

TOWARD DATA LAKES FOR CRISIS MANAGEMENT

A. Krtalić^{1*}, A. Kuveždić Divjak¹, A. Miletić¹

¹ Faculty of Geodesy, University of Zagreb, 10000 Zagreb, Croatia -
(andrija.krtalic@geof.unizg.hr, ana.kuvezdic.divjak@geof.unizg.hr, andrea.miletic@geof.unizg.hr)

KEY WORDS: Data Lakes, Crisis Management, Humanitarian Demining, Heat Islands, Volcanic Activity

ABSTRACT:

The content of the data lake comes (is filled) from different sources, and different users (experts in various fields) of the same data can download and analyse the same data for their (different) needs and analysis. Big Data about the human environment and the effect of natural and human-caused disasters (in this case: heat islands, earthquakes and lava flows, and landmine contamination) on that environment have been available to many people for years and are the subject of discussions, but there are still numerous research challenges in the form of structuring and storing data and analysis results. This implies certain requirements for efficient integration, access and querying of the various data in the data repository for the described purpose. Data lakes and data warehouses are offered as solutions to this problem. Well-designed data lakes can be a basic building block for different solutions in the analysis of the effects of disasters on the environment, and high-quality data warehouses for modelling future potential disasters in the same area. This paper presents certain personal observations and certain proposals for the creation of efficient data lakes and data warehouses (based on many years of work on problems in areas: humanitarian demining, heat islands and volcanic activity) for the needs of decision-making in crisis based on examples from practice. The goal is to influence the development of a unique framework for the creation and maintenance of a data lake, in terms of its better utilization so that it does not become a data swamp.

1. INTRODUCTION

Extreme events related to climate (weather) changes and social unrest in the world, such as floods, heat islands, forest fires and military conflicts, are becoming more frequent and intense all over the globe (Climate Change, 2014; Ikeda and Nagasaka, 2011; Angra, and Sapountzaki, 2022; Paolini and Santamouris, 2023; United Nations, 2015; Landmine Monitor, 2022). Such events greatly affect the population that is affected by such phenomena and the quality of their life. Open data (i.e. data that's freely used, shared and built-on by anyone, anywhere, for any purpose), plays a crucial role in addressing these challenges, as it provides access to valuable information and insights for decision-making, research, and collaboration.

Extreme weather events and natural disasters, including those mentioned above, are global risks with significant effects on people (UNISDR, 2016). Between 1998 and 2017 climate-related and geophysical disasters killed 1.3 million people and left a further 4.4 billion injured, homeless, displaced or in need of emergency assistance (UNISDR, 2018). If some add to that the fact that as of October 2022, a total of 67 countries and other areas were known or suspected to be contaminated with anti-personnel mines (Landmine Monitor, 2022), the need for high-quality data and adequate storage is evident. To address these risks, it is necessary to collect, analyze, and model data about environmental phenomena, enabling damage control and scenario planning for future events. This requires continuous improvement in data collection, storage, and processing capabilities, as well as the integration of diverse datasets from various sources, including open data.

The concept of a data lake, a centralized repository for storing large volumes of raw and unprocessed data, is particularly relevant in this context. Data lakes facilitate the storage and

management of diverse data types, including open data, obtained from measurements, sensors, social media, news, government, and private sources (Climate Change, 2014). By integrating different types of data, including structured, semi-structured, and unstructured data, within a data lake, organizations can gain comprehensive insights into the interconnectedness of environmental phenomena and their impacts on society.

However, the diversity and scale of data in data lakes pose significant challenges and open issues, such as data integration, data quality, and data interpretation (Abadi et al., 2016). Open data, with its principles of accessibility and collaboration, can contribute to addressing these challenges by providing high-quality data from various domains. By leveraging open data within data lakes, researchers and practitioners can access and interpret the data needed for understanding the origin and development of events or phenomena, supporting the development of decision-making support systems.

In this paper, we aim to address the needs for the creation and organization of a data lake system through examples from fields such as humanitarian demining, heat islands, volcanic eruptions (and earthquakes), floods, and forest fires. By examining the learner's perspective, we emphasize the importance of data lakes in retrieving and integrating data on specific types of disasters. Furthermore, we identify significant challenges and problems implied by the implementation of data lakes, including the integration of open data, and offer insights into potential solutions.

2. MATERIALS AND METHODS

The increased frequency of disasters caused by climate change and human activity on the earth's surface, such as floods, forest

* Corresponding author

fires, occurrence of urban heat islands, earthquakes, volcanic activity and military mining, are a great challenge for the known models of decision-making systems (disaster risk reduction management). Due to the different causes and consequences of disasters, the formal adoption of an integrated disaster risk reduction and human development system remains fragmented due to a lack of legislative and policy frameworks (Raikes et al., 2019). Furthermore, different disasters are approached differently, but, on the other hand, the result of one can affect the other. For example, floods and fires can greatly affect the situation in a mine-suspected area. The approach to a particular phenomenon is also different. In the management models of floods, earthquakes and humanitarian demining, everything is (more) risk-oriented, while the management of forest fires is oriented towards and still dominates the modelling of current crisis management. However, in addition to the specific needs and data that should be known or collected to monitor and analyze a single disaster, there are also the same data that are used to monitor several of them. For example, the digital terrain model, which is necessary for the management of all the mentioned disasters.

