Article

Ontology-Based Data Observatory for Formal Knowledge Representation of UXO Using Advanced Semantic Web Technologies

Marko Horvat [1,*], **Andrija Krtalić** [2], **Amila Akagić** [3] **and Igor Mekterović** [1]

1   Department of Applied Computing, Faculty of Electrical Engineering and Computing, University of Zagreb, Unska 3, HR-10000 Zagreb, Croatia; igor.mekterovic@fer.hr
2   Institute of Cartography and Photogrammetry, Faculty of Geodesy, University of Zagreb, Kačićeva 26, HR-10000 Zagreb, Croatia; kandrija@geof.hr
3   Faculty of Electrical Engineering, University of Sarajevo, Zmaja od Bosne 8, BH-71000 Sarajevo, Bosnia and Herzegovina; aakagic@etf.unsa.ba
*   Correspondence: marko.horvat3@fer.hr

Abstract: As landmines and other unexploded ordnances (UXOs) present a great risk to civilians and infrastructure, humanitarian demining is an essential component of any post-conflict reconstruction. This paper introduces the Minefield Observatory, a novel web-based datastore service that semantically integrates diverse data in humanitarian demining to comprehensively and formally describe suspected minefields. Because of the high heterogeneity and isolation of the available minefield datasets, extracting relevant information to determine the optimal course of demining efforts is time-consuming, labor-intensive and requires highly specialized knowledge. Data consolidation and artificial intelligence techniques are used to convert unstructured data sources and store them in an ontology-based knowledge database that can be efficiently accessed through a Semantic Web application serving as the Minefield Observatory user interface. The MINEONT+ ontology was developed to integrate diverse mine scene information obtained through non-technical surveys and remote sensing, such as aerial and hyperspectral satellite imagery, indicators of mine presence and absence, contextual data, terrain analysis information, and battlefield reports. The Minefield Observatory uses the Microdata API to embed this dataset into dynamic HTML5 content, allowing seamless usage in a user-centric web tool. A use-case example was provided demonstrating the viability of the proposed approach.

Keywords: semantic web; advanced web applications; HTML5 microdata; ontology; data integration; humanitarian demining; suspected hazardous area

1. Introduction

Military and civilian approaches to demining (humanitarian mine action or mine action [1]) differ significantly. Civilian demining begins when the conflict ends, and all actions are aimed at minimizing the risk with the goal of a 100% clearance rate of land mines [2]. The development of successful planning and implementation of demining projects depends on the quality and adequate use of all available information on the location of the Suspected Hazardous Area (SHA) [3,4]. These plans are implemented by national centers for demining [5] (so-called Mine-Action Centers, MACs) under the auspices of governments, which do not deal with demining per se, but collect and analyze all the data they can find by interviewing returnees in SHA, general assessments of mine actions [1,5], non-technical surveys [6], technical surveys [7], analytical assessments of military maps, reading biographies of military commanders, and interpreting and analyzing multi-sensory images of SHA. In [2], it was stated that these are long and expensive processes that ultimately do not provide enough accurate information. That is why demining experts need to speed up the mine action process to quickly identify SHA to avoid accidents as

much as possible. For this purpose, this paper introduces a novel ontology-based approach to storing and using all data for humanitarian demining.

At a time when data are increasingly influencing automated decision-making processes in various sectors, the complex and important field of humanitarian mining remains relatively unaffected and still relies on traditional methods of data mining and knowledge discovery. However, the heterogeneity, diversity, and—above all—the sheer size of data obtained from various sensors pose a significant challenge and make the extraction of relevant, usable information for non-technical time and labor-intensive. The nature of the data necessary to conduct a successful demining is very diverse and includes not only data from multispectral, hyperspectral, ground-penetrating radar or magnetic sensors from remote sensing techniques but also information derived from different unstructured and semi-structured documents such as hand-drawn minefield charts or reports of previous minefield accidents, among others [8–13].

Therefore, the significance of semantic capture and formal and structured representation of SHA has become extremely important as it contributes significantly to improving operational efficiency and facilitating more effective decision-making processes in humanitarian demining. In this context, semantic technologies have proven to be an indispensable tool in various other domains (per examples [14–16]), and they are also well-aligned with current trends for upcoming industry standards [17]. They enable the construction of knowledge bases characterized by linking concepts according to rules governing the relationships between different concepts. The overall goal of these tasks is to project ontology-related textual information into distribution vectors so that a variety of applications—from semantic similarity measurement and supervised learning based on ontology annotations to knowledge discovery and reasoning—can be performed with greater efficiency.

For these reasons, two original and innovative concepts are presented in this paper within the areas of advanced web technologies and artificial intelligence to improve the non-technical survey and humanitarian demining. These two contributions are (1) the Minefield Observatory and (2) the MINEONT+ ontology.

The Minefield Observatory is a novel concept based on advanced web technologies and formal knowledge representation methods. This approach shows how we may use the potential of semantic web technologies to solve the difficulties of data integration in minefields. Our solution uses DevOps principles and cloud-based infrastructure to transform unstructured and semi-structured data into machine-readable formats to access the data through HTML5 web application endpoints. This results in creating a unified web-based data repository that significantly improves current methods and practices for managing and interpreting mine scene data.

The data observatory web platform, such as the Minefield Observatory, represents a substantial advancement in data management and representation. It provides a basis for organizing and understanding massive amounts of unstructured and semi-structured data from multiple sources. This is especially important in a non-technical survey for mine action, as data from many sources, such as reports, field notes, maps, hyperspectral airborne images, and various ground observation sensor readings, must be merged and interpreted. A key feature of data observatories is their capacity to gather and make sense of such complicated and varied information. Establishing a single perspective of the data makes it easier to identify patterns, trends, and insights, resulting in more informed decisions and actions.

At the center of the Minefield Observatory is the MINEONT+ ontology, which we developed to bring together diverse mine scene information. This ontology includes a number of aspects, such as concepts represented in multimodal aerial and hyper-spectral satellite imagery, indicators of the presence or absence of mines, contextual data, terrain analysis information, and battlefield experiential knowledge. Using the Microdata API, we can embed a minefield record dataset into an HTML5 page, employing the latest web technology standards and thereby facilitating its use in a user-facing tool.

The remainder of the paper is organized as follows: Section 2 sets the stage and introduces contemporary data platforms, explaining the concepts of data observatories,

data lakes and optimization of data lake performance using ontologies in the context of data engineering. In particular, Section 2.1. describes data platform functions with their different advantages and drawbacks. Section 2.2 explains how data lakes may be used in cloud-based data platform architectures to ingest data in raw format. Section 2.3 explains how the introduction of ontologies as a tool for formal knowledge representation can assist in data transformation to consolidate diversely structured datasets. Section 3 provides an overview of related work. Section 4 describes the MINEONT+ ontology developed to formally describe, share, and semantically enrich the high-level description of a SHA. The developed ontology is used for data consolidation of diversely structured minefield documents in a non-technical survey. This section also presents the structure of the Microdata API developed to share the content of the Minefield Observatory's knowledge base utilizing the MINEONT+ ontology. Section 5 presents the structure of the Minefield Observatory and the methodology for evaluating the observatory's knowledge base, as well as the results of this evaluation. To demonstrate the feasibility of the proposed approach, Section 6 provides a use-case with a real-life example of minefield records ingestion into the Minefield Observatory using deep learning algorithms and export of such that may be provided by using the Microdata API. Section 7 provides a discussion about the Minefield Observatory approach and deliberates on the benefits and limitations of this study. Finally, Section 8 concludes the paper and proposes possible future research directions.

2. Data Observatories as Advanced Web-Based Data Management Platforms

To understand data observatories, their function, and benefits over traditional and monolithic repositories of diverse datasets such as data warehouses, it is necessary to first define the concepts of data platforms [18], data lakes [19], and ontology-optimized data lakes [20].

A data platform is a type of software designed for three main functions: collecting data, analyzing data, and managing data. These platforms are usually made to work as analytics platforms and to act as an integration layer. When we look at their structure, data platforms can be mainly of two types: (1) centralized data platform and (2) decentralized data platform, often called the "Data Mesh" [21,22].

The centralized data model is characterized by a monolithic architecture where data ownership, responsibility, and management reside primarily with a single, centralized team [23]. The obvious benefit of such an approach is that organizations can maintain consistent and standardized data across the board with centralized data storage. This promotes data integrity and reduces the risk of discrepancies or inconsistencies. However, this often leads to potential scalability issues and can hinder adaptability to changing business requirements or technological advances. While the unified approach provides streamlined management and standardization, it can also lead to scalability issues, especially as the volume and diversity of data increases [24]. In addition, adaptability within this structure can be limited, especially with rapid functional changes or evolving business needs. While the centralized nature provides consistency, it can occasionally act as a bottleneck that limits innovation and responsiveness to new data requirements.

The decentralized data platform or Data Mesh is a novel architectural approach that treats data as a product [25]. Rather than centralizing data in one place or team, responsibility for the data is decentralized and distributed across multiple teams or domains. This approach is consistent with the way modern software is developed and deployed, which is often in smaller, autonomous teams. The Data Mesh paradigm promotes the idea that teams that produce data should also take responsibility for, maintain, and deliver their data products, just as they would a software product. The data products are made available to data consumers through endpoints called Data Marts [26]. This shift in ownership and responsibility helps organizations scale their data infrastructure and practices while ensuring data quality and accessibility [27].

The main purpose of any data platform—and one of the main motivations behind developing the Data Mesh architecture—is to support a process called Data Democratization [28,29]. This means making digital information easy to access and use by all categories of users. The idea behind data democratization is to let people who are not experts collect and

analyze data on their own. The end goal is to create a system that provides data to new applications and services and functions as the main data integration layer. In addition, building a data platform usually involves setting up a data system or infrastructure [30]. On the other hand, using a data platform implies making applications and services that use this data system [30]. Data platforms are central in today's data industry strategies, helping to transfer data from where it is collected to its many uses. It is important for both experts and non-expert users alike to understand how these platforms work to make the most of our data-driven world.

2.1. Data Observatories Roles as Comprehensive Data Services

Data Observatory is a centralized web-based facility that supports data-driven research and decision-making, providing an integrated environment where data are not only accessible but also usable and interpretable [31–33]. Key to this process is the wide range of functions that a Data Observatory performs.

The key functions of data observatories, separated into core and additional functions, are listed and described in Table 1. The core functions are fundamental to its operation and form the backbone of data observatory services, while the additional functions are not essential to the basic operations.

Table 1. The core and additional functions of Data Observatories.

Function	Description
Core Functions	
Data Collection and Storage	Aggregates diverse data from various sources, providing a central repository for data-driven research and decision-making.
Data Integration	Provides services for integrating data from multiple sources, allowing researchers to combine and analyze data from different experiments, simulations, or observations.
Data Standardization	Ensures that data are consistent and compatible with other data, involving tasks such as converting data to a common format or unit of measurement.
Data Processing and Analysis	Utilizes specialized tools, algorithms, and models to transform raw data into actionable insights and generate new knowledge.
Data Visualization	Offers advanced visualization tools and technologies for exploring and analyzing data findings.
Additional Functions	
	Modeling and Simulation
Data Modeling and Simulation	Provides tools and expertise for building and running simulations, enabling researchers to test hypotheses and explore scenarios using data.
	Training Functions
Training and Support	Provides training and support to researchers through workshops, tutorials, and resources to assist researchers in their data-driven projects.
	Collaboration Functions
Data Sharing	Facilitates data sharing through APIs or other interfaces, promoting collaboration within and outside the organization.
Data Dissemination and Publication	Responsible for disseminating and publishing data, making it available through various means for further analysis and research.
	Data Governance and Security
Data Governance	Establishes robust policies and procedures for data use and access, addressing data privacy and security issues.
Data Security and Privacy	Implements stringent measures to protect the security and privacy of data, preventing unauthorized access and usage.

