Humanitarian Demining Using Data Observatories and Data Lakes

Conference Paper · May 2023

Marko Horvat
University of Zagreb Faculty of Electrical Engineering and Computing

Krtalić Andrija
University of Zagreb

Amila Akagic
University of Sarajevo

Klara Krmpotić
University of Zagreb

Humanitarian Demining Using Data Observatories and Data Lakes

Humanitarno razminiranje korištenjem opservatorija podataka i podatkovnih jezera

Marko Horvat[1], Andrija Krtalić[2], Amila Akagić[3], Klara Krmpotić[1], Sven Skender[1]

[1] Department of Applied Computing, Faculty of Electrical Engineering and Computing, University of Zagreb, Unska 3, HR-10000 Zagreb, Croatia

[2] Institute of Cartography and Photogrammetry, Faculty of Geodesy, University of Zagreb, Kačićeva 26, HR-10000 Zagreb, Croatia

[3] Faculty of Electrical Engineering, University of Sarajevo, Zmaja od Bosne 8, BH-71000 Sarajevo, Bosnia and Herzegovina

Abstract: Humanitarian demining is an indispensable part of any post-conflict reconstruction and rehabilitation efforts, as landmines and other unexploded ordnance (UXO) continue to pose a significant threat to civilians and infrastructure. Traditional non-technical survey demining methods are typically time-consuming, labor-intensive, and require professional familiarity with a number of different complex tools and involved procedures. The emergence of the latest artificial intelligence (AI) technologies for deep learning (DL) and data analysis techniques presents an opportunity to transform humanitarian demining by significantly improving its efficiency, speed, accuracy, and safety. One such approach is the development of a data observatory that implements data ingestion pipelines into a data lake using data transformation to improve the usability of unstructured data about a mine scene. This paper addresses the key characteristics and benefits of such a system and its potential to improve humanitarian demining.

Sažetak: Humanitarno razminiranje neizostavan je dio obnove bilo kojeg područja nakon završetka ratnih sukoba, s obzirom da mine i druga neeksplodirana ubojna sredstva (NUS) predstavljaju značajnu prijetnju civilima i infrastrukturi. Tradicionalne metode netehičnog izvida u razminiranju obično zahtijevaju veliku količinu vremena, rada, te iziskuju profesionalno poznavanje niza složenih alata i uključenih postupaka. Pojava novih tehnologija umjetne inteligencije (AI) za duboko učenje (DL) i tehnika analize podataka predstavlja priliku za transformaciju humanitarnog razminiranja znatnim poboljšanjem učinkovitosti, brzine, točnosti i sigurnosti. Jedan takav pristup je razvoj opservatorija podataka koji implementira cjevovode za unos podataka u podatkovno jezero korištenjem transformacije podataka kako bi se poboljšala upotrebljivost nestrukturiranih podataka o minskoj sceni. Ovaj rad bavi se ključnim svojstvima i prednostima takvog sustava kao i njegovim sposobnostima za unapređenje humanitarnog razminiranja.

Keywords: data transformation, data observatory, big data, non-technical survey, mine scene analysis, artificial intelligence

1. Introduction

Humanitarian demining is an indispensable part of post-conflict reconstruction and rehabilitation efforts, as millions of landmines and unexploded ordnance (UXO) continue to pose a significant threat to civilians, infrastructure, and socio-economic development [1]. Non-technical survey is an interdisciplinary process that allows to identify, understand, and prioritize the areas where UXO pose the greatest threat and to create mine presence probability maps, thus enabling a more organized, safe, targeted and efficient approach to mine action [2]. It typically involves the collection of relevant data from a variety of sources, such as remote sensing, mine action expert knowledge, formal and informal historical records, and interviews with local residents, former combatants, and eyewitnesses [3]. As new information becomes available, or as the situation on the ground evolves, the survey findings may need to be updated and refined. This adaptive approach ensures that mine action efforts remain responsive to changing circumstances and that resources are correctly redirected. It can be said that non-technical survey provides a comprehensive understanding of the complex and multifaceted problem of landmine and explosive remnants of war (ERW) contamination and is an indispensable part of any mine action process as a crucial foundation for the subsequent phases of clearance.

