFID journal*, vol. 22, no. 7, pp. 97–114, 2009.

[8] A. Sinha, P. Kumar, N. P. Rana, R. Islam, and Y. K. Dwivedi, "Impact of internet of things (iot) in disaster management: a task-technology fit perspective," *Annals of Operations Research*, vol. 283, pp. 759–794, Dec 2019.

[9] T. Comes, B. Mayag, and E. Negre, "Decision support for disaster risk management: Integrating vulnerabilities into early-warning systems," in *Information Systems for Crisis Response and Management in Mediterranean Countries* (C. Hanachi, F. Bénaben, and F. Charoy, eds.), (Cham), pp. 178–191, Springer International Publishing, 2014.

[10] R. Basher, "Global early warning systems for natural hazards: systematic and people-centred," *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, vol. 364, no. 1845, pp. 2167–2182, 2006.

[11] A. Andiojaya and H. Demirhan, "A bagging algorithm for the imputation of missing values in time series," *Expert Systems with Applications*, vol. 129, pp. 10 – 26, 2019.

[12] R. Hidayat, S. Sutanto, A. Hidayah, B. Ridwan, and A. Mulyana, "Development of a landslide early warning system in indonesia," *Geosciences*, vol. 9, p. 451, 10 2019.

[13] N. Muhamad, C. Lim, M. I. H. Reza, and J. J. Pereira, "Urban hazards management: A case study of langat river basin, peninsular malaysia," in *2015 International Conference on Space Science and Communication (IconSpace)*, pp. 438–443, 2015.

[14] V. Hristidis, S.-C. Chen, T. Li, S. Luis, and Y. Deng, "Survey of data management and analysis in disaster situations," *Journal of Systems and Software*, vol. 83, no. 10, pp. 1701 – 1714, 2010.

[15] E. W. Ngai, K.-l. K. Moon, S. Lam, E. S. K. Chin, and S. S. Tao, "Social media models, technologies, and applications," *Industrial management data systems*, vol. 115, no. 5, pp. 769–802, 2015.

[16] O. Aulov and M. Halem, "Human sensor networks for improved modeling of natural disasters," *Proceedings of the IEEE*, vol. 100, pp. 2812–2823, Oct 2012.

[17] L. Li, M. F. Goodchild, and B. Xu, "Spatial, temporal, and socioeconomic patterns in the use of twitter and flickr," *cartography and geographic information science*, vol. 40, no. 2, pp. 61–77, 2013.

[18] Y. Wang, T. Wang, X. Ye, J. Zhu, and J. Lee, "Using social media for emergency response and urban sustainability: A case study of the 2012 beijing rainstorm," *Sustainability*, vol. 8, no. 1, p. 25, 2015.

[19] P. Zhao, K. Qin, X. Ye, Y. Wang, and Y. Chen, "A trajectory clustering approach based on decision graph and data field for detecting hotspots," *International Journal of Geographical Information Science*, vol. 31, no. 6, pp. 1101–1127, 2017.

[20] E. Abana, C. Dayag, V. Valencia, P. Talosig, J. Ratilla, and G. Galat, "Road flood warning system with information dissemination via social media," *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 9, p. 4979, 12 2019.

[21] K. K. Scott and N. A. Errett, "Content, accessibility, and dissemination of disaster information via social media during the 2016 louisiana floods," *Journal of Public Health Management and Practice*, vol. 24, no. 4, 2018.

[22] J. Vongkusolkit and Q. Huang, "Situational awareness extraction: a comprehensive review of social media data classification during natural hazards," *Annals of GIS*, vol. 0, no. 0, pp. 1–24, 2020.

[23] M. C. B. R. J. Yin, A. Lampert and R. Power, "Using social media to enhance emergency situation awareness," *IEEE Intelligent Systems*, vol. 27, no. 6, pp. 52–59, 2012.

[24] H. Zade, K. Shah, V. Rangarajan, P. Kshirsagar, M. Imran, and K. Starbird, "From situational awareness to actionability: Towards improving the utility of social media data for crisis response," *Proc. ACM Hum.-Comput. Interact.*, vol. 2, Nov. 2018.

[25] A. Mukkamala and R. Beck, "Disaster management and social media use for decision making by humanitarian organizations," in *2016 49th Hawaii International Conference on System Sciences (HICSS)*, pp. 1379–1385, 2016.

[26] Z. Wang and X. Ye, "Social media analytics for natural disaster management," *International Journal of Geographical Information Science*, vol. 32, no. 1, pp. 49–72, 2018.

[27] R. Ranjan, K. Mitra, and D. Georgakopoulos, "Mediawise cloud content orchestrator," *Journal of Internet Services and Applications*, vol. 4, p. 2, Jan 2013.

[28] W. OGC, "Semantic sensor network ontology," 2016.

[29] Envision, "Ls ontology," 2016. Available at µ`http://envision.brgm-rec.fr/LS-Ontologies.aspx`.

[30] R. G. Raskin and M. J. Pan, "Knowledge representation in the semantic web for earth and environmental terminology (sweet)," *Computers & Geosciences*, vol. 31, no. 9, pp. 1119 – 1125, 2005. Application of XML in the Geosciences.