One of the most important functions is the collection and storage of data, which often includes large amounts of diverse data from a variety of sources. These include data derived from experimental results, simulated models, and field observations. The ability to aggregate such disparate data into a single integrated digital platform is a critical feature of a data observatory. After the data have been ingested, it must be prepared for further processing. This step involves organizing, cleaning, standardizing, and maintaining data in a suitable form for later use.

Once the data have been prepared according to the specific requirements of the domain of usage—such as the field of humanitarian demining—it can be processed and analyzed. A data observatory provides the necessary tools and expertise to facilitate this. It uses specialized algorithms and models to transform complex data into tangible insights and generate new knowledge. This process is critical to transforming raw data into information that can drive research and decision-making.

Equally important is the role of the data observatory in visualizing and communicating data. By providing advanced visualization tools and technologies, researchers are able to explore and analyze data more effectively. In addition, these tools help communicate their findings to other researchers, various stakeholders, or the public.

During the data observatory's development phase, a set of standard reports is usually created. These reports are intended to address the general requirements of most users by providing a comprehensive overview of the data collected and analyzed at the observatory. Standard reports typically include key metrics and trends that are required for a comprehensive understanding of the problem domain. The standardization of these reports ensures that all users have access to consistent and reliable data, which is crucial for comparative analysis and benchmarking.

However, as stakeholders interact with the observatory and become more familiar with the available data, they often recognize the need for more specialized reports. These specialized reports address specific questions identified after the initial production phase. The observatory's structure must be adaptable to enable the creation of these customized reports, providing users with the information they require to make informed decisions. This adaptability ensures that the observatory remains relevant and useful to its users over time.

Data observatories also emphasize data sharing and provide access to data through Application Programming Interfaces (APIs) or API gateways. In some cases, data can also be exported in various formats for further analysis. The ability to share data efficiently allows researchers to collaborate more effectively within their organizations and with external stakeholders. Mutual relationships between the Core and Additional functions of data observatories are illustrated as a UML activity in Figure 1.
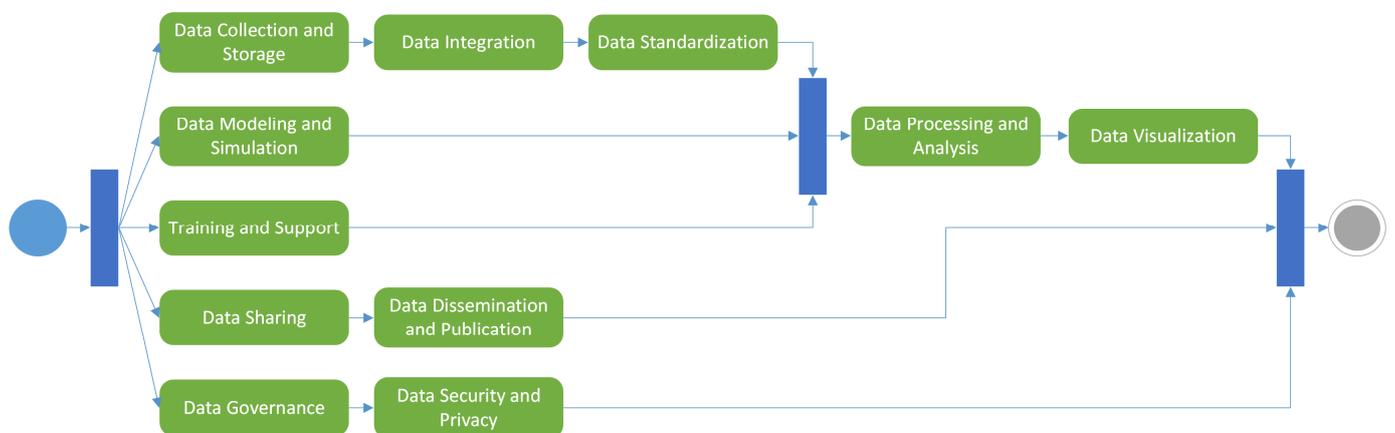
Figure 1. A UML activity diagram of the core (data collection and storage, data integration, data standardization, data processing and analysis, and data visualization) and additional functions (training and support, data sharing, data dissemination and publication, data governance, data security and privacy) of data observatories with their dependence interrelationships.

As can be seen in Figure 1, in addition to core functions, a data observatory also performs several additional tasks. These include curating and managing data to ensure that it is organized, cleaned, and maintained in a form suitable for analysis and use. Tasks such as data standardization, data integration, and metadata creation may be part of this process.

Data observatories often provide training functions that support researchers by offering workshops, tutorials, and resources to assist researchers with their data-driven projects. This subset of additional functions ensures that users can take full advantage of the data, tools, and technologies made available by the observatory.

Another role that data observatories often fulfill is facilitating collaboration and networking among researchers. Organizing events, meetings, or other opportunities for researchers to network fosters an environment where ideas, expertise, and data can be shared. Data integration is another important service provided by data observatories, allowing researchers to combine and analyze data from different experiments, simulations, or observations. They also provide standardization of data to ensure its consistency and compatibility with other data.

Data observatories also facilitate data modeling and simulation by providing tools and expertise for creating and running simulations. This enables researchers to test hypotheses and explore scenarios using data.

Finally, the set of functions defining data security and privacy is of great importance. Data observatories take strict measures to protect the security and privacy of data. In this respect, data governance is an important area of responsibility for a data observatory. This includes establishing sound policies and procedures for data use and access and addressing privacy and data security issues. This is important to preserve the integrity of the data and ensure its appropriate use. The function of data dissemination and publication is integral to their role, making data available through various channels for further analysis and research.

2.2. Data Lakes as Large Storage Repositories of Differently Structured Data

The data lake is essentially a large storage space designed to hold both structured and unstructured data, storing it in its raw format [34–37]. One of the most important features of a data lake is its ability to store data without a predetermined schema or structure. This aspect provides flexibility and facilitates ingesting data from a wide range of sources in its original (i.e., raw or native) format.

A landing area, often referred to as a staging area, is the Initial section within a data lake into which the data are first loaded [38]. This area serves as a temporary storage location for raw data ingested from various sources. In a typical workflow, the data passes through this landing area into the data lake, where it is stored in its original format without any significant transformation or processing. The main function of the landing area is to act as a buffer zone for incoming data so that it can be captured in its unaltered state. This is particularly important to preserve the integrity and granularity of the original data, which can be crucial for subsequent comprehensive analysis. By storing data in this unprocessed form, the data lake ensures that all original data attributes are preserved, enabling a wide range of uses and analyses.

From the landing area, the data can then be processed, transformed, or moved to other areas within the data lake for more structured storage and analysis. For example, the data can be cleansed, categorized, and then transferred to a formalized storage area within the data lake, where it can be organized in a more query-friendly format. Alternatively, the data in the landing area can be accessed directly and analyzed using advanced analytics and machine learning tools, especially when raw and unstructured data analysis is required.

In addition to its storage capabilities, a data lake also offers scalability and flexibility so that new data sources can be included, or larger amounts of data can be ingested as needed. This makes data lakes a suitable solution for companies that generate large amounts of data and need a scalable solution for storing and managing it. By storing data in a centralized, scalable, and cost-effective manner, data lakes enable organizations to access and analyze

their data with much less effort in support of microservice architectures. This also facilitates insight extraction and data-driven decision-making.

2.3. Improving Data Lakes with Ontologies

To see how data lakes can be improved with formal knowledge representation methods such as ontologies, it is first necessary to define the associated terms. By definition, ontology is a formal, explicit specification of a shared conceptualization of a domain of interest [39]. An ontology contains a set of concepts within a domain and the relationships between these concepts. It is a formal vocabulary of terms representing concepts in a particular domain, along with the relationships between them. Importantly, ontologies are used to reason about the described concepts, i.e., to derive implicit knowledge from the existing explicit knowledge and automated inference utilizing reasoning engines [40]. In addition, ontologies are often used to provide a common understanding of a domain between different people or systems [41–43].

Depending on their complexity, scope and purpose, ontologies are divided into (1) general and (2) specialized domain ontologies. A general ontology provides a foundational knowledge framework that can be applied to different domains. It contains basic concepts and relationships that are universally applicable, such as spatiotemporal relationships, logical concepts, or basic entities such as objects, events, and actions. On the other hand, a specialized domain ontology is tailored to a specific area of knowledge or field of study. It captures the unique concepts, relationships and specific rules relevant to that domain.

Typically, ontologies are stored in knowledge bases (KBs), which are repositories that facilitate the formal representation of domain-specific knowledge through concepts, relationships, and constraints defined in the vocabulary defined by the utilized ontologies. By using automated reasoning services, KB facilitates automated reasoning processes by providing the necessary infrastructure to infer new knowledge, validate existing information, and execute complex queries [44–47].

In the context of improving data lakes, particularly for their application in non-technical surveys for mine action, it is beneficial to consider adding an ontology layer. This layer represents a major advance in the way we manage and use large data sets such as data lakes [47,48].

The ontology layer enables the formal representation of knowledge and transforms the raw data stored in the data lake into structured, meaningful information. Ontologies provide a common vocabulary that defines types, properties, and relationships between entities in a specific domain. In the context of demining, these could be concepts represented in multimodal aerial and hyperspectral satellite imagery, indicators of the presence or absence of mines, contextual data, terrain analysis information, and experiential knowledge about the battlefield.

Beyond the organization of knowledge, introducing an ontology layer also opens the possibility of automated reasoning using expert systems. These systems can use the structured knowledge provided by the ontology to use existing declarative knowledge, or facts, and imperative knowledge in the form of rules to draw conclusions and thus derive new knowledge. This functionality is particularly valuable in the context of demining operations, where efficient data discovery and accurate decision-making are important.

3. Related Work

In [31], the Forest Observatory is presented, a linked data store designed to integrate and represent wildlife data semantically. It focuses on an extensive wildlife sanctuary where a large amount of diverse Internet of Things (IoT) data are generated. This research uses semantic web technologies to address the problem of heterogeneous and isolated wildlife data. The Forest Observatory Ontology (FOO) is developed to semantically model and link data sources, enhancing data accessibility and enabling complex queries about wildlife. The paper evaluates the ontology and its application, demonstrating its potential in aiding wildlife research and decision-making.

Regarding the rationale behind data observatories, a paper [32] examines the concept of the Web Observatory, a global infrastructure project designed to promote the sharing and use of web-related datasets and analytics applications for research and business. It emphasizes the need for a distributed infrastructure for big data analytics that enables the retrieval of common datasets without replication and the reuse of analytics tools for different datasets. The project aims to bridge the gap between big data analytics and the web of big data, facilitating innovation and digital literacy by making data accessible to a wider audience.