The Sendai Framework for Disaster Risk Reduction involves action according to four priorities, namely: understanding disaster risk; strengthening crisis risk management; investing sufficient funds in disaster risk reduction; improve the readiness to react to threats and consequences (UNISDR, 2015). The integration and structuring of several different data sets that will be used to monitor and respond to several types of disasters can help the most in this. Instead of fragmented silos of data, it is more advisable to form one set of easy data, from which different experts will draw data for their needs. The integration and structuring of several different data sets that will be used to monitor and respond to several types of disasters can help the most in this. Instead of fragmented silos of data, it is more advisable to create a single set of data from which different experts will draw data for their needs.

Responses to threats and mitigation of the consequences of disasters can be structural (e.g. compensation for damages and victims, evacuation shelters, establishment of) or non-structural (e.g. construction and other laws, laws according to which damages are repaired, public awareness programs) and can be reactive (i.e. analysis of the impact of phenomena on the environment) or proactive (i.e. modelling of similar situations for the purpose of prevention and reduction of the risk of a similar occurrence again) (according to Raikes et al., 2019).

Understanding the origin and development of a certain event or phenomenon is important for determining the variety of responses and reactions to them, and the implications they have for the development of decision-making support systems in these situations (Wahl, 2018). A data lake (Dixon, 2010; Managing, 2015) is a popular and practical concept for collecting and storing data in various fields of human activity. This mainly concerns the academic community, data lakes should provide solutions for managing data in several different areas (e.g. floods, earthquakes, heat islands, humanitarian relief) at once (Wieder and Nolte, 2022).

Like data warehouses (Devlin and Murphy, 1988; Chaudhary et al., 2011), data lakes aim to integrate heterogeneous data from different sources into a single, homogeneous data management system. In this way, data users can overcome the limitations of different and isolated data silos and implement unified data management.

2.1 Data Lakes and Crisis Management

Data lakes can play a valuable role in crisis management by providing a flexible and scalable data storage solution that can accommodate the diverse and dynamic nature of crisis-related data. Here are a few ways data lakes can support crisis management:

- **Data Integration:** During a crisis, data from various sources such as social media feeds, sensor data, emergency calls, news reports, and public health data need to be collected and integrated quickly. A data lake allows for the ingestion and storage of these different types of data in their raw format, enabling organizations to centralize and consolidate information from multiple sources.
- **Real-Time Data Processing:** Data lakes can handle high-velocity data streams and perform real-time processing and analytics on the incoming data. This capability is beneficial in crisis management scenarios where rapid decision-making and situational awareness are crucial. Real-time data processing can help identify patterns, trends, and anomalies, enabling organizations to respond more effectively to the crisis.
- **Advanced Analytics:** Data lakes provide a platform for performing advanced analytics on crisis-related data. By applying machine learning, natural language processing, and other analytical techniques, organizations can gain deeper insights into the crisis, identify potential risks, predict future developments, and support evidence-based decision-making.
- **Collaborative Data Sharing:** Data lakes can facilitate data sharing and collaboration among different organizations involved in crisis management, such as government agencies, emergency services, healthcare providers, and humanitarian organizations. By establishing data access controls and protocols, stakeholders can securely share relevant data, enhancing coordination and information exchange.
- **Historical Data Analysis:** Data lakes store data in its raw and unprocessed form, allowing historical data to be retained for future analysis. This can support post-crisis evaluations, performance assessments, and the development of strategies for improving crisis response and preparedness.

Implementing a data lake for crisis management requires careful planning, data governance, and security measures to ensure the accuracy, reliability, and privacy of the data being stored and processed. Organizations should establish appropriate data management practices, data quality controls, and security protocols to maximize the benefits of using a data lake in crisis management scenarios.

2.2 Data for Mine Action

67 countries in the world are facing the problem of mines left over after war conflicts at their territories (Landmine Monitor, 2022). After the end of the conflict, each national demining center is tasked with collecting, reviewing, and analyzing all available data on suspected hazardous areas. "A Suspected Hazardous Area (SHA) is an area where there is reasonable suspicion of mine/Explosive Remnants of War (ERW) contamination on the basis of indirect evidence of the presence of mines/ERW" (IMAS 04.10, 2003). After that, these areas

should be defined and marked in order to reduce the risk to the population. In many cases, warring parties are not inclined to provide complete information about mined areas, so there is a need to collect additional information from the depth of the SHA. On the basis of these and existing data, the entire area is then analyzed and decisions are made about the final borders of the SHA and how to approach mine removal. The impact of landmines and ERW goes beyond individual accidents and casualties, especially in tourist countries such as Croatia (tourism accounts for over 20% of GDP). Many former battlefield areas have created acute social, economic and environmental damage by denying access to productive areas for civilian use.

Although there are a large number of land (underground) mine detectors today, demining per se has not progressed much since World War II, except for machine demining, which cannot be carried out in all terrains. In addition, there was the development of a system for collecting and analyzing additional information from the depth of the SHA for the purpose of its reduction (CROMAC, 2010; Krtalić and Bajić, 2019) and its better definition. On the operation of such a system, the needs for collecting data for the purpose of activities in humanitarian demining will be presented. The Advanced Intelligence Decision Support System (AIDSS) technology is the first mine action technology in humanitarian demining (IMAS 04.10, 2003) to combine methods of remote sensing with advanced intelligence in a successful operational system in nontechnical surveys (Bajić, 2010; CROMAC, 2010; Krtalić and Bajić, 2019). The AI DSS is not a detector of mines, but a set of tools and methods to be used by experienced operators and experts from Mine Action Centres in order to help area reduction from remote sensing data and expert knowledge. The constituent parts of AI DSS technology are shown in Figure 1. The main role in the whole process is played by national Mine Action Centers (MACs). They set requirements for (additional) data collection on the basis of which they will perform analyzes and obtain results.