[31] R. Ranjan, "Streaming big data processing in datacenter clouds," *IEEE Cloud Computing*, vol. 1, no. 1, pp. 78–83, 2014.

[32] L. Wang and R. Ranjan, "Processing distributed internet of things data in clouds," *IEEE Cloud Computing*, vol. 2, no. 1, pp. 76–80, 2015.

[33] E. Bertino, S. Nepal, and R. Ranjan, "Building sensor-based big data cyberinfrastructures," *IEEE Cloud Computing*, vol. 2, no. 5, pp. 64–69, 2015.

[34] C. for Research on the Epidemiology of Disasters (CRED), "Annual disaster statistical review 2016: the numbers and trends," 2016. Available at µ`https://www.preventionweb.net/go/56027`.

[35] B. T. Pham, D. Tien Bui, H. R. Pourghasemi, P. Indra, and M. B. Dholakia, "Landslide susceptibility assesssment in the uttarakhand area (india) using gis: a comparison study of prediction capability of naïve bayes, multilayer perceptron neural networks, and functional trees methods," *Theoretical and Applied Climatology*, vol. 128, pp. 255–273, Apr 2017.

[36] G. S. of India, "Landslide hazard," 2020. Available at µ`https://www.gsi.gov.in`.

[37] S. Park, H. Lim, B. Tamang, J. Jin, S. Lee, S. Chang, and Y. Kim, "A study on the slope failure monitoring of a model slope by the application of a displacement sensor," *Journal of Sensors*, vol. 2019, p. 7570517, Dec 2019.

[38] A. Adeel, M. Gogate, S. Farooq, C. Ieracitano, K. Dashtipour, H. Larijani, and A. Hussain, *A Survey on the Role of Wireless Sensor Networks and IoT in Disaster Management*, pp. 57–66. Singapore: Springer Singapore, 2019.

[39] M. Y. S. Maryam Mohsin, "10 social media statistics you need to know in 2020 [infographic]," 2020.

[40] Zephoria, "The top 20 valuable facebook statistics – updated may 2020," 2020.

[41] Twitter, "Quarterly results," 2020. Available at μ`https://investor.twitterinc.com/financial-information/quarterly-results/default.aspx`.

[42] N. G. Abdulhamid, D. A. Ayoung, A. Kashefi, and B. Sigweni, "A survey of social media use in emergency situations: A literature review," *Information Development*, vol. 0, no. 0, p. 0266666920913894, 0.

[43] T. H. Nazer, G. Xue, Y. Ji, and H. Liu, "Intelligent disaster response via social media analysis a survey," *SIGKDD Explor. Newsl.*, vol. 19, p. 46–59, Sept. 2017.

[44] Z. Wang and X. Ye, "Social media analytics for natural disaster management," *International Journal of Geographical Information Science*, vol. 32, no. 1, pp. 49–72, 2018.

[45] A. Saroj and S. Pal, "Use of social media in crisis management: A survey," *International Journal of Disaster Risk Reduction*, vol. 48, p. 101584, 2020.

[46] K. Munir and M. S. Anjum, "The use of ontologies for effective knowledge modelling and information retrieval," *Applied Computing and Informatics*, vol. 14, no. 2, pp. 116 – 126, 2018.

[47] O. Goonetilleke, T. Sellis, X. Zhang, and S. Sathe, "Twitter analytics: a big data management perspective," *ACM SIGKDD Explorations Newsletter*, vol. 16, no. 1, pp. 11–20, 2014.

[48] V. Tablan, K. Bontcheva, I. Roberts, and H. Cunningham, "Mímir: An open-source semantic search framework for interactive information seeking and discovery," *Web Semantics: Science, Services and Agents on the World Wide Web*, vol. 30, pp. 52–68, 2015.

[49] D. Maynard, I. Roberts, M. A. Greenwood, D. Rout, and K. Bontcheva, "A framework for real-time semantic social media analysis," *Web Semantics: Science, Services and Agents on the World Wide Web*, vol. 44, pp. 75–88, 2017.

[50] A. Khoshkbarforoushha and R. Ranjan, "Resource and performance distribution prediction for large scale analytics queries," in *Proceedings of the 7th ACM/SPEC on International Conference on Performance Engineering*, ICPE '16, (New York, NY, USA), p. 49–54, Association for Computing Machinery, 2016.

[51] M. Wang, R. Ranjan, P. P. Jayaraman, P. Strazdins, P. Burnap, O. Rana, and D. Georgakopulos, "A case for understanding end-to-end performance of topic detection and tracking based big data applications in the cloud," in *Internet of Things. IoT Infrastructures* (B. Mandler, J. Marquez-Barja, M. E. Mitre Campista, D. Cagáňová, H. Chaouchi, S. Zeadally, M. Badra, S. Giordano, M. Fazio, A. Somov, and R.-L. Vieriu, eds.), (Cham), pp. 315–325, Springer International Publishing, 2016.

[52] M. Armbrust, R. S. Xin, C. Lian, Y. Huai, D. Liu, J. K. Bradley, X. Meng, T. Kaftan, M. J. Franklin, A. Ghodsi, and M. Zaharia, "Spark sql: Relational data processing in spark," in *Proceedings of the 2015 ACM SIGMOD International Conference on Management of Data*, SIGMOD '15, (New York, NY, USA), p. 1383–1394, Association for Computing Machinery, 2015.