In [47], researchers focused on the development of a semantic data lake to address the challenges of analyzing data across multiple enterprise collaboration systems (ECS). This data lake ingests data in real time and uses ontology-based data access for harmonization, allowing for efficient analysis and integration of diverse data structures and formats. The approach improves business intelligence by allowing identical SPARQL queries to be run across multiple systems, simplifying data access and analysis.

The paper [48] presents a methodology for enhancing data lakes with semantic layers. It demonstrates the process of converting data from different sources into a unified knowledge graph that facilitates advanced data queries and exploration. The approach includes the integration of entity linking techniques for text data, using a domain-specific ontology, and using RDF Mapping Language (RML) for data transformation. The result is a comprehensive knowledge graph that combines structured, semi-structured and unstructured data, offering significant benefits for data analysis and business intelligence. Thus, incorporating an ontology layer into a data lake represents a significant upgrade.

In [49], the creation of a comprehensive database with around 1000 entries for ammunition, mainly from the Second World War, is presented. This database, developed in collaboration with SENSYS and using the MuniMan software (Information about the MuniMan software is available at: https://sensysmagnetometer.com/), also includes 250 datasets describing the modern Warsaw Pact and NATO ammunition. The corpus covers regions in Europe, North Africa, and Southeast Asia and provides detailed information on ammunition types, periods of use, countries that manufacture ammunition and identification data. This tool was designed to help experts identify unexploded ordnance and support decision-making in ordnance disposal. However, the database does not include an advanced data infrastructure that utilizes web-based technologies, nor does it make use of ontologies or other technologies to formally represent knowledge about unexploded ordnance (UXO).

The development of a sophisticated decision support system for humanitarian demining has been previously described in [12,50]. The Advanced Intelligence Decision Support System (AIDSS) integrates various advanced technologies, including remote sensing and data fusion, to support decision-making processes in identifying and clearing mine-affected areas. The AIDSS aims to provide a reliable, efficient tool to support the process of making decisions about suspected hazardous areas based on the methodology scientifically developed and validated in the EU FP5 SMART project and upgraded in the EU FP7 TIRAMISU project. The paper emphasizes the potential applications of this system in non-technical surveys, demonstrating its innovative contributions to improving the efficiency and safety of humanitarian demining operations.

In our previous research, the authors introduced MINEONT, a novel ontology developed for mine action, particularly in the context of non-technical surveys for humanitarian demining [51]. MINEONT has been developed using the OWL-DL 2 formalism, which provides an expressive and formal representation of mine action concepts, high-level semantics, geospatial metadata and information from remote sensing non-technical surveys. This ontology supports data such as multi-sensory imagery, mine presence indicators and mine action expertise. The paper discusses the construction of this ontology and its potential for improving decision support systems in the field of mine action and highlights its advantages over existing methods.

The authors have also explored new approaches to improve the efficiency, accuracy, and safety of humanitarian demining in previously published research [33]. To improve the process of the non-technical survey in humanitarian demining, the introduction of a data observatory and data lake system was proposed. This system would be able to process large amounts of unstructured data to improve decision-making through the use of artificial intelligence, deep learning and data analysis techniques. Key features include data ingestion pipelines, transformation techniques, and blockchain technology for data integrity, interoperability, data analysis, security, and data governance. The envisioned approach has the potential to transform humanitarian demining and make it more effective and safer. Furthermore, in our latest study [52] we have proposed to use blockchain, specifically Filecoin and NFT.Storage (NFT.Storage is a freely available decentralized storage service available at: https://nft.storage/), for storing landmine and UXO locations. This novel, interdisciplinary approach ensures secure, reliable, and decentralized data storage for humanitarian demining. The use of blockchain databases for tracking UXO information overcomes challenges such as scalability, integration, legal compliance, and cost. Using blockchain for minefield records can improve the efficiency, accessibility, and safety of demining operations. In this research, we have outlined the concept, its implementation, and potential benefits and highlighted the role of blockchain in improving demining processes.

4. MINEONT+ Ontology

The MINEONT+ ontology, which stands for "MINE-action ONTology Plus," is a core ontology specifically designed to formally represent knowledge in the aerial non-technical survey domain. This ontology, expressed in the OWL-DL 2 formalism, provides a comprehensive and formal representation of concepts related to non-technical surveys in humanitarian demining. Its design ensures an accurate representation of relevant knowledge in this domain and functions as the fundamental component of the Minefield data observatory.

More precisely, the MINEONT+ ontology model encapsulates high-level semantics, geospatial metadata, and information for a non-technical survey acquired through various sensors included in remote sensing. The ontology's vocabulary defines formal concepts, which include but are not limited to multisensory aerial and satellite imagery, indicators of mine presence and absence, terrain analysis information, and formalized knowledge of humanitarian demining specialists. The MINEONT+ model presented in this research was developed as a continuation of the previous ontology [51], which had more limited possibilities to present knowledge about different types of UXO.

The construction of the MINEONT+ knowledge database, a prerequisite for utilizing the ontology for minefield information consolidation and formalization, involves two phases: data acquisition and data processing. As already explained, inputs for constructing the knowledge database for non-technical surveys can come from numerous sources, such as minefield records, mine accident maps, interviews, surveyor reports, military maps, and more. These data inputs are characterized by diverse formats, types, and structuring levels, presenting a challenge in harmonizing and integrating this disparate information into an integrated KB. After acquiring and adequately processing the required data, it can be stored in the ontology model of the integrated knowledge database.

The MINEONT+ ontology model, is designed to provide a formal and comprehensive representation of a minefield containing different UXOs. As can be seen in Figure 2, the model is structured around several key concepts and their interrelationships. First-level concepts that are directly subsumed under owl:Thing class are: Minefield, MinefieldRecord, and MinefieldDrawing. Additional key concepts are: DeminingOperation, MachineLearningAlgorithm, GeospatialData, MinefieldIncident, OrientationPointList, OrientationPoint, UXOList, UXO, and co:ListItem.
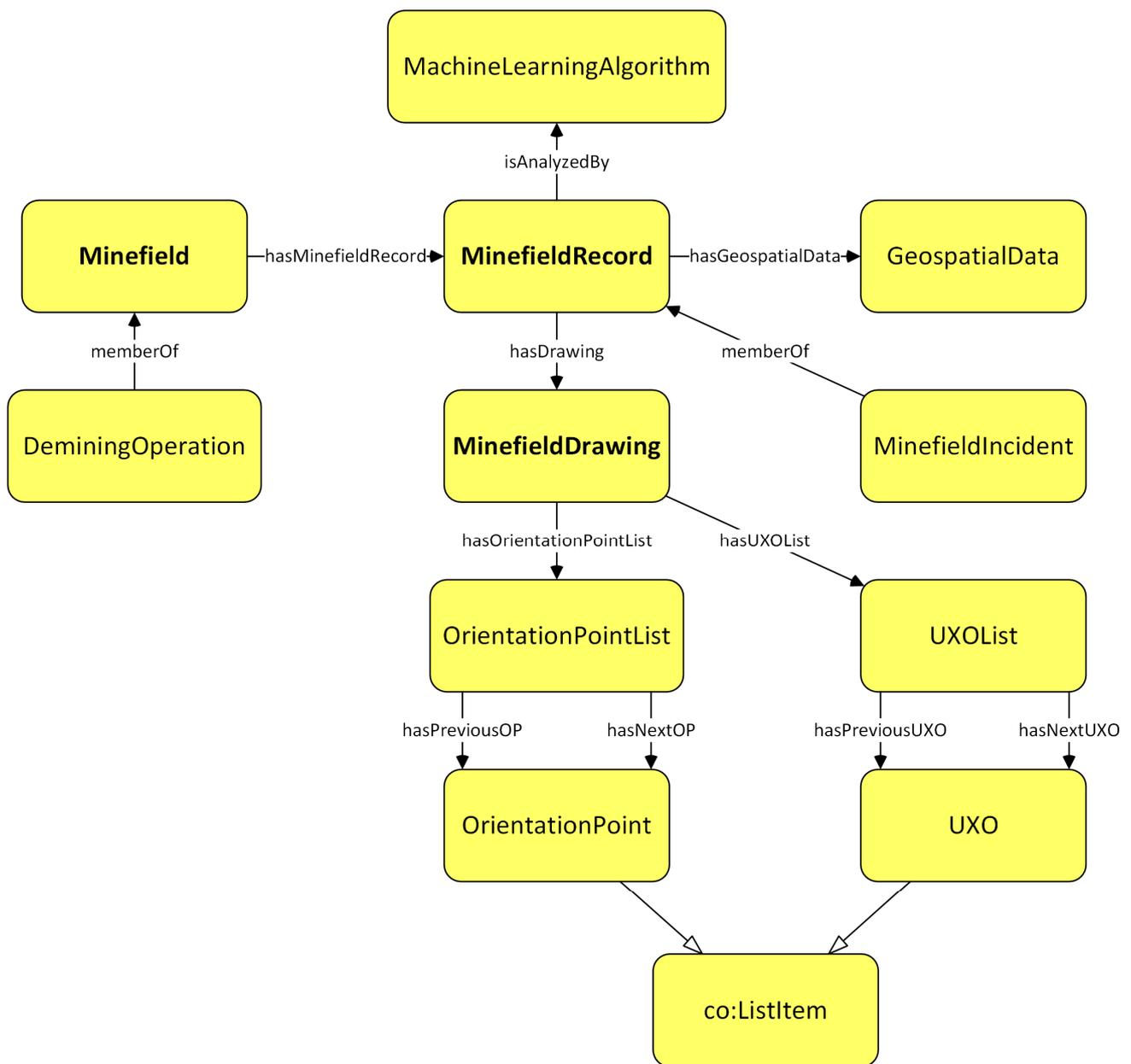
Figure 2. The MINEONT+ model with top-level concepts and the most important object properties. The model is defined in OWL 2 DL format.

At the highest level is the Minefield concept, which can contain many Minefield-Record instances in the knowledge database. Each MinefieldRecord has a GeospatialData instance. Furthermore, data stored in a MinefieldRecord instance can be analyzed by one or many machine learning algorithms. This relationship is captured by isAnalyzedBy object relationship between MinefieldRecord and MachineLearningAlgorithm concepts. An instance of a MinefieldRecord is related to one or more instances of MinefieldDrawing, OrientationPointList, and UXOList classes through their respective object relationships. In formal OWL 2 DL terms, this can be expressed as:

$$\text{MinefieldRecord} \equiv \exists_{\geq 1}(\text{MinefieldDrawing}) \sqcap \text{OrientationPointList} \sqcap \text{UXOList} \quad (1)$$

The concept "Minefield" has functional attributes dataLastInspected and riskLevel indicating the data and time of a particular landmine location's last inspection and estimated risk level. This concept also has attributes of location, name, and status for the name of the loca-

tion, the minefield and its status or state, respectively. Concept GeospatialData has functional attributes to store the minefield's geographic latitude, longitude, and inclination level.

Each UXO in the MinefieldDrawing is represented in the knowledge database ABox as exactly one instance of the UXO concept. This concept can have labeled attributes such as name, description, and quantity.

The UXOs are organized into sequences, with the first member of the sequence attached to a UXOList individual using the co:firstItem object relation. Each subsequent member in the list is linked to the previous one with the co:nextItem object relation until the last member is denoted using the co:lastItem object relation. Each list item also has its index that uniquely identifies it.