The mine action experts conduct MIS data analysis and derive general and specific requirements about missing information and data. The data available in MIS can be: positions of known fortification facilities, minefield records (a document on the method of laying mines, and the number of specified mines in a minefield), mine accidents. MIS data was collected from interviews with returnees to the war zone, from military maps, from monographs of military commanders, and from field reconnaissance (MAC scouts). The reconstruction of the mine scene is carried out, by MAC experts, on the basis of the known and mentioned information about it, introducing into consideration the visual basis such as digital orthophoto (DOP), geographical maps/aerial and satellite images, digital surface model (DSM). From them, information is obtained about the configuration of the terrain (which is an important piece of information in combat tactics) and connected with contextual data (the method of demining also depends on the configuration of the terrain). Aerial data collection missions are carried out after the analysis of all existing data and established requirements for them. The processing and interpretation of aerial and satellite images, for the purpose of finding the remains of military fortifications (mining indicators (Krtalić and Bajić, 2019)) is done by mine scene interpreters (persons who know the way of warfare in a certain area and the methods of digital image processing). The next stage of the AI DSS technology process is the fusion of all data, based on multi-criteria analysis (5 stage in Figure 1). The results of this fusion are the thematic maps of the SHA in the form of a mine hazard

map (Krtalić et al., 2018). The final stage of the application of AI DSS technology is the return to the beginning, that is, the verification of the obtained results (claims) by experts in MAC. Based on this check, the data enters or does not enter the MIS (Bajić et al. 2011).

The AI DSS technology is based on existing MIS data of national MACs, and on new evidence from aerial and satellite images (Figure 1). That complex system consisting of three modules (Krtalić and Bajić, 2019):

- Module for analytical evaluation (MIS) of data
- Data collection module
- Data preprocessing and processing module,

within which the described phases of AI DSS take place.

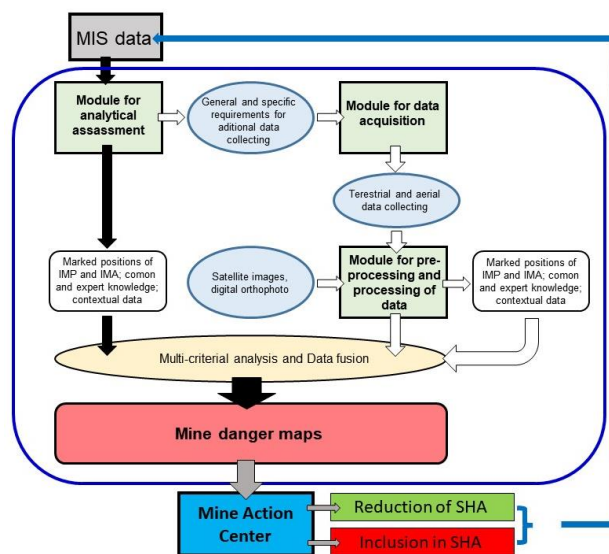


Figure 1. Scheme of AI DSS methodology (Krtalić and Bajić, 2019).

2.3 Data for the Surface Urban Heat Islands

The surface urban heat island (SUHI) is a phenomenon that represents the difference in land surface temperature (LST) between urban and non-urban surfaces (Zhou et al. 2019) and appears in a different way than the urban heat island (UHI) (Roth et al., 1989; Tran et al., 2006). UHI is caused by differences in radiative cooling between urban and rural areas during the night, while SUHI is caused by differences in radiative surface heating between urban and rural areas during the day (Choi et al., 2014). UHIs can be classified into two major categories (Zhou et al. 2019) based on how they are formed and according to their height. These categories are: 'air' (or 'atmospheric') and 'surface' UHI (Oke, 1982). The UHI is usually detected by measuring air temperature data in situ, from ground meteorological stations (Nichol et al., 2009). Still, the in situ measured air temperatures usually can't provide enough spatial details for detection of UHI. On the contrary, the SUHI is primarily detected and measured by satellite thermal remote sensing data. It offers the ability to study the urban thermal environment at various spatial and temporal scales (Pal et al., 2012, Zhou et al. 2019). Areas with high LST mainly contain tall buildings (residential buildings or industrial complexes) with low vegetation cover (Li et al. 2013). Vegetation cover helps in selective absorption and reflection of incident radiation and thus regulates heat exchange (Gallo et al.

1995). Vegetated urban areas reduce the possibility of SUHI formation, and the removal of existing vegetated areas contributes to the formation of new SUHIs (Foley et al., 2005). Research has established that LST is strongly dependent on the health of the vegetation (Yue et al., 2007), the healthier vegetation, the LST values are lower, and vice versa. LST and NDVI have a direct and inversely proportional relationship with the level of environmental criticality of the analyzed area (Senanayake et al. 2013a). Based on this fact, the deductive environmental criticality index (ECI) was defined (Figure 2):