[53] M. FERNÁNDEZ-LÓPEZ and A. GÓMEZ-PÉREZ, "Overview and analysis of methodologies for building ontologies," *The Knowledge Engineering Review*, vol. 17, no. 2, p. 129–156, 2002.

[54] M. Fernández-López, A. Gómez-Pérez, and N. Juristo, "Methontology: From ontological art towards ontological engineering," in *Proceedings of the Ontological Engineering AAAI-97 Spring Symposium Series*, American Asociation for Artificial Intelligence, March 1997. Ontology Engineering Group ? OEG.

[55] D. Wang, M. T. Amin, S. Li, T. Abdelzaher, L. Kaplan, S. Gu, C. Pan, H. Liu, C. C. Aggarwal, R. Ganti, X. Wang, P. Mohapatra, B. Szymanski, and H. Le, "Using humans as sensors: An estimation-theoretic perspective," in *IPSN-14 Proceedings of the 13th International Symposium on Information Processing in Sensor Networks*, pp. 35–46, April 2014.

[56] J. Yick, B. Mukherjee, and D. Ghosal, "Wireless sensor network survey," *Computer Networks*, vol. 52, no. 12, pp. 2292 – 2330, 2008.

[57] H. Shen, "Discussion and analysis of the crowdsourcing mode of public participation in emergency management," in *2015 8th International Symposium on Computational Intelligence and Design (ISCID)*, vol. 2, pp. 610–613, Dec 2015.

[58] A. Immonen, P. Pääkkönen, and E. Ovaska, "Evaluating the quality of social media data in big data architecture," *IEEE Access*, vol. 3, pp. 2028–2043, 2015.

[59] A. Steinberg, C. Wukich, and H. Wu, "Central social media actors in disaster information networks," *International journal of mass emergencies and disasters*, vol. 34, pp. 47–74, 03 2016.

[60] J. Rogstadius, M. Vukovic, C. A. Teixeira, V. Kostakos, E. Karapanos, and J. A. Laredo, "Crisistracker: Crowdsourced social media curation for disaster awareness," *IBM Journal of Research and Development*, vol. 57, pp. 4:1–4:13, Sep. 2013.

[61] S. E. Middleton, L. Middleton, and S. Modafferi, "Real-time crisis mapping of natural disasters using social media," *IEEE Intelligent Systems*, vol. 29, pp. 9–17, Mar 2014.

[62] S. L. Kuriakose, G. Sankar, and C. Muraleedharan, "History of landslide susceptibility and a chorology of landslide-prone areas in the western ghats of kerala, india," *Environmental Geology*, vol. 57, pp. 1553–1568, Jun 2009.

[63] A. Olteanu, C. Castillo, F. Diaz, and S. Vieweg, "Crisislex: A lexicon for collecting and filtering microblogged communications in crises," in *Proceedings of the Eighth International Conference on Weblogs and Social Media, ICWSM 2014,*

Ann Arbor, Michigan, USA, June 1-4, 2014 (E. Adar, P. Resnick, M. D. Choudhury, B. Hogan, and A. H. Oh, eds.), The AAAI Press, 2014.

[64] H. To, S. Agrawal, S. H. Kim, and C. Shahabi, "On identifying disaster-related tweets: Matching-based or learning-based?," *CoRR*, vol. abs/1705.02009, 2017.

[65] S. Lomborg and A. Bechmann, "Using apis for data collection on social media," *The Information Society*, vol. 30, no. 4, pp. 256–265, 2014.

[66] F. Eight, "Data for everyone," 2019. Available at µ`https://www.figure-eight.com/data-for-everyone/`.

[67] M. Imran, S. Elbassuoni, C. Castillo, F. Diaz, and P. Meier, "Practical extraction of disaster-relevant information from social media," in *Proceedings of the 22nd International Conference on World Wide Web*, pp. 1021–1024, ACM, 2013.

[68] B. Takahashi, E. C. Tandoc, and C. Carmichael, "Communicating on twitter during a disaster: An analysis of tweets during typhoon haiyan in the philippines," *Computers in Human Behavior*, vol. 50, pp. 392 – 398, 2015.

[69] K. Rudra, N. Ganguly, P. Goyal, and S. Ghosh, "Extracting and summarizing situational information from the twitter social media during disasters," *ACM Trans. Web*, vol. 12, July 2018.

[70] G. Zamarreño-Aramendia, F. J. Cristòfol, J. De-San-eugenio vela, and X. Ginesta, "Social-media analysis for disaster prevention: Forest fire in artenara and valleseco, Canary Islands," *Journal of Open Innovation: Technology, Market, and Complexity*, vol. 6, no. 4, pp. 1–18, 2020.

[71] Z. Cheng, J. Caverlee, and K. Lee, "You are where you tweet: a content-based approach to geo-locating twitter users," in *Proceedings of the 19th ACM international conference on Information and knowledge management*, pp. 759–768, ACM, 2010.

[72] R. Li, S. Wang, H. Deng, R. Wang, and K. C.-C. Chang, "Towards social user profiling: unified and discriminative influence model for inferring home locations," in *Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining*, pp. 1023–1031, ACM, 2012.