However, traditional approach to non-technical survey is time-consuming, labor-intensive, and requires experienced mine action domain professionals. The emergence of new artificial intelligence (AI) technologies and data analysis techniques presents an opportunity to revolutionize the overall humanitarian demining process by significantly improving its efficiency, accuracy, and speed [4]. This paper explores the potential of using a data observatory and data lake-based approach to non-technical survey for humanitarian demining, discussing the benefits and implications of such a system.

The reminder of the paper is organized as follows: The next section explains the proposal of data observatories and data lakes for the purpose of non-technical survey and humanitarian demining, including the description of the modules of such a concept. Section 3 details the data pipeline implemented as part of the proposed data observatory, including data input, management, analysis, and data sharing. Section 4 highlights the advantages of using blockchain technology in data lakes for humanitarian demining, such as immutability, transparency, and trustworthiness. Section 5 lists all benefits of developing a data observatory for humanitarian demining. Finally, the last section summarizes the proposed approach and emphasizes the potential impact of the proposed solution on improving the efficiency, speed, accessibility, and overall cost of humanitarian demining.

2. Data Observatories and Data Lakes: A New Approach to Humanitarian Demining

A data observatory is an organized, centralized platform designed to collect, store, process and analyze vast amounts of data from multiple sources [5]. The underlying premise of any data observatory is to provide stakeholders with access to comprehensive, up-to-date, and actionable information that can be used to enhance decision-making processes, help in data mining and discovery of new hidden information in existing datasets, and improve overall efficiency in various domains, including humanitarian demining [6]. The data observatory ingests the collected data into a data lake. Data lakes are large-scale, scalable and centralized repositories that can accommodate vast amounts of structured and unstructured data in native formats [7]. They are particularly well-suited for storing and processing diverse data types, such as textual descriptions, document scans, satellite or drone multispectral and hyperspectral images, and geospatial information. By implementing a data ingestion pipeline into a data lake, humanitarian demining efforts can leverage this scalable, cost-effective storage solution to accommodate the ever-increasing volume and variety of data.

In the proposal outlined in this paper, within the context of non-technical survey for humanitarian demining, a data observatory would gather information from various sources, such as satellite imagery, drone footage, ground sensors, and historical records, to create a comprehensive picture of a mine-affected area. This structure not only enhances decision-making processes but also drives innovation in humanitarian demining. By combining the power of data lakes with advanced data transformation techniques, the data observatory approach significantly could improve the usability and value of unstructured data for demining operations, ultimately contributing to more efficient and effective clearance of landmines and unexploded ordnance.

Unstructured data about a Suspected Hazardous Area (SHA), such as different satellite, aerial drone images or field reports, can be challenging to process and analyze due to its inherent complexity and lack of standardization. Data transformation techniques, including natural language processing (NLP), computer vision, and machine learning (ML) algorithms, are essential for converting this raw data into a more structured and usable format [4].

In the context of non-technical survey and humanitarian demining, data transformation techniques can facilitate the extraction of valuable information from diverse data sources. For example, different computer vision and deep learning algorithms can be applied to satellite imagery to identify potential mine-contaminated areas, while computer vision object classification and NLP together can help process and analyze textual data from hand-drawn minefield records, historical records and expert reports [4]. These techniques can be employed to automatically identify and categorize essential information within the data, such as the types of landmines or UXO, their locations, and potential hazards [4].

The Mine Action Data Observatory (MADO) concept consists of four main components (as depicted in Figure 1):

1. A component for ingestion, processing, and transformation of unstructured, non-technical survey data about mine scene.
2. A knowledge base component with a formal vocabulary for describing mine scene semantics using ontologies and knowledge graphs.
3. A component with a service for automatic inference from the knowledge base, exploratory search of the knowledge base, and generation of new knowledge.
4. An image processing component for localization and classification of mine scene objects using DL models.