$$ECI = \frac{LST}{NDVI} \quad (1).$$

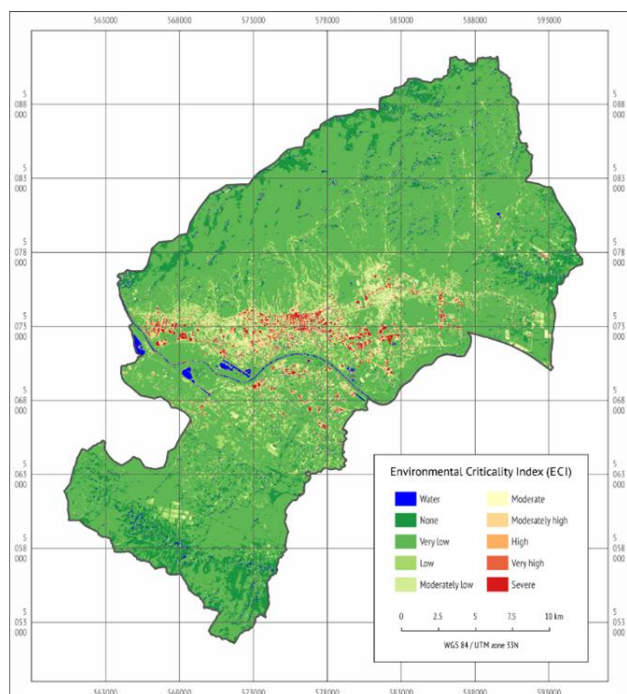


Figure 2. Environmental criticality index (ECI) in the city of Zagreb during 2017 (Krtalić et al., 2020).

In this way, the level of combined environmental criticality can be identified based on LST and vegetation area. The ECI value indicates the impact of SUHI on urban areas, that is, where they have the greatest impact in terms of land cover.

2.4 Data for the the Monitoring of Volcanic Activity

After a series of large volcanic eruptions around the world (Kilauea in 2021 and Mauna Loa in Hawaii in 2022, Shiveluch in Russia in 2023, of the Hunga-Tonga-Hunga-Ha’apai volcano in Tonga in 2022, La Palma in Spain in 2021, Etna in Italy, in 2022) and their wide consequences on the local population (fatalities, tsunamis, concentration of SO₂, disappearance of residential and agricultural areas), the need for information in order to establish a quality management system in emergency situations it is greater than ever. The use of remote sensing methods, which are based on the collection of data from the air and the processing of such collected data, enables access to reliable data in a fast and economical way. Sensors installed in satellite systems enabled the collection of information about physical, chemical and biological systems on the Earth’s surface, as well as the monitoring of changes and assessment of

phenomena in real time and subsequently. In this way, the problem of lack of adequate data for crisis management was partially eliminated. For a complete data solution, in situ data on the conditions of the volcanic area before and after the eruption are needed.

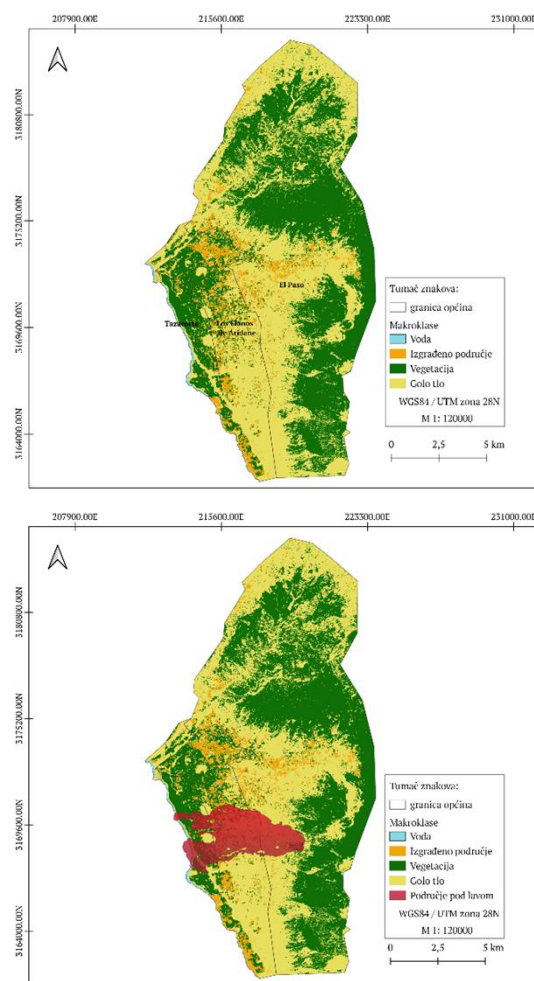


Figure 3. Classification of the surface of La Palma Island before (upper image) after (lower image) the volcano eruption in 2021 (Ilić, 2022).

Regarding the frequency of eruptions, volcanoes are classified into three categories: active, dormant and extinguished (Rothery, 2016). An active volcano is a volcano whose activity was visible in previous years. A dormant volcano is one that has not erupted recently but may do so soon. While a volcano that will never erupt again (after the last known eruption) is called extinct. Types of lava (magma that pours from a volcano onto the surface of the earth) differ according to the composition from which they are built and the speed of the flow expansion (Harris et al., 2017). Volcanoes can erupt with a small explosion, with great ferocity or simply with the flow of lava. This depends on the amount of magma and the concentration of gases under the volcano. In weaker eruptions, only magma is poured out; in the case of explosive eruptions with a lot of gas, magma is thrown, and volcanic ash is ejected high into the atmosphere.