[73] A. Crooks, A. Croitoru, A. Stefanidis, and J. Radzikowski, "# earthquake: Twitter as a distributed sensor system," *Transactions in GIS*, vol. 17, no. 1, pp. 124–147, 2013.

[74] J. P. De Albuquerque, B. Herfort, A. Brenning, and A. Zipf, "A geographic approach for combining social media and authoritative data towards identifying useful information for disaster management," *International Journal of Geographical Information Science*, vol. 29, no. 4, pp. 667–689, 2015.

[75] X. Guan and C. Chen, "Using social media data to understand and assess disasters," *Natural hazards*, vol. 74, no. 2, pp. 837–850, 2014.

[76] J. Chae, D. Thom, Y. Jang, S. Kim, T. Ertl, and D. S. Ebert, "Public behavior response analysis in disaster events utilizing visual analytics of microblog data," *Computers & Graphics*, vol. 38, pp. 51–60, 2014.

[77] Y. Kryvasheyeu, H. Chen, N. Obradovich, E. Moro, P. Van Hentenryck, J. Fowler, and M. Cebrian, "Rapid assessment of disaster damage using social media activity," *Science advances*, vol. 2, no. 3, p. e1500779, 2016.

[78] J. Yin, S. Karimi, A. Lampert, M. Cameron, B. Robinson, and R. Power, "Using social media to enhance emergency situation awareness," in *Twenty-Fourth International Joint Conference on Artificial Intelligence*, 2015.

[79] Z. Wang, X. Ye, and M.-H. Tsou, "Spatial, temporal, and content analysis of twitter for wildfire hazards," *Natural Hazards*, vol. 83, no. 1, pp. 523–540, 2016.

[80] C. Granell and F. O. Ostermann, "Beyond data collection: Objectives and methods of research using vgi and geo-social media for disaster management," *Computers, Environment and Urban Systems*, vol. 59, pp. 231–243, 2016.

[81] Q. Huang and Y. Xiao, "Geographic situational awareness: mining tweets for disaster preparedness, emergency response, impact, and recovery," *ISPRS International Journal of Geo-Information*, vol. 4, no. 3, pp. 1549–1568, 2015.

[82] kinsta, "Wild and interesting facebook statistics and facts (2020)," 2020. Available at μ`https://kinsta.com/blog/facebook-statistics/`.

[83] H. Cunningham, V. Tablan, A. Roberts, and K. Bontcheva, "Getting more out of biomedical documents with gate's full lifecycle open source text analytics," *PLoS computational biology*, vol. 9, no. 2, p. e1002854, 2013.

[84] GATE, "The gate cloud paralleliser (gcp)," 2018. Available at μ`https://gate.ac.uk/gcp/`.

[85] H. Cunningham, V. Tablan, I. Roberts, M. A. Greenwood, and N. Aswani, *Information Extraction and Semantic Annotation for Multi-Paradigm Information Management*, pp. 307–327. Berlin, Heidelberg: Springer Berlin Heidelberg, 2011.

[86] Y. Kim, J. Huang, and S. Emery, "Garbage in, garbage out: data collection, quality assessment and reporting standards for social media data use in health research, infodemiology and digital disease detection," *Journal of medical Internet research*, vol. 18, no. 2, 2016.

[87] S. Kar, H. S. Al-Olimat, K. Thirunarayan, V. Shalin, A. Sheth, and S. Parthasarathy, "D-record: Disaster response and relief coordination pipeline," in *Proceedings of the ACM SIGSPATIAL International Workshop on Advances in Resilient and Intelligent Cities (ARIC 2018)*, Association for Computing Machinery, 2018.

[88] S. Wang, "A cybergis framework for the synthesis of cyberinfrastructure, gis, and spatial analysis," *Annals of the Association of American Geographers*, vol. 100, no. 3, pp. 535–557, 2010.

[89] S. Wang, L. Anselin, B. Bhaduri, C. Crosby, M. F. Goodchild, Y. Liu, and T. L. Nyerges, "Cybergis software: a synthetic review and integration roadmap," *International Journal of Geographical Information Science*, vol. 27, no. 11, pp. 2122–2145, 2013.

[90] C. Slamet, A. Rahman, A. Sutedi, W. Darmalaksana, M. A. Ramdhani, and D. S. Maylawati, "Social media-based identifier for natural disaster," in *IOP Conference Series: Materials Science and Engineering*, vol. 288, p. 012039, IOP Publishing, 2018.

[91] D. Yates and S. Paquette, "Emergency knowledge management and social media technologies: A case study of the 2010 haitian earthquake," in *Proceedings of the 73rd ASIS&T Annual Meeting on Navigating Streams in an Information Ecosystem-Volume 47*, p. 42, American Society for Information Science, 2010.

[92] F. Amato, V. Moscato, A. Picariello, and G. Sperli'ì, "Extreme events management using multimedia social networks," *Future Generation Computer Systems*, vol. 94, pp. 444 – 452, 2019.

[93] F. Amato, V. Moscato, A. Picariello, and G. Sperlí, "Multimedia social network modeling: A proposal," in *2016 IEEE Tenth International Conference on Semantic Computing (ICSC)*, pp. 448–453, 2016.