In OWL 2 DL, the relationships between UXO, UXOList, and the co:firstItem, co:nextItem, and co:lastItem object relations can be expressed as:

$$\text{UXOList} \equiv \exists(\text{co:firstItem.UXO}) \sqcap \exists(\text{co:nextItem.UXO}) \sqcap \exists(\text{co:lastItem.UXO}) \quad (2)$$

Each MinefieldRecord is divided into one or more sequences of UXO individuals, which contain a particular explosive device or remnant of war. Sequences have their index and can be numbered. Similarly, a single MinefieldRecord instance can have at least one OrientationPoint individual. UXOs and OrientationPoints are hierarchically organized into sequences.

In OWL 2 DL, the relationships between concepts MinefieldRecord, UXO, and OrientationPoint can be expressed as:

$$\text{MinefieldRecord} \equiv \exists_{\geq 1}(\text{UXO.Sequence}) \sqcap \exists_{\geq 1}(\text{cOrientationPoint.Sequence}) \quad (3)$$

Each MinefieldRecord is associated with one or more UXO sequences and one or more OrientationPoint sequences. By defining an instance of the MinefieldIncident concept, each minefield can be associated with one or more minefield incidents. Each individual has attributes indicentType, casualties, incidentData and reportBy. Because demining operations can be very complex and time-consuming, the attributes progress, operatingOrganization, startDate, endDate, and operationName can be used to describe them. The MINEONT+ model with top-level concepts and the most important object properties is shown in a diagram in Figure 2.

These formal expressions in OWL 2 DL define the relationships between the key concepts in the MINEONT+ model, providing a foundation for automated reasoning and query processing.

4.1. Minefield Observatory Microdata Schema

In the context of web applications, the structuring of a Microdata schema is important for the practical implementation of the Minefield Observatory system. By using the capabilities of HTML5 standard together with the Microdata API, minefield records stored in the data observatory can be presented and exported in a standardized, open, and easily accessible format. The structured schema based on Microdata enables the transfer of complex datasets in machine-readable formats that facilitate their use in other academic research and demining applications.

The Microdata API is critical for connecting the data observatory to other information systems and sending detailed information about minefields. This integration is essential for making the knowledge database content accessible to other systems, thereby expanding the reach and impact of the data collected in humanitarian demining efforts.

The usage of the Microdata API within the Minefield observatory always involves three steps. First, the selection of the concepts from the MINEONT+ ontology will be represented. Second, the data must be fetched from the KB using SPARQL queries. In addition, finally, a new HTML5 document is generated, and the semantic data are embedded within the HTML document structure.

The Microdata schema developed for the Minefield Observatory is comprised of 8 key elements:

- Minefield (itemscope): Represents the overall context of the data.
- Location (itemprop): Specifies the geographical coordinates of the minefield.
- Status (itemprop): Indicates whether the minefield is active or inactive.
- DateLastInspected (itemprop): Marks the date when the minefield was last inspected.
- RiskLevel (itemprop): Describes the assessed level of risk associated with the minefield.
- DeminingOperation (itemprop): Details an event that encompasses efforts to clear the minefield.
- MinefieldIncident (itemprop): Describes incidents that have occurred within the minefield.
- WarRemnant (itemprop): Provides information on remnants of war found within or around the minefield.

The structure of the Microdata schema is shown in Figure 3. An example of schema usage with HTML5 code is shown in Section 6.2.
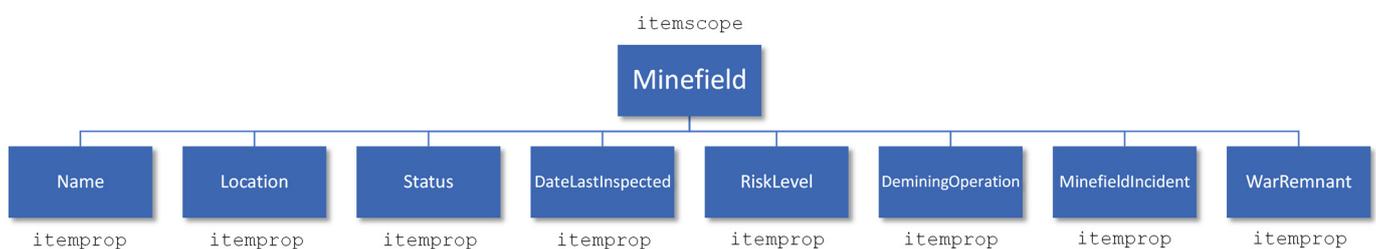
Figure 3. The HTML5 Microdata schema developed for the Minefield Observatory.

Within the top-level Minefield element (itemscope), the subsumed element (itemprop) "name" specifies the name of the minefield. The geographical coordinates (latitude, longitude) are given under "location" to identify the exact location. The "status" itemprop element indicates the current state of the minefield. The "DateLastInspected" itemprop element notes the date of the last inspection of a particular minefield, while "RiskLevel" itemprop element indicates the associated risk. The "DeminingOperation" is itemprop element group and contains comprehensive details of ongoing demining activities. The "MinefieldIncident" itemprop element group records historical data on incidents, and "War-Remnant" itemprop group describes any war remnants found, including their discovery and disposal status. Each element is crucial for a detailed representation of the minefield. War remnant is a larger term that encompasses UXOs and other related non-explosive material that might be left in the ground after cessation of hostilities.

Within the "DeminingOperation" group, marked by itemprop attribute, each subsumed element has a specific purpose: "operationName" assigns a label to the demining operation, "startDate" and "endDate" denote the operation's commencement and conclusion dates, respectively, "operatingOrganization" describes the entity overseeing the operation, and progress reflects its current state or completion level. Similarly, in the "MinefieldIncident" group, the element "incidentDate" specifies the date of the incident, "incidentType" describes the nature of the incident, "casualties" details the impact or outcome of the incident in terms of human harm, and "reportBy" identifies the organization or authority that reported the incident. Finally, the "WarRemnant" itemprop group element "type" identifies the kind of war remnant, "foundDate" indicates the date the item was discovered, and "disposalStatus" describes the current status of its disposal process.

5. Minefield Observatory Structure

This section explains the architecture of the Minefield Observatory and the integration of its different components, including the ontology knowledge database. A thorough evaluation of the Minefield Observatory is conducted using a task-based methodology and applying it to a real-world scenario representative of its intended use. This methodology enables the evaluation of the functionality and effectiveness of the observatory in a real-world environment and gives us insights into its performance and utility for UXO demining.

The ontology-based paradigm for minefield annotation consists of terminological and assertional knowledge about high-level SHA description and a reasoning engine. These two types of knowledge are the basic components of a knowledge-based system based on Description Logics (DLs) [44] as a set of structured knowledge-representation formalisms with decidable reasoning algorithms. DLs represent important notions about a domain as concept and role descriptions. To achieve this, DLs use a set of concept and role constructors on the basic elements of a domain-specific alphabet. This alphabet consists of a set of individuals (objects) constituting the domain, a set of atomic concepts describing the individuals and a set of atomic roles assigned to the individuals. The concept and role constructors that are employed indicate the expressive power and the name of the specific DL. Here, we use $\mathcal{SHOIN}(\mathcal{D})$, on which OWL 2 DL is based that employs concept negation, intersection, and union: existential and universal quantifiers, transitive and inverse roles, role hierarchy and a number of restrictions. Since OWL Lite is semantically very limited and OWL 2 Full is undecidable, OWL 2 DL represents a compromise between adequate expressivity and guaranteed decidability. Most importantly, a variety of tools for knowledge engineering exist [53,54] that allow construction, management, reuse, and reasoning with OWL-based ontologies. As such, OWL 2 DL is a suitable ontology language for representation and reasoning about high-level minefield description.
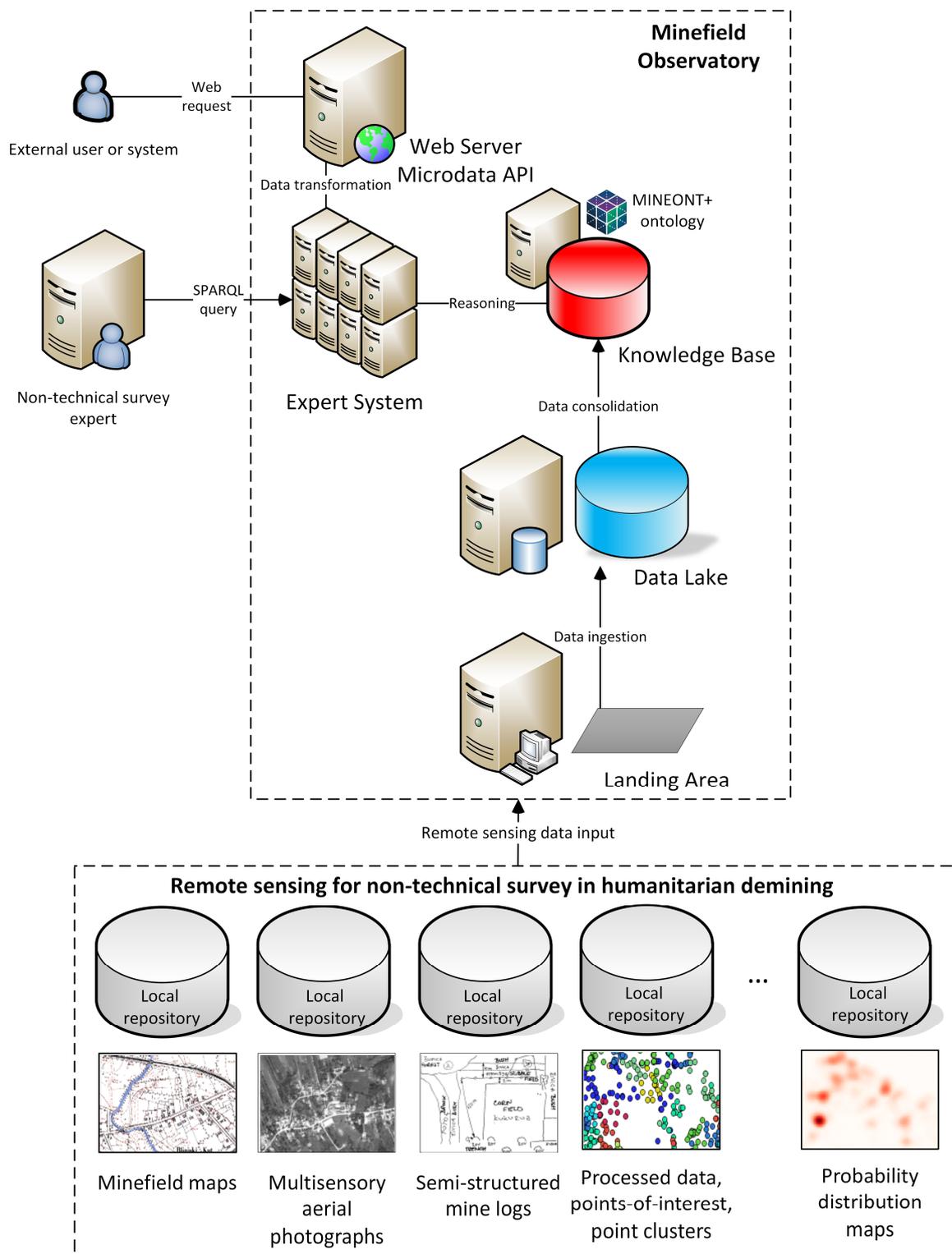
In the context of humanitarian demining, data originates from a variety of sources, such as remote sensing data, multimodal aerial and hyperspectral satellite imagery, indicators of the presence or absence of mines, contextual data, terrain analysis information and battlefield experiential knowledge and can be collected and stored in the data lake. This diversity and richness of data sources is particularly beneficial in the context of non-technical surveys for demining, as it provides a comprehensive picture of the mine area. However, in the context of data engineering efficiency, such a variety of data sizes, types and formats is not desirable. Therefore, a solution that includes data lakes is a desirable choice.