Monitoring of volcanoes can be carried out using: a permanent network of GPS points (with the remark that such a network does not provide sufficient spatial coverage for better

monitoring of surface deformations caused by the eruption (González et al., 2015)); classification of multispectral images (determining the conditions before and after the eruption, Figure 3); integration of satellite multispectral images and a digital terrain model (on the basis of terrain slope data, the lava flow can be predicted, Tsang et al., 2020); thermal images (monitoring of thermal anomalies on them, Francis and Oppenheimer, 2003); radar images (analyses of radar images and radar interferometry, González et al., 2015); monitoring the spread of volcanic gas and ash clouds (using data from the Geostationary Operational Environmental Satellite (GOES)); measuring the concentration of sulphur dioxide in the atmosphere (the explosive power of the eruption can be determined from the concentration of volcanic gases in the atmosphere, Theys et al., 2019).

2.5 Structuring Data Lake for Crisis Management

This paper aims to help researchers design and create a data lake according to their needs (area of interest) and discover open questions and directions for future data lake research. For this purpose, the results of previous research on specific crisis situations were used, instead of scientific assumptions (Hai et al., 2015). As stated earlier, there are also the same data type that are used to monitor several kinds of disaster. Some examples are listed in Table 1. Following the approach presented in Table 1, tentative structure for a data lake in the context of crisis management can be proposed:

1. Input Data Sources:

- **National Databases:** Incorporate structured input data from various national databases such as cadastre, mining information systems, pedological data, and data on forests.
- **Public Databases and Providers:** Include data from public databases and providers, such as portals with satellite images and results of remote sensing methods (e.g., NASA, Copernicus).
- **Field Data:** Collect unstructured input data from the field, including real-time observations and reports on current crisis conditions.

2. Data Processing and Integration:

- **Data Preprocessing:** Clean, transform, and standardize the collected data to ensure consistency and compatibility.
- **Integration and Fusion:** Integrate heterogeneous data from different sources, including structured data from national databases, public databases, and providers, as well as unstructured field data.
- **Multispectral Image Processing:** Perform image classification and analysis to derive relevant information about conditions before and after disasters, such as volcanic eruptions.
- **Terrain Analysis:** Utilize digital terrain models and terrain slope data to predict lava flow patterns and assess the impact of volcanic activity.
- **Environmental Criticality Index (ECI):** Calculate the ECI using land surface temperature (LST) and vegetation data to identify areas with high environmental criticality, such as surface urban heat islands.

3. Storage and Data Management:

- **Structured Output Data:** Store structured output data generated from data processing and analysis, including mine hazard maps, thematic maps of disaster areas, and other relevant results.
- **Data Warehouses:** Establish data warehouses to store and organize the structured output data, allowing for efficient retrieval and utilization by national and local communities.
- **Metadata Management:** Maintain comprehensive metadata to document the characteristics, origin, and processing history of the data within the data lake.
- **Data Governance:** Implement data governance practices to ensure data quality, security, privacy, and compliance with relevant regulations.

4. Data Retrieval and Utilization:

- **User-specific Data Extraction:** Enable users to retrieve specific data subsets relevant to their respective

Humanitarian demining	Urban Heat Islands	Volcanic Activity	Floods	Wildfire
Administrative data (borders of cities, municipalities, cadastral units, etc.)	Administrative data (borders of cities, municipalities, cadastral units, etc.)	Administrative data (borders of cities, municipalities, cadastral units, etc.)	Administrative data (borders of cities, municipalities, cadastral units, etc.)	Administrative data (borders of cities, municipalities, cadastral units, etc.)
Information about the parties to the conflict (types of weapons, tactics, methods of laying mines)				
Military data (units in areas of military action, data from military maps, minefield records, etc.)				
Spatial data (aerial and satellite images, DOP; DTM, terrain slope, etc.)	Spatial data (aerial and satellite images, DOP; DTM, terrain slope, etc.)	Spatial data (aerial and satellite images, DOP; DTM, terrain slope, etc.)	Spatial data (aerial and satellite images, DOP; DTM, terrain slope, etc.)	Spatial data (aerial and satellite images, DOP; DTM, terrain slope, etc.)
Temporal data on the situation in the mine suspected area (possible floods, fires, etc.)	Temporal data on the situation in the mine suspected area (possible floods, fires, etc.)	Temporal data of previous activities and consequences (repetition frequency and intensity, for the purpose of modeling them when they happen again)	Temporal data on previous floods (for the purpose of preventing new ones and modeling them if they happen again)	Temporal data on previous fires (for the purpose of preventing new ones and modeling them if they happen again)
Land cover data (soil type, types of vegetation, water, etc.)	Land cover data (soil type, types of vegetation, water, etc.)	Land cover data (soil type, types of vegetation, water, etc.)	Land cover data (soil type, types of vegetation, water, etc.)	Land cover data (soil type, types of vegetation, water, etc.)

Table 1. Types of data used to monitor and analyze different types of crises.