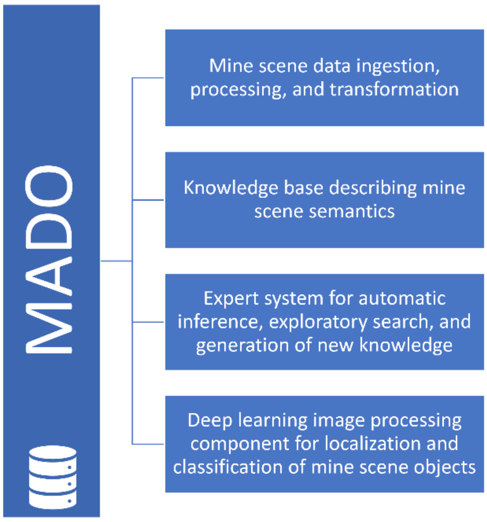

Figure 1. Four components of Mine Action Data Observatory (MADO).

The proposed data platform will leverage existing best practices in humanitarian demining and the knowledge base will include partially structured data on UXO obtained from any available sensor or through sensor fusion. Examples of such data include: Records of interviews with populations living near SHA, mine logs from completed mining or demining operations, mine accident logs, aerial and satellite imagery, scans of military maps, GIS mine explosive device data, and the results of analyzes of any other available data.

Currently, these data sources are not linked and there is no model that offers the possibility of integrated search, analysis, and discovery of new knowledge in the field of humanitarian demining.

3. Non-technical Survey Data Input, Management, Analysis, and Sharing Pipeline

A data observatory typically consists of several key components that work in concert to facilitate data input, management, analysis, and sharing. In [5], the components of a typical data observatory are identified as:

1) data ingestion,

2) data storage,

3) data transformation,

4) data analysis and visualization, and

5) data sharing and collaboration.

The data ingestion pipeline, as illustrated in Figure 2, is the first component of a mine action data observatory, and involves collecting raw data from various sources, such as aerial and ground-based sensors, satellite imagery, historical records, and expert knowledge.

A data ingestion pipeline is a series of processes that facilitate the collection, preprocessing, and storage of raw data from disparate sources. In the context of humanitarian demining, this may include data from aerial and ground-based sensors, satellite imagery, historical records, and expert knowledge. Data ingestion pipelines streamline the process of acquiring and integrating data, ensuring that it is readily available for analysis. Data ingestion pipelines are designed to streamline the acquisition and integration of data, ensuring that it is readily available for subsequent analysis.

Unstructured data, such as satellite images or field reports, can be challenging to process and analyze due to its inherent complexity and lack of standardization. Data transformation techniques, including natural language processing, computer vision, and machine learning algorithms, are essential for converting this raw data into a more structured and usable format. One such study was proposed in [9], where an approach for automatic detection of mine presence indicators in hazardous areas from aerial surveying images in a safe and explainable manner was presented.

As technology continues to advance and new data sources and formats emerge, the synergy between data ingestion pipelines and data lakes will only become more critical to the success of data observatories in humanitarian demining. The integration of emerging technologies, such as IoT devices, AI, and ML algorithms, will further enhance the capabilities of these components, enabling more sophisticated data collection, advanced data transformation techniques, real-time data processing, enhanced data integration and interoperability, data analysis, as well as security and data governance.

The ongoing development of AI and ML algorithms will facilitate more advanced data transformation techniques, allowing stakeholders to extract even greater insights from the diverse data sources collected by the data ingestion pipeline. This can help identify patterns, trends, and relationships that might otherwise be overlooked, contributing to more effective demining strategies.

In the context of real-time data processing, the integration of IoT devices and real-time data processing tools into the data ingestion pipeline will enable the continuous collection and analysis of data from mine-affected areas. This can provide stakeholders with up-to-date information about mine scenes, allowing them to respond more quickly and effectively to emerging risks and opportunities.