A data lake can often serve as a central repository within a data observatory where raw data collected from these various sources can be stored. This raw data can then be accessed and examined by researchers and data scientists to gain insights and support decision-making processes. In addition, a data lake can also be used to store the results of data analyses so that researchers and analysts can access and use them later.

The Architecture of the Minefield Observatory containing a landing area and data lake, is shown in Figure 4.

The KB for ontological representation of minefield records within the Observatory has two main components. The terminological component (TBox) describes the relevant notions of the application domain by stating the properties of concepts and roles and their interrelations. Tbox contains an ontological representation of the knowledge about SHAs. The assertional component (Abox) is a formal set of assertions describing specific information or semantics in terms of terminological knowledge. Abox describes a concrete world by stating individuals and their specific properties and interrelations.

The annotation process of a suspected minefield begins with the identification of concepts in its content that can be observed by subjects and deemed important (by demining experts) for the description of SHA. After a concept is recognized, an equivalent concept must also be identified in the ontology used for SHA representation. Tbox must define all concepts that exist in the SHA semantics. After an equivalent concept has been found, a new individual is created, associated with the minefield and stored in Abox. This process is repeated for all minefields in consideration within the Observatory KB. Retrieving knowledge about mine records assets in the proposed architecture can be easily achieved by using semantic query languages such as the SPARQL query language [55]. The next figure (Figure 5) illustrates a SPARQL 1.1 query that might be posed by an expert system using the MINEONT+. In this example, instances of MinefieldRecord class are retrieved from the KB, specifically their labels, creation dates, and validity statuses.

Figure 4. Architecture of the Minefield Observatory containing a landing area and data lake in the context of non-technical survey and humanitarian demining. The landing area receives differently structured and semi-structured data acquired from remote sensing of a SHA. The data lake is utilized as a repository of differently structured raw data in native formats. After the data have been consolidated, the knowledge base (KB) and MINEONT+ ontology layer above the data lake perform semantic integration of the acquired data.

```
PREFIX ex: <http://example.org/mineont#>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>

SELECT ?minefieldRecord ?label ?dateCreated ?isValid
WHERE {
    ?minefieldRecord a ex:MinefieldRecord .
    OPTIONAL { ?minefieldRecord ex:label ?label . }
    OPTIONAL { ?minefieldRecord ex:dateCreated ?dateCreated . }
    OPTIONAL { ?minefieldRecord ex:isValid ?isValid . }
}
ORDER BY ?dateCreated
LIMIT 10
```

Figure 5. A SPARQL 1.1 query for retrieval of MinefieldRecord instances from the KB.

Prefixes ex: and xsd: are defined to provide the URI references for ontology classes and XML Schema Definition (XSD) properties in the SPARQL query, respectively. Variables (?minefieldRecord, ?label, ?dateCreated, ?isValid) are selected, representing the minefield record ID, its label, the date it was created, and its validity status. Clause OPTIONAL is used to account for the possibility that some records might not have all the properties ("label", "dateCreated", "isValid"). Results are ordered in ascending order by the ?dateCreated property to get a chronological list of minefield records. This query may be executed in the Protégé ontology editor [56,57] extended with the Jess rule engine [58]. The forward chaining search strategy should be used to maximize the number of returned tuples and multimedia documents [59].

6. Use-Case Example

Minefield records are the most reliable source of information regarding minefield presence, location, shape, and content (Figure 6). The Croatian MAC experts analyzed 122 mine records from Croatia's municipality Blinjski Kut and attempted to locate their positions in space using the information extracted from them [60]. These mine records contain 39 different types of information about the laid minefield, which are classified into five categories: (1) cartographical data (name and scale of the map, etc.), (2) data for orientation and positioning of minefield (coordinate of referent point, etc.), (3) type and number of mines in minefield, (4) characteristics of minefield (type and dimension of minefield), and (5) information about mines placement (date, military unit, responsible person, etc.). All of the above information is required for experts to analyze the mine scene and determine the SHA.

Examples of minefield records associated with the Blinjski Kut minefield are shown in Figure 6. The document sets contain hundreds of similar semi-structured and hand-drawn records for several different locations in Croatia. The minefield records represent the painstaking efforts of demining teams and local communities to document the presence and location of landmines within the Blinjski Kut minefield. Each record in the collection provides crucial information about the type of landmine, its condition, and any additional relevant details that aid in the safe removal and clearance of the area. The records in Figure 6 highlight the diverse range of landmines present in the Blinjski Kut minefield. Some entries detail traditional anti-personnel mines, while others document more sophisticated anti-tank mines. The variation in explosive devices underscores the complexity of the demining task at hand [61].
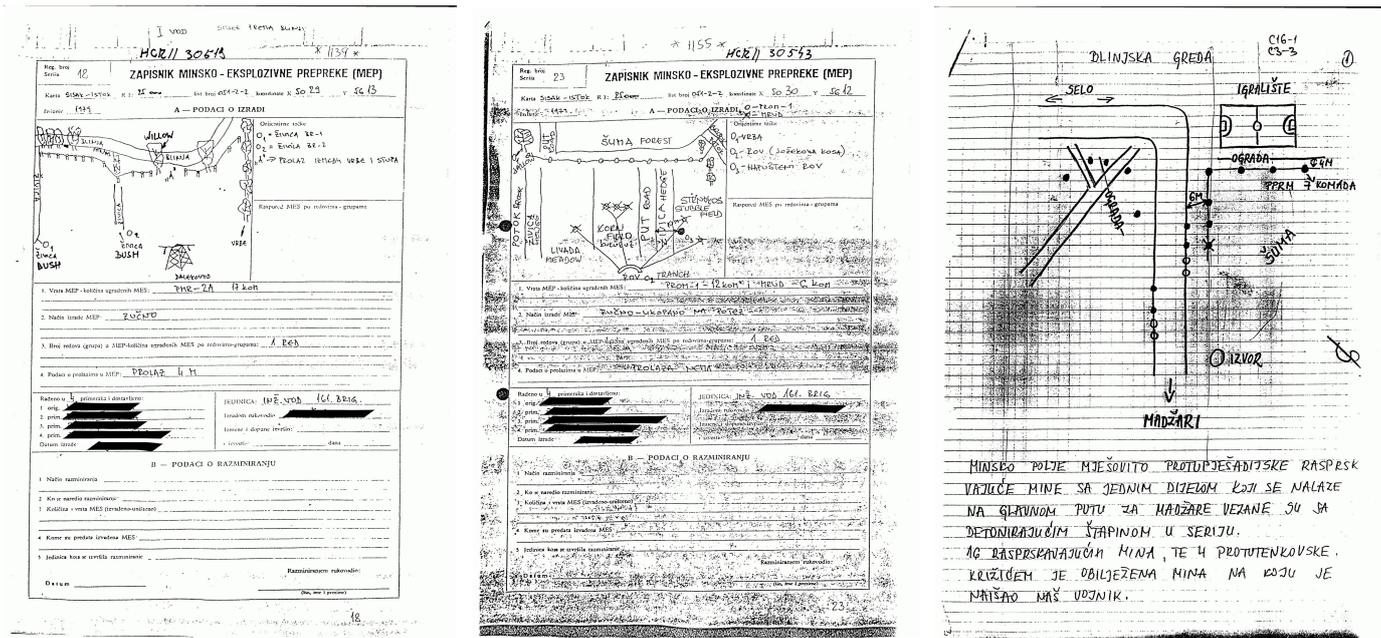
Figure 6. Examples of minefield records associated with the Blinjski Kut minefield. The actual set of documents contains hundreds of similar semi-structured and hand-drawn records. Personalized information was blacked out.

As already explained, the minefield records serve as a historical record of the minefield's composition and play a crucial role in ongoing demining operations [62]. Demining teams use this documentation to track their progress, adjust their strategies as needed, and ensure that each section of the minefield is thoroughly cleared before declaring it safe for public use. As such, this information represents a valuable source of information that should be extracted and organized in a suitable manner.

The semi-structured nature of these records reflects the challenging conditions under which demining operations are conducted [63]. In many cases, demining teams use a combination of sketches, written descriptions, and photographs to create a comprehensive overview of the minefield. This information is vital for developing effective demining strategies, ensuring the safety of personnel involved, and mitigating the risk to local populations.

6.1. Deep Learning-Based Mine Records Ingestion

The hand-drawn sketches accompanying the records provide visual representations of the minefield layout [64]. These sketches often include symbols and annotations that convey the location of each mine, the depth at which they are buried, and any observed changes in their condition over time. Such information is invaluable for creating detailed maps that guide demining efforts and help minimize the risk of accidents during the clearance process.

The high-level overview of the entire process of text extraction is shown in Figure 7. The process of extracting data from images featuring minefield records begins by leveraging advanced techniques such as identifying potential regions of interest within the documents. This is accomplished through the application of the YOLO ("You Only Look Once") deep learning model, which not only identifies but also labels these potential regions [65]. Subsequently, these labeled regions are cropped into smaller, more manageable images. Figure 7 illustrates a snapshot of this initial stage, with each region uniquely colored to signify distinct labels.
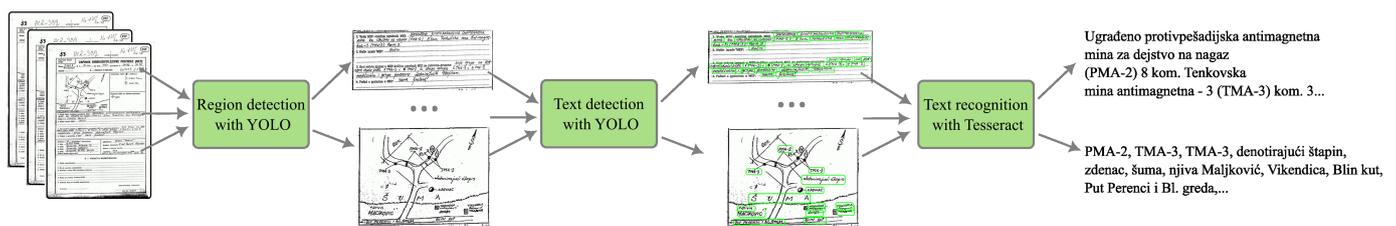
Figure 7. The high-level overview of the text extraction process using computer vision methods based on deep learning network architectures.

Breaking down the extraction task into smaller, more focused activities enables the utilization of various tools for addressing specific aspects of the process. The effectiveness of this segmentation is evident in the example image, where each color-coded region represents a different label associated with the content it encapsulates.