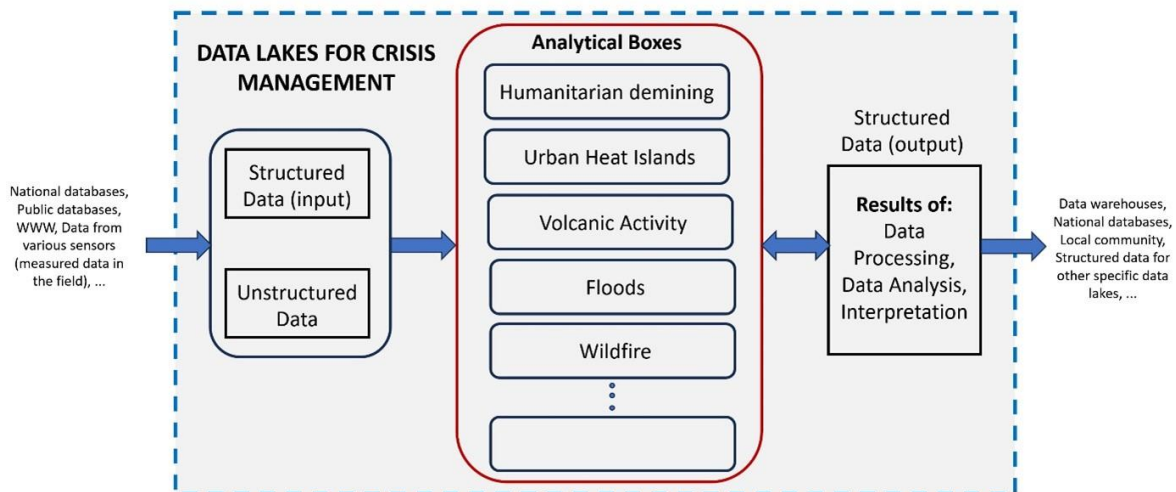


Figure 4. A possible structure of a potential data lake for crisis management.

respective domains, such as demining, urban planning, or disaster response.

- **Cross-Domain Analysis:** Facilitate the analysis of interrelated phenomena by allowing users to access and combine data from different disaster types, enabling comprehensive and integrated insights.
- **Decision Support Systems:** Support the development of decision support systems by providing access to timely and accurate data for proactive and reactive responses to crises.
- **Collaboration and Knowledge Sharing:** Promote collaboration among researchers, practitioners, and stakeholders by providing a shared platform for data sharing, analysis, and knowledge exchange.

In order to provide the best possible response to potential threats and crises, the data lake for crisis management (Figure 4) must be filled with structured input data from various national databases (cadastre, mining information system (if such exists), pedological data, data on forests, ...), public databases and providers data (portals with satellite images, results of remote sensing methods (e.g. NASA, Copernicus)). For monitoring the course of the crisis and modelling its development and further movement (for example, the spread of fire and forecasting the area that will be affected by flooding), it is necessary to collect data from the field. These data must also 'flow' into such a lake. This is very valuable unstructured input data about current conditions in crisis areas. According to the above, it is clear that such a lake contains different types of structured and unstructured input data. From the data body structured in this way, each user pulls data for a specific type of disaster or crisis situation. In this way, the results of data processing within a particular disaster can be easily used in the analysis of other crisis situations. For example, floods and torrents can cause landmines to be moved to areas outside of war zones where experts in humanitarian demining do not expect them. Furthermore, burned surfaces will also behave differently in floods, that is, water will move differently over fire-reduced vegetation. The results of the analyzes within the thus formed lake are stored as structured output data that can be stored in the form of data warehouses, they can be used by the national and local community for their needs in crisis situations.

3. CONCLUSION

In conclusion, the increased frequency and intensity of disasters caused by climate change and human activities pose significant challenges for decision-making systems and crisis management. Addressing these challenges requires the integration and structuring of various data sets to monitor and respond to multiple types of disasters simultaneously. Data lakes, along with the principles of open data, provide a practical solution for managing and analyzing heterogeneous data from diverse sources.

The research presented in this paper highlights the importance of data lakes in specific domains, such as humanitarian demining, surface urban heat islands, and monitoring volcanic activity. In the context of demining, the Advanced Intelligence Decision Support System (AIDSS) technology combines remote sensing methods with advanced intelligence to support area reduction and mine hazard mapping. For surface urban heat islands, satellite thermal remote sensing data and environmental criticality index (ECI) enable the study of urban thermal environments and the identification of areas with high land surface temperature. Additionally, the monitoring of volcanic activity involves various data sources, including multispectral images, digital terrain models, thermal images, radar images, and the measurement of volcanic gases.

The proposed structure of the data lake for crisis management ensures the collection and analysis of all available data on a particular disaster, as well as the analysis of the impact of crises on each other. Such a centralized system of data and results of various analyzes would enable a comprehensive insight into a particular situation and the impact of a particular crisis on the environment. Furthermore, a data lake organized in this way has a strict division into structured data (which can flow continuously from various verified databases) and unstructured data that is refreshed with each input so as not to make a swamp out of a lake. For this purpose, analyzes are provided, which structure the data and transform them into useful output data that can be stored in data warehouses and further used for various purposes. The proposed organization of data lakes can also be networked with similar lakes that have an influence on each other and thus form a 'system of connected data vessels' for analyzes of wider areas in wider social interests.

To effectively manage crises and mitigate their consequences, the design and creation of a data lake tailored to specific needs is crucial. Such a data lake should incorporate structured input data from national databases, public databases, and data providers, as well as unstructured data collected from the field. By structuring and integrating these diverse data types within a data lake, different users can extract data specific to their respective domains, facilitating the analysis and modeling of various crisis situations. The results of these analyses can be stored as structured output data, which can be utilized by national and local communities for their crisis management needs.

Overall, the utilization of data lakes, coupled with open data principles, enables comprehensive and integrated data management for crisis situations. By adopting data lakes as a foundational infrastructure, researchers and practitioners can leverage the power of diverse data sources to enhance decision-making, understand the origin and development of events, and develop effective strategies for disaster risk reduction and human development. Further research and exploration in this field will contribute to refining data lake architectures, addressing open issues, and advancing the field of crisis management through data-driven approaches.