[94] F. Amato, V. Moscato, A. Picariello, and G. Sperlí, "Diffusion algorithms in multimedia social networks: A preliminary model," in *Proceedings of the 2017 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining 2017*, ASONAM '17, (New York, NY, USA), p. 844–851, Association for Computing Machinery, 2017.

[95] L. Derczynski, D. Maynard, G. Rizzo, M. van Erp, G. Gorrell, R. Troncy, J. Petrak, and K. Bontcheva, "Analysis of named entity recognition and linking for tweets," *Information Processing Management*, vol. 51, no. 2, pp. 32 – 49, 2015.

[96] T. Sakaki, M. Okazaki, and Y. Matsuo, "Tweet analysis for real-time event detection and earthquake reporting system development," *IEEE Transactions on Knowledge and Data Engineering*, vol. 25, no. 4, pp. 919–931, 2013.

[97] J. Chae, D. Thom, H. Bosch, Y. Jang, R. Maciejewski, D. S. Ebert, and T. Ertl, "Spatiotemporal social media analytics for abnormal event detection and examination using seasonal-trend decomposition," in *Visual Analytics Science and Technology (VAST), 2012 IEEE Conference on*, pp. 143–152, IEEE, 2012.

[98] T. Reuter and P. Cimiano, "Event-based classification of social media streams," in *Proceedings of the 2nd ACM International Conference on Multimedia Retrieval*, p. 22, ACM, 2012.

[99] H. Becker, M. Naaman, and L. Gravano, "Learning similarity metrics for event identification in social media," in *Proceedings of the third ACM international conference on Web search and data mining*, pp. 291–300, ACM, 2010.

[100] N. R. Adam, B. Shafiq, and R. Staffin, "Spatial computing and social media in the context of disaster management," *IEEE Intelligent Systems*, vol. 27, no. 6, pp. 90–96, 2012.

[101] Z. Xu, H. Zhang, V. Sugumaran, K.-K. R. Choo, L. Mei, and Y. Zhu, "Participatory sensing-based semantic and spatial analysis of urban emergency events using mobile social media," *EURASIP Journal on Wireless Communications and Networking*, vol. 2016, no. 1, p. 44, 2016.

[102] R. Albtoush, R. Dobrescu, and F. Ionescou, "A hierarchical model for emergency management systems," *University" Politehnica" of Bucharest Scientific Bulletin, Series C: Electrical Engineering*, vol. 73, no. 2, pp. 53–62, 2011.

[103] C. for Research on the Epidemiology of Disasters CRED, "Em-dat the international disaster database." Available at μ`https://www.emdat.be/classification`.

[104] H. Ke, D. Chen, T. Shah, X. Liu, X. Zhang, L. Zhang, and X. Li, "Cloud-aided online eeg classification system for brain healthcare: A case study of depression evaluation with a lightweight cnn," *Software: Practice and Experience*, vol. 0, no. 0.

[105] Y. Tang, D. Chen, L. Wang, A. Y. Zomaya, J. Chen, and H. Liu, "Bayesian tensor factorization for multi-way analysis of multi-dimensional eeg," *Neurocomputing*, vol. 318, pp. 162 – 174, 2018.

[106] D. Chen, Y. Hu, L. Wang, A. Y. Zomaya, and X. Li, "H-parafac: Hierarchical parallel factor analysis of multidimensional big data," *IEEE Transactions on Parallel and Distributed Systems*, vol. 28, pp. 1091–1104, April 2017.

[107] D. Chen, X. Li, L. Wang, S. U. Khan, J. Wang, K. Zeng, and C. Cai, "Fast and scalable multi-way analysis of massive neural data," *IEEE Transactions on Computers*, vol. 64, pp. 707–719, March 2015.

[108] C. van Westen, *Landslide risk assessments for decision - making*, pp. 67–71. The World Bank, 7 2012.

[109] M. Kibanov, G. Stumme, I. Amin, and J. G. Lee, "Mining social media to inform peatland fire and haze disaster management," *CoRR*, vol. abs/1706.05406, 2017.

[110] F. E. Horita, J. a. P. d. Albuquerque, L. C. Degrossi, E. M. Mendiondo, and J. Ueyama, "Development of a spatial decision support system for flood risk management in brazil that combines volunteered geographic information with wireless sensor networks," *Comput. Geosci.*, vol. 80, pp. 84–94, July 2015.

[111] D. George, "Understanding structural and semantic heterogeneity in the context of database schema integration," 2006.

[112] W3C, "W3c semantic web activity," 2013. Available at μ`https://www.w3.org/2001/sw/`.

[113] W3C, "W3c data activity building the web of data," 2013. Available at µ`https://www.w3.org/2013/data/`.

[114] W3C, "Resource description framework (rdf): Concepts and abstract syntax," 2004. Available at µ`https://www.w3.org/TR/2004/REC-rdf-concepts-20040210/`.

[115] W3C, "Web ontology language (owl)," 2012. Available at µ`https://www.w3.org/OWL/`.

[116] J.-P. Calbimonte, H. Jeung, O. Corcho, and K. Aberer, "Semantic sensor data search in a large-scale federated sensor network," *Proceedings of the 4th International Workshop on Semantic Sensor Networks*, vol. 839, pp. 23–38, 2011.