Following the successful detection and cropping of these regions, the subsequent step involves identifying specific textual information within them. This process allows for more targeted and efficient data extraction as the focus shifts to discerning relevant details embedded within each segmented region. The utilization of advanced technologies not only streamlines this identification process but also enhances the accuracy and precision of text extraction from the minefield records.

In essence, this multi-stage approach, incorporating YOLO-based region identification, text detection and subsequent text extraction, facilitates the overall extraction task and optimizes the workflow by breaking it down into more manageable components. The example image in Figure 8 visually represents how this method enhances the efficiency and effectiveness of data extraction from minefield records, contributing to a more robust and accurate analysis of the documented information.

In Figure 9, we depict the information extraction process. The initial step involves annotating minefield records with relevant labels. These labels correspond to designated regions of interest, numbered 1 to 9 in the image. Following labeling, the YOLO algorithm is employed to detect and crop all regions of interest into designated folders, allowing the creation of smaller datasets for subsequent analysis.

Moving to the second phase, we focus on text detection [66]. A specific region of interest is chosen, and annotations are added to identify the specific textual content for extraction, as demonstrated in Figure 9. The objective here is to detect individual words, numbers, or combinations of letters, words, and numbers. Once annotations are complete, YOLO is employed again to train on this data and detect text.

Upon implementation of the custom text detector, the process proceeds to the third phase of text recognition using the Tesseract OCR engine [67,68]. Tesseract boasts Unicode (UTF-8) support and can recognize over 100 languages, including Croatian, the language used in the minefield records. Detected regions undergo text recognition through Tesseract, and the results are formatted appropriately for storage, ready for later use in MINEONT+.

While this procedure applies to most regions in the minefield record, it may not be necessary for hand-drawn maps of the minefield. In such cases, the map can be saved in image format and described in MINEONT+. Visualization tools can then be employed to enhance information management.

Figure 8. An example of labeling the regions of the minefield document that will be split into many smaller images for further processing stages. Personalized information was blacked out.
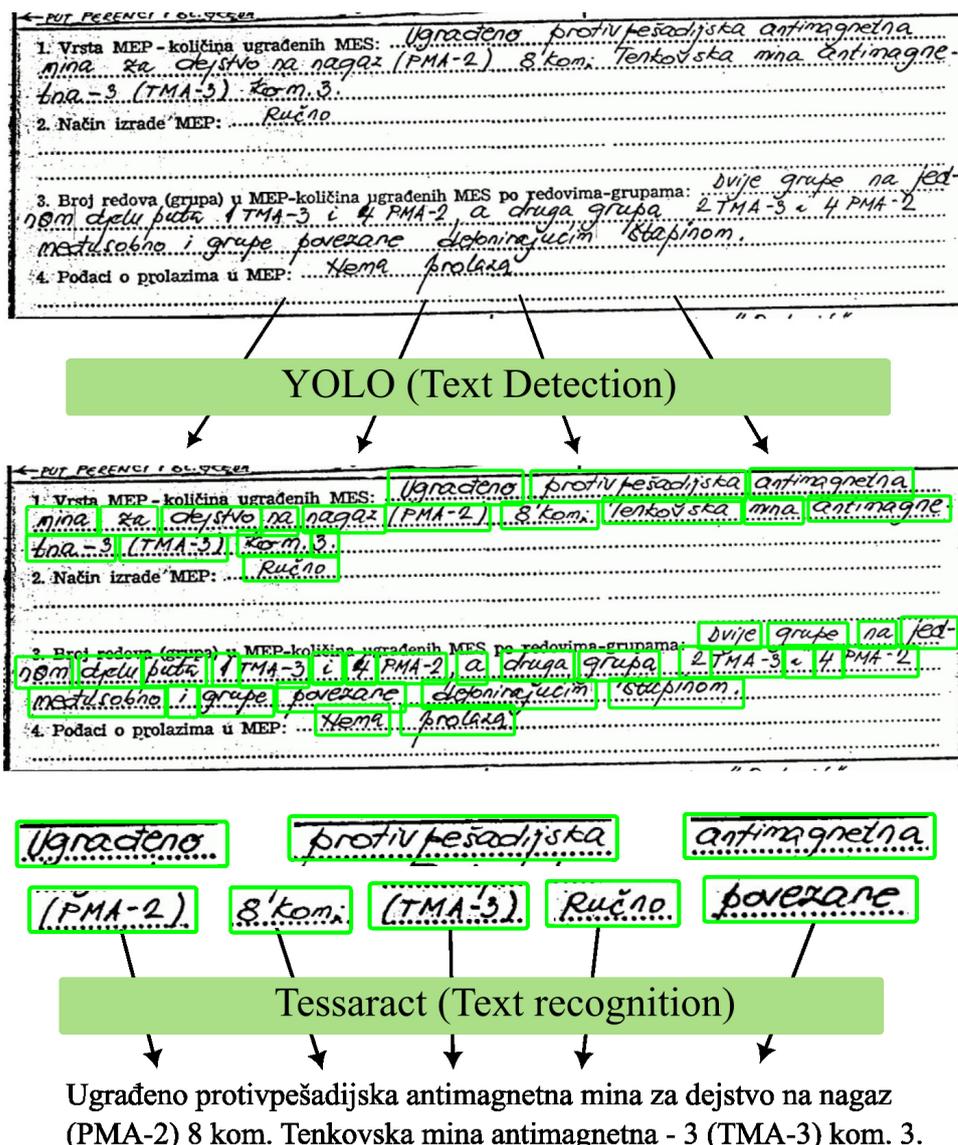
Figure 9. The overview of the information extraction process from the minefield records using computer vision methods.

6.2. Microdata API Example

A HTML5 representation of the Microdata schema for the Blinjski Kut minefield use-case near Sisak, Croatia, is shown as an example in Figure 10. This is only a partial representation of KB content for this use-case that can be provided by the data observatory Microdata API.

This Microdata API gateway is a key component of the minefield observatory, serving as an interface between external clients and internal observatory services. It optimizes client interactions by providing a single point of access to the information stored in the observatory, improving security, managing traffic, and allowing the observatory to provide a consistent and comprehensive data service. Users gain access to a wide range of detailed minefield information via this gateway, allowing for secure and efficient data retrieval and interaction.

```html
<div itemscope itemtype="http://schema.org/Minefield">
   <h1 itemprop="name">Blinjski Kut</h1>
   <div itemprop="location" itemscope itemtype="http://schema.org/Place">
       <meta itemprop="latitude" content="45.4667" />
       <meta itemprop="longitude" content="16.3783" />
   </div>
   <span itemprop="status">Active</span>
   <time itemprop="dateLastInspected" datetime="2003-03-10">March 10, 2003</time>
   <span itemprop="riskLevel">High</span>
   <!-- Additional details for Demining Operations -->
   <div itemprop="DeminingOperation" itemscope itemtype="http://schema.org/Event">
       <span itemprop="operationName">Operation Peace Return</span>
       <time itemprop="startDate" datetime="2003-04-01">April 1, 2003</time>
       <time itemprop="endDate" datetime="2003-12-31">December 31, 2003</time>
       <span itemprop="operatingOrganization">International Demining Group</span>
       <span itemprop="progress">In Progress</span>
   </div>
   <!-- Additional details for Incident Reports -->
   <div itemprop="MinefieldIncident" itemscope itemtype="http://schema.org/Event">
       <time itemprop="incidentDate" datetime="2003-02-15">February 15, 2003</time>
       <span itemprop="incidentType">Detonation</span>
       <span itemprop="casualties">2 injured</span>
       <span itemprop="reportBy">Ministry of the Interior of the Republic of Croatia</span>
   </div>
   <!-- Additional details for Remnants of War Data -->
   <div itemprop="WarRemnant" itemscope itemtype="http://schema.org/Product">
       <span itemprop="type">Unexploded Ordnance</span>
       <time itemprop="foundDate" datetime="2003-03-05">March 5, 2003</time>
       <span itemprop="disposalStatus">Pending</span>
   </div>
</div>
```

Figure 10. Microdata schema for the use-case of Blinjski Kut minefield near Sisak, Croatia, in HTML5 format.

Using the Microdata schema defined in Section 4.1 for the minefield Blinjski Kut example, the snippet in Figure 10 would read as follows: The minefield is located at latitude 45.4667, longitude 16.3783, and is currently active with a high-risk level. The last inspection was on 10 March 2003. Additionally, there is a demining operation titled "Operation Peace Return" conducted by the International Demining Group, which started on 1 April 2003 and has been ongoing as of 31 December 2003. An incident occurred on 15 February 2003, involving a detonation with two injuries reported by the Ministry of the Interior of the Republic of Croatia. There is also a remnant of war, specifically an unexploded ordnance, found on 5 March 2003, which is pending disposal.

7. Discussion

In the Discussion section, we critically examine the Minefield Data Observatory, highlighting its multiple benefits and acknowledging the inherent limitations of our study. Section 7.1 focuses on the significant benefits of using advanced knowledge representation methods like computer ontologies coupled with automated reasoning expert systems. Section 7.2 addresses the limitations of our study, in particular, the constraints of ontologies and the challenges associated with analytically evaluating data on SHA.

7.1. Benefits of the Minefield Data Observatory

The use of advanced knowledge representation methods such as computer ontologies combined with automated reasoning expert systems to formally define concepts and their mutual relationships in the domain of humanitarian demining represents the main benefit of the proposed system. This new approach allows for faster, simpler, and more accurate analysis of all existing semi-structured and heterogeneous information stored within a MAC mine information system. Formal representation of minefield records will also enable the automated discovery of new knowledge in the existing document repositories.

Furthermore, research into using decidable decision methods in computer vision to aid in the semantically rich interpretation of the processed mine scene represents another benefit of the Minefield Observatory system. There have been no significant developments in humanitarian demining in this area due to the very complex scene and objects (e.g., indicators of mine presence like a trench, infantry shelter, bunker, drywall, etc.) that need to be detected and reliably extracted. This tool will be accompanied by Standard Operating Procedures for each step of the process. As a result, MAC operatives do not need to be experts in every single technology-related step. They will benefit from the decision-making system for humanitarian demining. Furthermore, new members of the MAC can be easily educated on this new tool for recognizing the characteristics of SHA. In the end, MAC personnel will make the final decision on what constitutes a mine presence indicator.

7.2. Limitations of the Study

The first potential limitation of the study is related to the usage of ontologies in a supervised learning setting. The process of using an ontology-based KB always relies heavily on the expertise and contributions of domain experts, who must meticulously and manually populate KB with instances of ontology concepts by identifying the correct terminological concepts in the KB formal vocabulary. This requirement presents a major challenge as it requires assembling a group of expert individuals who have a deep understanding of the problem domain as well as the developed ontology vocabulary for the description of the domain. These experts must spend significant time and effort to accurately define and relate complex concepts, which can be labor-intensive and prone to errors. In addition, reliance on human input increases the risk of subjectivity and inconsistencies in the ontology, which can compromise the integrity and usefulness of the KB.

Additionally, once developed, ontologies are essentially static structures rarely changed. Large ontological structures, especially those with many concepts, relationships, and properties, may have scalability problems and be difficult to modify. As such, they may not capture all the dynamic data properties of minefields. Integrating changing terminologies into such rigid knowledge description frameworks may be challenging for practical purposes.