ACKNOWLEDGEMENTS

This research is part of the Twinning Open Data Operational project that has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement no. 857592. Work by Andrea Miletić was additionally supported by "Young Researchers' Career Development Project—Training of Doctoral Students" (DOK-01-2020) founded by Croatian Science Foundation.

REFERENCES

- Abadi, D. et al., 2016. The Beckman report on database research. *Commun. ACM*, 59(2):92-99.
- Angra, D., Sapountzaki, K., 2022. Climate Change Affecting Forest Fire and Flood Risk—Facts, Predictions, and Perceptions in Central and South Greece. *Sustainability* 2022, 14, 13395.
- Bajić, M., 2010. The advanced intelligence decision support system for the assessment of mine-suspected areas. *The Journal of ERW and Mine Action*, Issue, 14(3), 69–75.
- Bajić, M., Matic, Č., Krtalić, A., Čandar, Z., Vuletić, D., 2011. Research of the Mine Suspected Area, CROMAC-Centre for testing, development and training Ltd. (CROMAC-CTDT Ltd.), Zagreb, Sortina 1d, 10020 Zagreb, Croatia
- Burnhill, P., Mewissen, M., Rusbridge, A., 2014. Where data and journal content collide: what does it mean to 'publish your data', Presented at 'Dealing with Data Conference', 26 August 2014, University of Edinburgh Library. <https://www.era.lib.ed.ac.uk/handle/1842/9394> (14 July 2023)
- Chaudhary, S., Murala, D. P., Srivastav, V.K., 2011. A Critical Review of Data Warehouse, *Global Journal of Business Management and Information Technology*. Volume 1, Number 2 (2011), pp. 95-103.

Choi, Y.Y., Suh, M.S., Park, K.H., 2014. Assessment of Surface Urban Heat Islands over Three Megacities in East Asia Using Land Surface Temperature Data Retrieved from COMS, *Remote Sensing*, 6, 5852-5867.

Climate Change, 2014. Intergovernmental Panel on Climate Change (IPCC), *Climate Change 2014: Impacts, Adaptation, and Vulnerability. Part A: Global and Sectoral Aspects. Contribution of Working Group II to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*, (2014) (Retrieved from United Kingdom and New York, NY, USA).

CROMAC., 2010. Area reduction of MSA. Zagreb, Croatia: CROMAC, Retrieved August 5, 2018, https://www.hcr.hr/pdf/SOP%20REDUKCIJA%2012.10.2010.%20KONACAN_eng.pdf, (30 April 2023).

Čmrlec, K., 2019. Determining and Mapping of Thermal Islands in the City of Zagreb Using Satellite Data, (in Croatian), Diploma Thesis, Faculty of Geodesy University of Zagreb, Zagreb, Croatia.

Devlin, B. A., Murphy, P. T., 1988. An architecture for a business and information system. *IBM Syst. J.* 27, 60–80.

Dixon, J., 2010. Pentaho, Hadoop, and Data Lakes. <https://jamesdixon.wordpress.com/2010/10/14/pentaho-hadoop-and-data-lakes/> (April 22, 2023)

Foley, J.A., DeFries, R., Asner, G.P., Barford, C., Bonan, G., Carpenter, S.R., Chapin, F.S., Coe, M.T., Daily, G.C., Gibbs, H.K., Holloway, T., Howard, E.A., Kucharik, C.J., Monfreda, C., Patz, J.a., Prentice, C., Ramankutty, N., Snyder, P.K., 2005. Global consequences of land use, *Science*, 309 (5734), 570–574.

Francis, P., Oppenheimer, C., 2003. Applications of satellite remote sensing techniques to volcanology. *TERRA-1: Understanding the Terrestrial Environment: The Role of Earth Observations from Space*, 37.

Gallo, K.P., Tarpley J.D., McNab, A.L., Karl, T.R., 1995. Assessment of urban heat islands: a satellite perspective. *Atmos Res*, 37 (1-3), 37–43.

González, P. J., Bagnardi, M., Hooper, A. J., Larsen, Y., Marinkovic, P., Samsonov, S. V., Wright, T. J., 2015. The 2014–2015 eruption of Fogo volcano: Geodetic modeling of Sentinel-1 TOPS interferometry. *Geophysical research letters*, 42(21), 9239-9246.

Hai, R., Quix, C., Jarke, M., 2015. Data lake concept and systems: a survey. *arXiv preprint arXiv:2106.09592*

Ikeda, S., Nagasaka, T., 2011. An emergent framework of disaster risk governance towards innovating coping capability for reducing disaster risks in local communities, *Int. J. Disaster Risk Sci.* 2 (2) (2011) 1–9.

Ilić, V., 2022. Monitoring the movement of lava after the volcanic eruption in the Canary Islands in 2021, (in Croatian), Diploma Thesis, Faculty of Geodesy University of Zagreb, Zagreb, Croatia.