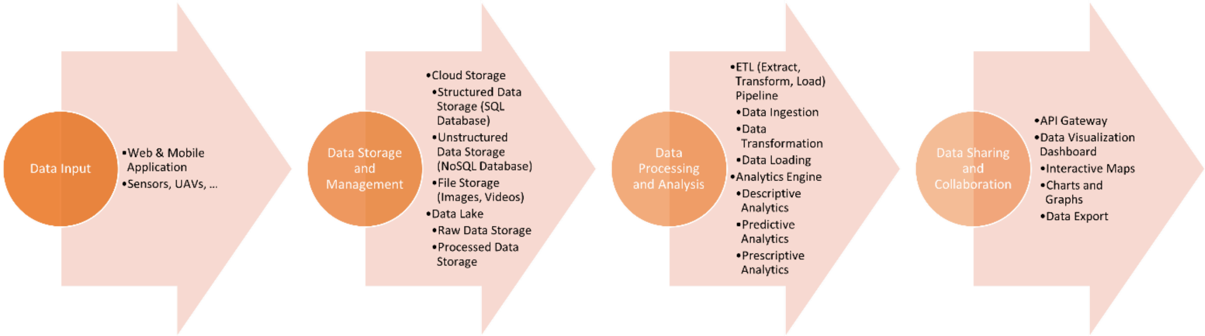

Figure 2. Data architecture pipeline of cloud-based software platform for data input, storage and management, processing and analysis, sharing and collaboration in non-technical survey.

As new data sources and formats become available, the integration capabilities of data ingestion pipelines and data lakes will need to evolve to accommodate these developments. This will ensure that data observatories remain a valuable resource for humanitarian demining efforts, promoting reporting, visualization, as well as collaboration, enhanced data integration and interoperability among different stakeholders, e.g., national Mine Action Centers (MACs) such as CROMAC, in the humanitarian demining operations.

In this regard, in our previous work we have developed a new core ontology called MINEONT (MINE-action ONTology) for formal knowledge representation and automated reasoning in demining [4]. This approach allows expressive representation of concepts in mine action, high-level semantics, geospatial metadata, and information obtained through various mine action surveys including multimodal data sources, such as aerial and satellite imagery, mine presence indicators, contextual data, and expert knowledge from humanitarian demining specialists [4]. The purpose of MINEONT is to be used to build a specialized AI-based decision support system that integrates large amounts of data and facilitates interpretation and decision-making processes as a data observatory for mine action [3][8].

4. Enhancing Mine Scene Description and Collaboration through Blockchain Technology

With the growing volume and variety of data associated with humanitarian demining, ensuring the security and proper data governance of data stored in data lakes will become increasingly important. Implementing robust data security measures and developing comprehensive data governance policies will be essential to safeguard sensitive information and maintain the trust of stakeholders involved in demining efforts.

To this regard we envision the use of blockchain technology to describe a mine scene. This approach has several significant benefits and could improve the overall efficiency and trustworthiness of humanitarian demining [10].

Blockchain technology ensures data integrity and reliability by securely storing geolocations, types, and other metadata of landmines and unexploded ordnance in an immutable and transparent ledger. By providing a shared, trustworthy source of information, the decentralized nature of blockchain facilitates collaboration among stakeholders such as governments, MACs, and international organizations. This allows for more efficient resource allocation, less duplication of effort, and better decision-making in demining operations. In addition, the verifiability provided by the blockchain enables stakeholders to trace the origin and history of mine scene data, which can aid in the establishment of trust in the information and the identification of potential inconsistencies in the landmine detection process.