[117] A. Laurent Lefort (CSIRO and W. S. S. N. I. Group), "Ontology for quantity kinds and units: units and quantities definitions," 2010.

[118] S. Lacasse, F. Nadim, S. Lacasse, and F. Nadim, *Landslide Risk Assessment and Mitigation Strategy*, pp. 31–61. Berlin, Heidelberg: Springer Berlin Heidelberg, 2009.

[119] H. Knublauch, "Ontology-driven software development in the context of the semantic web: An example scenario with," in *in Annex XVII (7) , and*, 2004.

[120] "Hazards and the built environment: Attaining built-in resilience," *Disaster Prevention and Management: An International Journal*, vol. 20, no. 2, pp. 215–216, 2011.

[121] M. C. Suárez-Figueroa, A. Gómez-Pérez, and M. Fernández-López, *The NeOn Methodology for Ontology Engineering*, pp. 9–34. Berlin, Heidelberg: Springer Berlin Heidelberg, 2012.

[122] J. Gill and B. Malamud, "Hazard interactions and interaction networks (cascades) within multi-hazard methodologies," *Earth System Dynamics*, vol. 7, pp. 659–679, 8 2016.

[123] J. C. Gill and B. D. Malamud, "Anthropogenic processes, natural hazards, and interactions in a multi-hazard framework," *Earth-Science Reviews*, vol. 166, pp. 246 – 269, 2017.

[124] OGC, "Observations and measurements," 2011. Available at µ`http://www.opengeospatial.org/standards/om`.

[125] OGC, "Sensor model language (sensorml)," 2011. Available at µ`http://www.opengeospatial.org/standards/sensorml`.

[126] OGC, "Sensor observation service," 2011. Available at µ`http://www.opengeospatial.org/standards/sos`.

[127] Y. Hong and R. F. Adler, "Towards an early-warning system for global landslides triggered by rainfall and earthquake," *International Journal of Remote Sensing*, vol. 28, no. 16, pp. 3713–3719, 2007.

[128] C. Pennington, K. Freeborough, C. Dashwood, T. Dijkstra, and K. Lawrie, "The national landslide database of great britain: Acquisition, communication and the role of social media," *Geomorphology*, vol. 249, pp. 44 – 51, 2015. Geohazard Databases: Concepts, Development, Applications.

[129] Q. Wei and Z. Jin, "Service discovery for internet of things: A context-awareness perspective," in *Proceedings of the Fourth Asia-Pacific Symposium on Internetware*, Internetware '12, (New York, NY, USA), pp. 25:1–25:6, ACM, 2012.

[130] M. B. Almeida and R. R. Barbosa, "Ontologies in knowledge management support: A case study," *Journal of the American Society for Information Science and Technology*, vol. 60, no. 10, pp. 2032–2047.

[131] M. Uschold and M. Gruninger, "Ontologies: principles, methods and applications," *The Knowledge Engineering Review*, vol. 11, no. 2, p. 93–136, 1996.

[132] M. Gruninger and M. S. Fox, "The role of competency questions in enterprise engineering," in *PROCEEDINGS OF THE IFIP WG5.7 WORKSHOP ON BENCHMARKING - THEORY AND PRACTICE*, 1994.

[133] Y. Zhen, X. Li, Y. Zhang, L. Zeng, Q. Ou, and X. Yin, "Transmission tower protection system based on internet of things in smart grid," in *2012 7th International Conference on Computer Science Education (ICCSE)*, pp. 863–867, July 2012.

[134] D. M. Ward, "The effect of weather on grid systems and the reliability of electricity supply," *Climatic Change*, vol. 121, no. 1, pp. 103–113, 2013.

[135] C. MacLennan, "Invergarry landslide causes power supply disruption for thousands of islanders," 2018.

[136] T. N. Thailand, "Evacuations considered as landslides fell power poles and close road at egat's lampang mine," 2018. Available at µhttps://www.nationthailand.com/national/30341199.

[137] Y. Yin, F. Wang, and P. Sun, "Landslide hazards triggered by the 2008 wenchuan earthquake, sichuan, china," *Landslides*, vol. 6, no. 2, pp. 139–152, 2009.

[138] M. Butler, M. Angelopoulos, and D. Mahy, "Efficient iot-enabled landslide monitoring," in *2019 IEEE 5th World Forum on Internet of Things (WF-IoT)*, pp. 171–176, April 2019.

[139] G. Mei, N. Xu, J. Qin, B. Wang, and P. Qi, "A survey of internet of things (iot) for geo-hazards prevention: Applications, technologies, and challenges," *IEEE Internet of Things Journal*, vol. 0, pp. 1–1, 2019.

[140] K. Dai, S. Chen, and D. Smith, "Vibration analyses of electrical transmission spun-cast concrete poles for health monitoring," in *Nondestructive Characterization for Composite Materials, Aerospace Engineering, Civil Infrastructure, and Homeland Security 2012* (A. L. Gyekenyesi, ed.), vol. 8347, pp. 15 – 22, International Society for Optics and Photonics, SPIE, 2012.

[141] S. Zahoor and R. N. Mir, “Resource management in pervasive internet of things: A survey,” *Journal of King Saud University - Computer and Information Sciences*, 2018.