Another limitation of the study is related to the analytical assessment of all available data on SHA, which consists of in-depth, comprehensive analysis and interpretation of all previously collected data stored in the mine information system MAC. The most important goals are the spatial positioning and contextual interpretation of all the mentioned data. Based on this, general and special requirements for collecting additional data are determined (in cases where the existing data are insufficient for the safe positioning of minefields or mine-explosive devices). The results are highly dependent on the expert military knowledge, skills, and affinities of the researchers of the analytical expert team. For this reason, it is of crucial importance to define well every data that enters the system and its links with all other objects in the system. In this way, the expert will be able to gain comprehensive insight into the situation on the battlefield and define the borders of the SHA more confidently, that is, identify what data he lacks in order to do so. The limitations of these procedures will largely depend on the ability to properly define each object and establish connections between them. The main challenge is to make all the data comparable so that they can all be used together.

Finally, in the context of applying deep learning for extracting text from minefield records, YOLO is a powerful object detection algorithm [69–71]; however, it has certain limitations when used for extracting text from documents such as minefield records [72]. YOLO may struggle with very small text, text in low-resolution images, and when there is a significant amount of noise in the image. If the minefield records contain too small or poorly defined text, YOLO might have difficulty accurately detecting and extracting it. YOLO is designed to detect objects with regular shapes and might face challenges when dealing with irregular text layouts or non-standard orientations. Minefield records may have text arranged in unconventional patterns, making it harder for YOLO to reliably

capture. In such cases, a pre-processing step is necessary to align images to an angle more suitable for extraction [73,74]. This will enable extracting text with regular shapes.

8. Conclusions and Future Work

Data observatories are multifunctional web-based platforms that provide a wide range of services and resources to support data-driven research and decision-making. However, the benefits of data observatories, data lakes, and other data engineering techniques that use advanced web technologies have not been employed in humanitarian demining.

This paper has successfully introduced the Minefield Observatory, an innovative web-based datastore service. It has effectively integrated a wide array of non-technical survey and humanitarian demining data and provided a comprehensive and formal representation of minefields through the MINEONT+ ontology. This approach greatly simplifies the process of extracting relevant information from different sensor datasets, thereby increasing the efficiency of demining efforts. The integration of diversely structured remote sensing datasets and the innovative use of the Microdata API for seamless user interaction show the robustness and utility of the proposed observatory concept.

The expected main outcome of the Minefield Observatory system is a toolbox for storage, special requirements for collecting data on SHA; assessment and analysis of all available data; extraction and delineation of mine presence indicators; producing mine hazard maps. All functions of the system could be available on one integrated workstation. The secondary outcome will be a functional multisensory UAV system (multispectral and thermal) for data collection from the depth of SHA with Standard Operational Procedures (SOP) to prepare and implement the UAV data collection flight mission.

Future developments will focus on extending the capabilities of the proposed observatory, potentially integrating it with other software applications, and utilizing it in advanced artificial intelligence systems to further improve its benefits in the domain of humanitarian demining.

References

1. *IMAS 04.10*; Glossary of Mine Action Terms, Definitions and Abbreviations, 2nd ed. UNMAS: New York, NY, USA, 2023. Available online: https://www.mineactionstandards.org/fileadmin/uploads/imas/Standards/English/IMAS_04.10_Ed.2_Am.11.pdf (accessed on 4 February 2024).
2. Maathuis, B.H.P. Remote Sensing Based Detection of Minefields. *Geocarto Int.* **2003**, *18*, 51–60. [CrossRef]
3. Bajić, M.; Matić, C.; Krtalić, A.; Candjar, Z.; Vuletic, D. *Research of the Mine Suspected Area*; HCR Centre for Testing, Development and Training Ltd.: Zagreb, Croatia, 2011; ISBN 978-953-99879-7-6. Available online: https://www.ctro.hr/publications (accessed on 4 February 2024).
4. Matić, Č.; Laura, D.; Turšić, R.; Krtalić, A. *Analytical Assessment for the Process of Collecting Additional Data on a Suspected Hazardous Area in Humanitarian Demining*; CROMAC-CTDT Ltd.: Zagreb, Croatia, 2014. Available online: https://www.ctro.hr/publications (accessed on 4 February 2024).
5. Geneva International Centre for Humanitarian Demining. A Guide to the International Mine Action Standards. 2006. Available online: https://www.files.ethz.ch/isn/26813/Guide_IMAS_2006.pdf (accessed on 4 February 2024).
6. *IMAS 08.10*; Non-Technical Survey, 1st ed. UNMAS: New York, NY, USA, 2019. Available online: https://www.mineactionstandards.org/fileadmin/uploads/imas/Standards/English/IMAS_08.10_Ed.1_Am.4.pdf (accessed on 4 February 2024).

7. *IMAS 08.20*; Technical Survey, 2nd ed. UNMAS: New York, NY, USA, 2019. Available online: https://www.mineactionstandards.org/fileadmin/uploads/imas/Standards/English/IMAS_08.20_Ed.1_Am.4.pdf (accessed on 4 February 2024).

8. Dorn, A.W. Eliminating Hidden Killers: How Can Technology Help Humanitarian Demining? *Stability Int. J. Secur. Dev.* **2019**, *8*, 1–17. [CrossRef]

9. Ibrahim, N.; Fahs, S.; AlZoubi, A. Land Cover Analysis Using Satellite Imagery for Humanitarian Mine Action and ERW Survey. In Proceedings of the Multimodal Image Exploitation and Learning 2021, Online, 12–16 April 2021; SPIE: Bellingham, WA, USA, 2021; Volume 11734, p. 1173402.

10. Bajić, M. Advanced Intelligence Decision Support System for the Assessment of Mine Suspected Areas. *J. ERW Mine Action* **2010**, *14*, 28.

11. Krtalić, A.; Racetin, I.; Gajski, D. The Indicators of Mine Presence and Absence in Airborne and Satellite Non-Technical Survey. In Proceedings of the 15th International Symposium "Mine Action 2018", Slano, Croatia, 9–12 April 2018; pp. 57–61.

12. Krtalić, A.; Bajić, M. Development of the TIRAMISU Advanced Intelligence Decision Support System. *Eur. J. Remote Sens.* **2019**, *52*, 40–55. [CrossRef]

13. Meurer, H.; Wehner, M.; Schillberg, S.; Hund-Rinke, K.; Kühn, C.; Raven, N.; Wirtz, T. An Emerging Remote Sensing Technology and Its Potential Impact on Mine Action. In Proceedings of the 7th International Symposium Humanitarian Demining, Sibenik, Croatia, 26–29 April 2010; p. 66.

14. Rosati, R.; Romeo, L.; Cecchini, G.; Tonetto, F.; Viti, P.; Mancini, A.; Frontoni, E. From Knowledge-Based to Big Data Analytic Model: A Novel IoT and Machine Learning Based Decision Support System for Predictive Maintenance in Industry 4.0. *J. Intell. Manuf.* **2023**, *34*, 107–121. [CrossRef]

15. Cho, S.; May, G.; Tourkogiorgis, I.; Perez, R.; Lazaro, O.; de La Maza, B.; Kiritsis, D. A Hybrid Machine Learning Approach for Predictive Maintenance in Smart Factories of the Future. In *Advances in Production Management Systems. Smart Manufacturing for Industry 4.0: IFIP WG 5.7 International Conference, APMS 2018*; Springer International Publishing: Cham, Switzerland, 2018; pp. 311–317.

16. De Luca, R.; Ferraro, A.; Galli, A.; Gallo, M.; Moscato, V.; Sperli, G. A Deep Attention Based Approach for Predictive Maintenance Applications in IoT Scenarios. *J. Manuf. Technol. Manag.* **2023**, *34*, 535–556. [CrossRef]

17. Çınar, Z.M.; Nuhu, A.A.; Zeeshan, Q.; Korhan, O.; Asmael, M.; Safaei, B. Machine Learning in Predictive Maintenance Towards Sustainable Smart Manufacturing in Industry 4.0. *Sustainability* **2020**, *12*, 8211. [CrossRef]

18. Özkula, S.M.; Reilly, P.J.; Hayes, J. Easy Data, Same Old Platforms? A Systematic Review of Digital Activism Methodologies. *Inf. Commun. Soc.* **2023**, *26*, 1470–1489. [CrossRef]

19. Li, G.; Hu, W.; You, T. Data Lake Development Status and Outlook. In Proceedings of the Third International Conference on Green Communication, Network, and Internet of Things (CNIoT 2023), Hainan, China, 17–21 December 2023; SPIE: Bellingham, WA, USA, 2023; Volume 12814, p. 128142G. [CrossRef]

20. Venugopal, V.E.; Srinivasa, S.; Ramanathan, C. Ontology Augmented Data Lake System for Policy Support. In *Big Data Analytics in Astronomy, Science, and Engineering, Proceedings of the 10th International Conference on Big Data Analytics, BDA 2022, Aizu, Japan, 5–7 December 2022*; Springer Nature: Cham, Switzerland, 2023; Volume 13830, p. 3.

21. Machado, I.A.; Costa, C.; Santos, M.Y. Data Mesh: Concepts and Principles of a Paradigm Shift in Data Architectures. *Procedia Comput. Sci.* **2022**, *196*, 263–271. [CrossRef]

22. Araújo Machado, I.; Costa, C.; Santos, M.Y. Advancing Data Architectures with Data Mesh Implementations. In *International Conference on Advanced Information Systems Engineering*; Springer International Publishing: Cham, Switzerland, 2022; pp. 10–18.

23. Errami, S.A.; Hajji, H.; El Kadi, K.A.; Badir, H. Spatial Big Data Architecture: From Data Warehouses and Data Lakes to the LakeHouse. *J. Parallel Distrib. Comput.* **2023**, *176*, 70–79. [CrossRef]

24. Zhang, R.; Indulska, M.; Sadiq, S. Discovering Data Quality Problems: The Case of Repurposed Data. *Bus. Inf. Syst. Eng.* **2019**, *61*, 575–593. [CrossRef]

25. Butte, V.K.; Butte, S. Enterprise Data Strategy: A Decentralized Data Mesh Approach. In Proceedings of the 2022 International Conference on Data Analytics for Business and Industry (ICDABI), Sakhir, Bahrain, 25–26 October 2022; IEEE: New York, NY, USA, 2022; pp. 62–66.

26. Khoshbakht, F.; Shiranzaei, A.; Quadri, S.M.K. Design & Develop: Data Warehouse & Data Mart for Business Organization. *Int. J. Intell. Syst. Appl. Eng.* **2023**, *11*, 260–265.

27. Zhang, H.; Ren, S.; Li, X.; Baharin, H.; Alghamdi, A.; Alghamdi, O.A. Developing Scalable Management Information System with Big Financial Data Using Data Mart and Mining Architecture. *Inf. Process. Manag.* **2023**, *60*, 103326. [CrossRef]

28. Kulkarni, A.; Chong, D.; Batarseh, F.A. Foundations of Data Imbalance and Solutions for a Data Democracy. In *Data Democracy*; Academic Press: London, UK, 2020; pp. 83–106.

29. Lefebvre, H.; Legner, C.; Fadler, M. Data Democratization: Toward a Deeper Understanding. In Proceedings of the International Conference on Information Systems (ICIS), Online, 11–12 December 2021.