5. Benefits of Developing a Data Observatory for Humanitarian Demining

In summary, the use of a data observatory and data lake system for non-technical survey in humanitarian demining, as proposed in the paper, offers several key benefits compared to the traditional approach:

1. Improved operational Efficiency: By automating the data collection, storage, and processing tasks, the mine action data observatory can significantly reduce the time and resources required to analyze and act upon mine-related data, leading to more efficient demining operations.
2. Enhanced decision-making: A data observatory provides mine action centers and other stakeholders with access to comprehensive, up-to-date information about mine scenes, enabling them to make more informed decisions regarding demining priorities and resource allocation.
3. Enhanced Accuracy: Advanced data analysis techniques and AI algorithms can help detect landmines and UXO with greater precision, thereby reducing the risk of false positives or negatives.
4. Increased Safety: Advanced data transformation techniques can help identify high-risk areas provide demining teams with real-time, accurate information about mine locations and potential hazards, allowing them to prioritize and execute their operations more safely.
5. Scalability and Flexibility: The data lake architecture is highly scalable and can accommodate the growing volume and variety of data sources, facilitating the integration of new technologies and methodologies in humanitarian demining as they become available.
6. Interoperability and collaboration: The centralized data observatory can serve as a central hub facilitating data sharing and joint analysis efforts as well as sharing best practices among various stakeholders, including governments, non-governmental organizations, and research institutions.

6. Conclusion

The integration of data observatories, data lakes, and advanced data analysis techniques has the potential to revolutionize humanitarian demining by offering a more efficient, accurate, and safer alternative to traditional methods. The design, development and deployment of Mine Action Data Observatory (MADO) can significantly contribute to the broader goals of post-conflict reconstruction and rehabilitation, ultimately saving lives and paving the way for sustainable development in mine-affected regions.

The development of a data observatory that implements data ingestion pipelines into a data lake and employs data transformation techniques to improve the usability of unstructured data about a mine scene is a promising approach for enhancing humanitarian demining efforts. By leveraging the power of data lakes, advanced data transformation techniques, and the centralization offered by a data observatory, stakeholders can unlock valuable insights, streamline operations, and ultimately save lives.

The integration of emerging technologies, such as artificial intelligence, machine learning, and computer vision, will only serve to further refine the capabilities of data observatories in the field of humanitarian demining. As these technologies advance, the potential for more accurate, efficient, and safer demining operations increases, paving the way for a future where the threat of landmines and unexploded ordnance is effectively mitigated.

References

[1] Humanitarian Demining, Geneva International Centre for, (2016). A Guide to Developing National Mine Action

Standards," Global CWD Repository, 1317. Available: 01-April-2023, https://commons.lib.jmu.edu/cisr-globalcwd/1317

[2] Bajić, M. (2010). The advanced intelligence decision support system for the assessment of mine-suspected areas. The Journal of Conventional Weapons Destruction, 14(3), 28.

[3] Krtalić, A., & Bajić, M. (2019). Development of the TIRAMISU advanced intelligence decision support system. European Journal of Remote Sensing, 52(1), 40-55.

[4] Horvat, M., Krtalić, A., Bajić, M., Muštra, M., Laura, D., & Gold, H. (2022). MINEONT: A proposal for a core ontology in the aerial non-technical survey domain. Book of papers of 18th International Symposium MINE ACTION 2022, 47-51.

[5] Imperial College London, Data Observatory, Available: 01-March-2023, https://www.imperial.ac.uk/data-science/facilities/data-observatory/

[6] Miloslavskaya, N., & Tolstoy, A. (2016). Big data, fast data and data lake concepts. Procedia Computer Science, 88, 300-305.

[7] Khine, P. P., & Wang, Z. S. (2018). Data lake: a new ideology in big data era. In ITM web of conferences (Vol. 17, p. 03025). EDP Sciences.

[8] Bajić, M., Gold, H., Horvat, M., Krtalić, A., Laura, D., & Muštra, M. (2021). The novel paradigm for a decision support system of the aerial non-technical survey. Book of papers of 17th International Symposium MINE ACTION 2021, 62-68.

[9] Osmankovic, D., Akagic, A., Krivic, S., Uzunovic, T. and Velagic, J. (2022). Towards Safe and Explainable Humanitarian Demining with Deep Learning. Book of papers of 18th International Symposium MINE ACTION 2022.

[10] Nassar, M., Salah, K., ur Rehman, M. H., & Svetinovic, D. (2020). Blockchain for explainable and trustworthy artificial intelligence. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 10(1), e1340