[142] T. Baker, E. Ugljanin, N. Faci, M. Sellami, Z. Maamar, and E. Kajan, “Everything as a resource: Foundations and illustration through internet-of-things,” *Computers in Industry*, vol. 94, pp. 62 – 74, 2018.

[143] S. Oueida, Y. Kotb, M. Aloqaily, Y. Jararweh, and T. Baker, “An edge computing based smart healthcare framework for resource management,” *Sensors (Basel, Switzerland)*, vol. 18, p. 4307, Dec 2018. 30563267[pmid].

[144] M. Khalaf, A. J. Hussain, D. Al-Jumeily, T. Baker, R. Keight, P. Lisboa, P. Fergus, and A. S. Al Kafri, “A data science methodology based on machine learning algorithms for flood severity prediction,” in *2018 IEEE Congress on Evolutionary Computation (CEC)*, pp. 1–8, July 2018.

[145] N. Abbas, M. Asim, N. Tariq, T. Baker, and S. Abbas, “A mechanism for securing iot-enabled applications at the fog layer,” *Journal of Sensor and Actuator Networks*, vol. 8, no. 1, 2019.

[146] T. Baker, M. Asim, MacDermott, F. Iqbal, F. Kamoun, B. Shah, O. Alfandi, and M. Hammoudeh, “A secure fog-based platform for scada-based iot critical infrastructure,” *Software: Practice and Experience*, vol. n/a, no. n/a.

[147] S. Sun, W. Song, A. Y. Zomaya, Y. Xiang, K.-K. R. Choo, T. Shah, and L. Wang, “Associative retrieval in spatial big data based on spreading activation with semantic ontology,” *Future Generation Computer Systems*, vol. 76, pp. 499 – 509, 2017.

[148] G. Babitski, S. Bergweiler, O. Grebner, D. Oberle, H. Paulheim, and F. Probst, “Soknos - using semantic technologies in disaster management software,” in *ESWC*, 2011.

[149] S. Liu, C. Brewster, and D. Shaw, “Ontologies for crisis management: a review of state of the art in ontology design and usability,” in *Proceedings of the 10th International ISCRAM Conference ? Baden-Baden, Germany, May 2013*, pp. 349–359, 2013.

[150] Z. Fushen, Z. Shaobo, Y. Simin, W. Chaolin, and H. Quanyi, “Ontology-based representation of meteorological disaster system and its application in emergency management: Illustration with a simulation case study of comprehensive risk assessment,” *Kybernetes*, vol. 45, pp. 798–814, Jan 2016.

[151] F. Guzzetti, S. Peruccacci, M. Rossi, and C. P. Stark, “Rainfall thresholds for the initiation of landslides in central and southern europe,” *Meteorology and Atmospheric Physics*, vol. 98, pp. 239–267, Dec 2007.

[152] T. Teja, A. Dikshit, and N. Satyam, “Determination of rainfall thresholds for landslide prediction using an algorithm-based approach: Case study in the darjeeling himalayas, india,” *Geosciences (Switzerland)*, vol. 9, p. 302, 07 2019.

[153] G. Martelloni, S. Segoni, R. Fanti, and F. Catani, "Rainfall thresholds for the forecasting of landslide occurrence at regional scale," *Landslides*, vol. 9, pp. 485–495, Dec 2012.

[154] D. Schachinger, W. Kastner, and S. Gaida, "Ontology-based abstraction layer for smart grid interaction in building energy management systems," in *2016 IEEE International Energy Conference (ENERGYCON)*, pp. 1–6, April 2016.

[155] J. Cuenca, F. Larrinaga, and E. Curry, "A Unified Semantic Ontology for Energy Management Applications," in *2nd International Workshop on Ontology Modularity, Contextuality, and Evolution (WOMoCoE 2017)*, 2017.

[156] J. García, F. J. García-Peñalvo, and R. Therón, "A survey on ontology metrics," in *Knowledge Management, Information Systems, E-Learning, and Sustainability Research* (M. D. Lytras, P. Ordonez De Pablos, A. Ziderman, A. Roulstone, H. Maurer, and J. B. Imber, eds.), (Berlin, Heidelberg), pp. 22–27, Springer Berlin Heidelberg, 2010.

[157] F. N. Kepler, C. Paz-Trillo, J. Riani, M. M. Ribeiro, K. V. Delgado, L. N. de Barros, and R. Wassermann, "Classifying ontologies," in *Proceedings of the Workshop on 2nd Workshop on Ontologies and their Applications co-located with the International Joint Conference IBERAMIA-SBIA-SBRN'06, October 23-27, 2006, Ribeirao Preto, SP, Brazil*, 2006.

[158] M. Horridge, M. E. n. Aranguren, J. Mortensen, M. Musen, and N. F. Noy, "Ontology design pattern language expressivity requirements," in *Proceedings of the 3rd International Conference on Ontology Patterns - Volume 929*, WOP'12, (Aachen, Germany, Germany), pp. 25–36, CEUR-WS.org, 2012.

[159] Protege, "Protege," 2018. Available at μ`https://protege.stanford.edu`.

[160] Y. Gu, Z. S. Qian, and F. Chen, "From twitter to detector: Real-time traffic incident detection using social media data," *Transportation research part C: emerging technologies*, vol. 67, pp. 321–342, 2016.