30. Jarvenpaa, S.L.; Essén, A. Data Sustainability: Data Governance in Data Infrastructures Across Technological and Human Generations. *Inf. Organ.* **2023**, *33*, 100449. [CrossRef]

31. Hamed, N.; Rana, O.; Orozco Ter Wengel, P.; Goossens, B.; Perera, C. Forest Observatory: A Resource of Integrated Wildlife Data. 2022. Available online: https://orca.cardiff.ac.uk/id/eprint/153362/ (accessed on 15 February 2024).

32. Tiropanis, T.; Hall, W.; Hendler, J.; de Larrinaga, C. The Web Observatory: A Middle Layer for Broad Data. *Big Data* **2014**, *2*, 129–133. [CrossRef] [PubMed]

33. Horvat, M.; Krtalić, A.; Akagić, A.; Krmpotić, K.; Skender, S. Humanitarian Demining Using Data Observatories and Data Lakes. In Proceedings of the19th International Symposium "Mine Action 2023", Vodice, Croatia, 3–5 May 2023; pp. 47–51.

34. Khine, P.P.; Wang, Z.S. Data Lake: A New Ideology in Big Data Era. *ITM Web Conf.* **2018**, *17*, 03025. [CrossRef]

35. Miloslavskaya, N.; Tolstoy, A. Big Data, Fast Data and Data Lake Concepts. *Procedia Comput. Sci.* **2016**, *88*, 300–305. [CrossRef]

36. Giebler, C.; Gröger, C.; Hoos, E.; Schwarz, H.; Mitschang, B. Leveraging the Data Lake: Current State and Challenges. In Proceedings of the 21st International Conference DaWaK 2019, Linz, Austria, 26–19 August 2019; Springer International Publishing: Cham, Switzerland, 2019; pp. 179–188.

37. Nargesian, F.; Zhu, E.; Miller, R.J.; Pu, K.Q.; Arocena, P.C. Data Lake Management: Challenges and Opportunities. *Proc. VLDB Endow.* **2019**, *12*, 1986–1989. [CrossRef]

38. Sawadogo, P.; Darmont, J. On Data Lake Architectures and Metadata Management. *J. Intell. Inf. Syst.* **2021**, *56*, 97–120. [CrossRef]

39. Staab, S.; Studer, R. (Eds.) *Handbook on Ontologies*; Springer: Berlin, Germany, 2009.

40. Möller, R.; Neumann, B. Ontology-Based Reasoning Techniques for Multimedia Interpretation and Retrieval. In *Semantic Multimedia and Ontologies: Theory and Applications*; Springer: London, UK, 2008; pp. 55–98.

41. Patel, A.; Debnath, N.C. A Comprehensive Overview of Ontology: Fundamental and Research Directions. *Curr. Mater. Sci. Formerly Recent Patents Mater. Sci.* **2024**, *17*, 2–20. [CrossRef]

42. Borgo, S.; Galton, A.; Kutz, O. Foundational Ontologies in Action. *Appl. Ontol.* **2022**, *17*, 1–16. [CrossRef]

43. Biagetti, M.T. Ontologies as Knowledge Organization Systems. *KO Knowl. Organ.* **2021**, *48*, 152–176. [CrossRef]

44. Baader, F.; Nutt, W. Basic Description Logics. In *Description Logic Handbook*; Baader, F., Calvanese, D., McGuinness, D.L., Nardi, D., Patel-Schneider, P.F., Eds.; Cambridge University Press: Cambridge, UK, 2002; pp. 47–100.

45. Gruber, T.R. The Role of Common Ontology in Achieving Sharable, Reusable Knowledge Bases. *KR* **1991**, *91*, 601–602.

46. Gruber, T.R. Toward Principles for the Design of Ontologies Used for Knowledge Sharing? *Int. J. Hum.-Comput. Stud.* **1995**, *43*, 907–928. [CrossRef]

47. Schwade, F.; Schubert, P. A Semantic Data Lake for Harmonizing Data from Cross-Platform Digital Workspaces Using Ontology-Based Data Access. In Proceedings of the AMCIS 2020—A Vision for the Future, Salt Lake City, UT, USA, 10–14 August 2020; Volume 2.

48. Käfer, T.; Umbrich, J.; Hogan, A.; Polleres, A. Towards a Dynamic Linked Data Observatory. In Proceedings of the LDOW at WWW, Lyon, France, 16 April 2012.

49. Fricke, G. UXO Field Identification Database: A Tool for UXO. In *Environmental Security and Public Safety: Problems and Needs in Conversion Policy and Research after 15 Years of Conversion in Central and Eastern Europe*; Springer: Dordrecht, The Netherlands, 2007; p. 171.

50. Bajić, M.; Gold, H.; Horvat, M.; Krtalić, A.; Laura, D.; Muštra, M. The Novel Paradigm for a Decision Support System of the Aerial Non-Technical Survey. In Proceedings of the 17th International Symposium "Mine Action 2021", Novi Vinodolski, Croatia, 16–18 June 2021; pp. 62–68.

51. Horvat, M.; Krtalić, A.; Bajić, M.; Muštra, M.; Laura, D.; Gold, H. MINEONT: A Proposal for a Core Ontology in the Aerial Non-Technical Survey Domain. In Proceedings of the 18th International Symposium "Mine Action 2022", Novi Vinodolski, Croatia, 16–18 June 2022; pp. 47–52.

52. Horvat, M.; Krmpotić, K.; Krtalić, A.; Akagić, A. Bridging Blockchain Technology and Humanitarian Demining: A Novel Concept for Decentralized Storage of Landmine and UXO Locations. In *Central European Conference on Information and Intelligent Systems*; Faculty of Organization and Informatics: Varazdin, Croatia, 2023; pp. 369–375.

53. Chen, X.; Jia, S.; Xiang, Y. A Review: Knowledge Reasoning Over Knowledge Graph. *Expert Syst. Appl.* **2020**, *141*, 112948. [CrossRef]

54. Munir, K.; Anjum, M.S. The Use of Ontologies for Effective Knowledge Modelling and Information Retrieval. *Appl. Comput. Inform.* **2018**, *14*, 116–126. [CrossRef]

55. Hogan, A.; Hogan, A. SPARQL Query Language. In *The Web of Data*; Springer: Cham, Switzerland, 2020; pp. 323–448.

56. Vigo, M.; Matentzoglu, N.; Jay, C.; Stevens, R. Comparing Ontology Authoring Workflows with Protégé: In the Laboratory, in the Tutorial and in the 'Wild'. *J. Web Semant.* **2019**, *57*, 100473. [CrossRef]

57. Tudorache, T.; Noy, N.F.; Tu, S.; Musen, M.A. Supporting Collaborative Ontology Development in Protégé. In Proceedings of the 7th International Semantic Web Conference, ISWC 2008, Karlsruhe, Germany, 26–30 October 2008; Springer: Berlin/Heidelberg, Germany, 2008; pp. 17–32.

58. Bak, J.; Jedrzejek, C.; Falkowski, M. Usage of the Jess Engine, Rules and Ontology to Query a Relational Database. In Proceedings of the International Symposium, RuleML 2009, Las Vegas, NV, USA, 5–7 November 2009; Springer: Berlin/Heidelberg, Germany, 2009; pp. 216–230.

59. Al-Ajlan, A. The Comparison Between Forward and Backward Chaining. *Int. J. Mach. Learn. Comput.* **2015**, *5*, 106. [CrossRef]

60. Krtalić, A.; Matić, Č. Statistical Processing of Minefield Records. In Proceedings of the International Symposium "Humanitarian Demining 2010", Šibenik, Croatia, 27–29 April 2010; pp. 78–80.

61. Osmankovic, D.; Akagic, A.; Krivic, S.; Uzunovic, T.; Velagic, J. Towards Safe and Explainable Humanitarian Demining with Deep Learning. In Proceedings of the 18th International Symposium "Mine Action 2022", Novi Vinodolski, Croatia, 16–18 June 2022; pp. 52–56.

62. Habib, M.K. Humanitarian Demining: Reality and the Challenge of Technology—The State of the Arts. *Int. J. Adv. Robot. Syst.* **2007**, *4*, 19. [CrossRef]

63. Chan, J.W.; Alegria, A.C.; Veratelli, M.G.; Folegani, M.; Sahli, H. Combined Spatial Point Pattern Analysis and Remote Sensing for Assessing Landmine Affected Areas. In Proceedings of the 2012 IEEE International Geoscience and Remote Sensing Symposium, Munich, Germany, 22–27 July 2012; IEEE: New York, NY, USA, 2012; pp. 5368–5371.

64. Long, S.; He, X.; Yao, C. Scene Text Detection and Recognition: The Deep Learning Era. *Int. J. Comput. Vis.* **2021**, *129*, 161–184. [CrossRef]

65. Redmon, J.; Divvala, S.; Girshick, R.; Farhadi, A. You Only Look Once: Unified, Real-Time Object Detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 26 June–1 July 2016; pp. 779–788.

66. Cao, L.; Li, H.; Xie, R.; Zhu, J. A Text Detection Algorithm for Image of Student Exercises Based on CTPN and Enhanced YOLOv3. *IEEE Access* **2020**, *8*, 176924–176934. [CrossRef]

67. Smith, R. An Overview of the Tesseract OCR Engine. In Proceedings of the Ninth International Conference on Document Analysis and Recognition (ICDAR 2007), Curitiba, Brazil, 23–26 September 2007; IEEE: New York, NY, USA, 2007; Volume 2, pp. 629–633.

68. Clausner, C.; Antonacopoulos, A.; Pletschacher, S. Efficient and Effective OCR Engine Training. *Int. J. Doc. Anal. Recognit. (IJDAR)* **2020**, *23*, 73–88. [CrossRef]

69. Diwan, T.; Anirudh, G.; Tembhurne, J.V. Object Detection Using YOLO: Challenges, Architectural Successors, Datasets and Applications. *Multimed. Tools Appl.* **2023**, *82*, 9243–9275. [CrossRef] [PubMed]

70. Jiang, P.; Ergu, D.; Liu, F.; Cai, Y.; Ma, B. A Review of Yolo algorithm developments. *Procedia Comput. Sci.* **2022**, *199*, 1066–1073. [CrossRef]

71. Brdjanin, A.; Dardagan, N.; Dzigal, D.; Akagic, A. Single Object Trackers in OpenCV: A Benchmark. In Proceedings of the 2020 International Conference on INnovations in Intelligent SysTems and Applications (INISTA), Novi Sad, Serbia, 24–26 August 2020; IEEE: New York, NY, USA, 2020; pp. 1–6.

72. Sirisha, U.; Praveen, S.P.; Srinivasu, P.N.; Barsocchi, P.; Bhoi, A.K. Statistical Analysis of Design Aspects of Various YOLO-Based Deep Learning Models for Object Detection. *Int. J. Comput. Intell. Syst.* **2023**, *16*, 126. [CrossRef]

73. Wang, X.; Zheng, S.; Zhang, C.; Li, R.; Gui, L. R-YOLO: A Real-Time Text Detector for Natural Scenes with Arbitrary Rotation. *Sensors* **2021**, *21*, 888. [CrossRef]

74. Cheng, Z.; Xu, Y.; Bai, F.; Niu, Y.; Pu, S.; Zhou, S. AON: Towards Arbitrarily-Oriented Text Recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 5571–5579.