- IMAS 04.10., 2003. Glossary of mine action terms, definitions and abbreviations. 2nd ed. 01 January, Amendment 7, August 2014. New York, USA, UNMAS
<https://www.mineactionstandards.org/standards/international-mine-action-standards-imas/imas-in-english/>, (30 April 2023)
- Krtalić, A., Bajić, M. 2019. Development of the TIRAMISU Advanced Intelligence Decision Support System, *European Journal of Remote Sensing*, 52:1, 40-55.
- Krtalić, A., Kuveždić Divjak, A., Župan, R., 2018. Visualization of Land Mine Danger, Svilaja Region (Croatia), *Journal of Maps*.
- Krtalić, A., Kuveždić Divjak, A., Čmrlec, K., 2020. Satellite-driven assessment of surface urban heat islands in the city of Zagreb, Croatia, *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, Volume V-3-2020, 2020 XXIV ISPRS Congress (2020 edition).
- Landmine Monitor, 2022. Landmine Monitor Report, International Campaign to Ban Landmines, Landmine Monitor 2022 (ICBL-CMC: November 2022) http://www.the-monitor.org/media/3352351/2022_Landmine_Monitor_web.pdf (14 July 2023).
- Li, Z-L, Tang, B-H., Wu, H., Ren, H., Yan, G., Wan, Z., Trigo, I.F., Sobrino, J.A., 2013. Satellite-derived land surface temperature: Current status and perspectives, *Remote Sensing of Environment*, 131, 14–37.
- Managing, F.H., 2015. Data lakes in big data era: What's a data lake and why has it become popular in data management ecosystem. In *Proceedings of the 2015 IEEE International Conference on Cyber Technology in Automation, Control, and Intelligent Systems (CYBER)*, Shenyang, China, 8–12 June 2015; pp. 820–824.
- Nichol, J.E., Fung, W.Y., Lam, K.-S., Wong, M.S., 2009. Urban heat island diagnosis using ASTER satellite images and „in situ“ air temperature, *Atmospheric Research*, 94, 276–284.
- Oke, T.R., 1982. The energetic basis of the urban heat island, *Quarterly Journal of the Royal Meteorological Society*, 108 (455), 1 – 24.
- Pal, S., Xueref-Remy, I., Ammoura, L., Chazette, P., Gibert, F., Royer, P., Dieudonne, E., Dupont, J.C., Haeffelin, M., Lac, C., Lopez, M., Morille, Y., Ravetta, F., 2012. Spatio-temporal variability of the atmospheric boundary layer depth over the Paris agglomeration: An assessment of the impact of the urban heat island intensity, *Atmospheric Environment*, 63, 261–275.
- Paolini, R., Santamouris, M., 2023. *Urban Climate Change and Heat Islands, Characterization, Impacts, and Mitigation*, Elsevier Inc., Radarweg 29, PO Box 211, 1000 AE Amsterdam, Netherlands.
- Raikes, J., Smith, T. F., Jacobson, C., & Baldwin, C., 2019. Pre-disaster planning and preparedness for floods and droughts: A systematic review. *International Journal of Disaster Risk Reduction*, 38, 101207.
- Roth, M., Oke, T.R., Emery, W.J., 1989. Satellite derived urban heat islands from three coastal cities and the utilization of such data in urban climatology, *International Journal of Remote Sensing*, 10, 1699–1720.
- Rothery, D., 2016. *Volcanoes, Earthquakes and Tsunamis: A Complete Introduction: Teach Yourself*, Hodder Education, part of Hachette UK, 338 Euston Road, London, UK.
- Senanayake, I.P., Welivitiya, W.D.D.P., Nadeeka, P.M., 2013a. Remote sensing based analysis of urban heat islands with vegetation cover in Colombo city, Sri Lanka using Landsat-7 ETM+ data, *Urban Climate*, 5, 19-35.
- Theys, N., Hedelt, P., De Smedt, I., Lerot, C., Yu, H., Vlietinck, J., Van Roozendael, M., 2019. Global monitoring of volcanic SO₂ degassing with unprecedented resolution from TROPOMI onboard Sentinel-5 Precursor. *Scientific reports*, 9(1), 1-10.
- Tran, H., Uchihama, D., Ochi, S., Yasuoka, Y., 2006. Assessment with satellite data of the urban heat island effects in Asian mega cities. *International Journal of Applied Earth Observation and Geoinformation*, 8, 34–48.
- Tsang, S. W., Lindsay, J. M., Coco, G., Deligne, N. I., 2020. The influence of surficial features in lava flow modelling. *Journal of Applied Volcanology*, 9(1), 1-12.
- United Nations, 2015: United Nations Office of Disaster Risk Reduction (UNISDR), *Global Assessment Report on Disaster Risk Reduction 2015* (Retrieved from Geneva, Switzerland).
- United Nations Office of Disaster Risk Reduction (UNISDR), 2015: Sendai framework for disaster risk reduction 2015-2030, Paper Presented at the Third UN World Conference, Sendai, Japan.
- UNISDR, 2016. Centre for Research on the Epidemiology of Disasters (CRED), United Nations Office of Disaster Risk Reduction (UNISDR), *The Human Cost of Weather-Related Disasters 1995-2015*, CRED & UNISDR, Geneva, Switzerland.
- UNISDR, 2018. *Economic Losses, Poverty & Disasters 1998-2017*, United Nations Office of Disaster Risk Reduction (UNISDR), CRED & UNISDR, Geneva, Switzerland.
- Zhou, D., Xiao, J., Bonafoni, S., Berger, C., Deilami, K., Zhou, Y., Froelking, S., Yao, R., Qiao, Z., Sobrino, J.A., 2019. *Satellite Remote Sensing of Surface Urban Heat Islands: Progress, Challenges, and Perspectives*. *Remote Sensing*, 11 (48), 36.
- Wahl, D., 2018. Transforming development and disaster risk, *Sustainability* 10 (1458), 2018 12.
- Wieder, P., Nolte, H., 2022. Toward data lakes as central building blocks for data management and analysis. *Front. Big Data* 5:945720.