[161] A. M. Stefano *et al.*, "Ears (earthquake alert and report system) a real time decision support system for earthquake crisis management," in *Proceedings of the 20th ACM SIGKDD international conference on knowledge discovery and data mining*, pp. 1749–1758, 2014.

[162] J. B. Houston and J. e. Hawthorne, "Social media and disasters: a functional framework for social media use in disaster planning, response, and research," *Disasters*, vol. 39, no. 1, pp. 1–22, 2015.

[163] L. R. Bahl, P. F. Brown, P. V. de Souza, and R. L. Mercer, "A tree-based statistical language model for natural language speech recognition," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 37, no. 7, pp. 1001–1008, 1989.

[164] D. M. Magerman, "Statistical decision-tree models for parsing," in *Proceedings of the 33rd annual meeting on Association for Computational Linguistics*, pp. 276–283, Association for Computational Linguistics, 1995.

[165] P. F. Brown, P. V. Desouza, R. L. Mercer, V. J. D. Pietra, and J. C. Lai, "Class-based n-gram models of natural language," *Computational linguistics*, vol. 18, no. 4, pp. 467–479, 1992.

[166] W. B. Cavnar, J. M. Trenkle, *et al.*, "N-gram-based text categorization," in *Proceedings of SDAIR-94, 3rd annual symposium on document analysis and information retrieval*, vol. 161175, Citeseer, 1994.

[167] T. Mikolov, M. Karafiát, L. Burget, J. Černockỳ, and S. Khudanpur, "Recurrent neural network based language model," in *Eleventh annual conference of the international speech communication association*, 2010.

[168] Y. Bengio, R. Ducharme, P. Vincent, and C. Jauvin, "A neural probabilistic language model," *Journal of machine learning research*, vol. 3, no. Feb, pp. 1137–1155, 2003.

[169] Y. Kim, Y. Jernite, D. Sontag, and A. M. Rush, "Character-aware neural language models," in *Thirtieth AAAI Conference on Artificial Intelligence*, 2016.

[170] R. Collobert and J. Weston, "A unified architecture for natural language processing: Deep neural networks with multitask learning," in *Proceedings of the 25th international conference on Machine learning*, pp. 160–167, ACM, 2008.

[171] Y. Kim, "Convolutional neural networks for sentence classification," *arXiv preprint arXiv:1408.5882*, 2014.

[172] E. Loper and S. Bird, "Nltk: the natural language toolkit," *arXiv preprint cs/0205028*, 2002.

[173] C. Manning, M. Surdeanu, J. Bauer, J. Finkel, S. Bethard, and D. McClosky, "The stanford corenlp natural language processing toolkit," in *Proceedings of 52nd annual meeting of the association for computational linguistics: system demonstrations*, pp. 55–60, 2014.

[174] "spacy - industrial-strength natural language processing in python." https://spacy.io/.

[175] R. Studer, V. R. Benjamins, and D. Fensel, "Knowledge engineering: Principles and methods," *Data Knowl. Eng.*, vol. 25, pp. 161–197, 1998.

[176] W3C, "Owl 2 web ontology language primer (second edition)," 2012. Available at μ`https://www.w3.org/TR/owl2-primer/`.

[177] W3C, "Sparql 1.1 query language, year = 2013, url = https://www.w3.org/TR/sparql11-query/, urldate = 2019-05-23, note = Available at μ`https://www.w3.org/TR/sparql11-query/`."

[178] S. Lai, L. Xu, K. Liu, and J. Zhao, "Recurrent convolutional neural networks for text classification," in *Twenty-ninth AAAI conference on artificial intelligence*, 2015.

[179] M. Abadi, A. Agarwal, P. Barham, E. Brevdo, and Z. Chen, "TensorFlow: Large-scale machine learning on heterogeneous systems." https://www.tensorflow.org/, 2015. Software available from tensorflow.org.

[180] E. Hovy, M. Marcus, M. Palmer, L. Ramshaw, and R. Weischedel, "Ontonotes: The 90\% solution," in *Proceedings of the human language technology conference of the NAACL, Companion Volume: Short Papers*, 2006.

[181] J. Pennington, R. Socher, and C. Manning, "Glove: Global vectors for word representation," in *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)*, pp. 1532–1543, 2014.

[182] Common Crawl, "Common Crawl." http://commoncrawl.org/.

[183] N. Kalchbrenner, E. Grefenstette, and P. Blunsom, "A convolutional neural network for modelling sentences," *arXiv preprint arXiv:1404.2188*, 2014.

[184] B. Qian, J. Su, Z. Wen, *et al.*, "Orchestrating the development lifecycle of machine learning-based iot applications: A taxonomy and survey," *ACM Computing Surveys (CSUR)*, vol. 53, no. 4, pp. 1–47, 2020.

[185] J. Kreps, N. Narkhede, J. Rao, *et al.*, "Kafka: A distributed messaging system for log processing," in *Proceedings of the NetDB*, vol. 11, pp. 1–7, 2011.

[186] F. E. Taylor, B. D. Malamud, K. Freeborough, and D. Demeritt, "Enriching great britain's national landslide database by searching newspaper archives," *Geomorphology*, vol. 249, pp. 52–68, 2015.