LLM-Driven Knowledge Graph Construction from Earth Observation Data for Extreme Events

Theodoros Aivalis[1,2], Iraklis A. Klampanos[1] and Antonis Troumpoukis[2]

[1]University of Glasgow, UK
[2]National Centre for Scientific Research "Demokritos", Greece

Abstract

The increasing frequency and severity of climate-related disasters call for more interpretable and actionable insights from Earth Observation (EO) data. In this work, we propose a novel framework that leverages multimodal Large Language Models (LLMs) to construct structured knowledge graphs (KGs) from heterogeneous disaster-related sources, including satellite imagery, textual reports, and geospatial metadata. By grounding these data streams in a domain-specific ontology, we produce semantically rich, human-aligned representations of extreme events so as to enable transparent reasoning and flexible querying across spatial, temporal, and socio-economic dimensions. We demonstrate the utility of our system through a detailed case study on flood events, supported by quantitative evaluations of the extracted triples and example KG-based queries. Our results show that this approach enables interpretable comparisons of disaster events, supports informed planning, and provides a reusable interface for downstream analysis in climate resilience and emergency response.

Keywords

Multimodal LLMs, KGs, Earth Observation, Satellite Imagery, Extreme Weather, Flood Events, Disaster Forecasting, Interpretability, Ontology-Guided Extraction, Semantic Querying

1. Introduction

In recent years, the scale and intensity of climate-related disasters have increased globally. From 1980 to 2024, the United States alone has experienced 403 weather and climate events, resulting in a total cost of over $2.9 trillion.[1] Notably, 2024 recorded 27 such billion-dollar disasters in the U.S., triple the historical annual average. On a global scale, the World Meteorological Organization documented more than 150 extreme weather events in 2024, including record-breaking floods, heatwaves, and hurricanes.[2] These disasters displaced over 800,000 people, the highest number since tracking began in 2008. The growing frequency and severity of such events underscore the need for more effective strategies to mitigate their impacts and support communities in better preparing for and responding to these challenges.

At the same time, the volume and variety of available data have expanded significantly. Earth Observation (EO) platforms, such as the European Union's Copernicus programme and Sentinel satellite missions, now produce vast amounts of high-resolution, multi-modal data daily.[3] This growth is supported by a rapid increase in satellite infrastructure. As of May 2025, approximately 11,700 active satellites orbit the Earth, with more than 2,800 launched in 2024 alone.[4] A significant portion of these are dedicated to EO, and forecasts suggest that even more EO satellites will be launched between 2024 and 2033, effectively tripling current capacity.[5] Nevertheless, the heterogeneity and scale of this data make it difficult to integrate and analyse effectively [1].[6] Unlocking its full potential remains a key challenge in disaster analysis and decision-making.

Workshop on AI-driven Data Engineering and Reusability for Earth and Space Sciences (DARES'25), co-located with the 28th European Conference on Artificial Intelligence (ECAI 2025), Bologna, Italy, October 25, 2025

teoaivalis@iit.demokritos.gr (T. Aivalis)
0009-0005-4452-9402 (T. Aivalis); 0000-0003-0478-4300 (I. A. Klampanos); 0000-0003-1078-8121 (A. Troumpoukis)

© 2025 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

[1]https://www.ncei.noaa.gov/access/billions, as viewed July 2025
[2]https://www.theguardian.com/environment/2025/mar/19/unprecedented-climate-disasters-extreme-weather-un-report, as viewed July 2025.
[3]https://dataspace.copernicus.eu/explore-data/data-collections/sentinel-data/sentinel-2, as viewed July 2025.
[4]https://www.livescience.com/how-many-satellites-orbit-earth, as viewed July 2025.
[5]https://news.satnews.com/2024/07/11/novaspace-earth-observation-satellites-set-to-triple-over-the-next-decade, as viewed July 2025.
[6]https://iaes.cgiar.org/spia/news/challenges-using-earth-observation-data-impact-evaluation, as viewed July 2025.

CEUR-WS.org/Vol-4128/paper5.pdf

CEUR Workshop Proceedings ceur-ws.org ISSN 1613-0073
published 2025-12-13

To address these issues, we propose a framework that integrates multi-source disaster data, specifically text descriptions from past events and satellite imagery, into a unified and structured representation. Recent advances in Large Language Models (LLMs), including their ability to process both textual and visual inputs, open new possibilities for extracting rich and interpretable information from heterogeneous sources. We extract key information from both modalities using multimodal LLMs. These outputs are then combined and encoded into a knowledge graph (KG), where each node represents aspects of the disaster such as location, impact, and timeline. Our goal is to compare a selected or current event with past ones by identifying structurally similar cases. This process enables decision-makers to draw on historical precedents to better understand the potential trajectory and consequences of new events. By representing disaster data in a structured and interpretable way, our approach supports more transparent and informed decision-making.

Contributions. The main contributions of this paper are summarised below:

- We introduce a framework that integrates multimodal disaster data into structured KGs grounded in a domain specific ontology.
- We leverage state-of-the-art multimodal LLMs to extract semantically aligned triples that describe disaster impacts, locations, and environmental context.
- We evaluate the semantic alignment and structure of the generated triples using cosine similarity and standard IR metrics, demonstrating that our method produces high-quality and interpretable representations across modalities.
- We showcase the practical utility of our KGs through structured queries on socio-economic, geographic, and environmental attributes, as well as event similarity retrieval, highlighting its potential for real-world disaster monitoring and response.

2. Related Work

2.1. Disaster Forecasting and Risk Assessment with EO data

EO plays a central role in monitoring and forecasting disasters, offering critical spatiotemporal insights for risk assessment, response, and mitigation. Central to this progress is the Copernicus Sentinel programme, operated by the European Space Agency, which delivers high-resolution, multi-sensor data for monitoring extreme weather events such as floods, droughts, wildfires, and storms. [7] The integration of radar, optical, and thermal imaging from Sentinel-1, -2, and -3 supports a range of applications, from early warning systems and damage assessment to long-term climate adaptation. Notably, the Copernicus Emergency Management Service (CEMS) demonstrates the practical use of EO data for rapid crisis mapping and emergency response.

While satellite imagery provides essential situational awareness, Merz et al. [2] point out that traditional early warning systems often fall short in anticipating the socio-economic impacts of disasters. They propose a more holistic approach that integrates hazard, exposure, and sensitivity data into predictive models to deliver actionable insights for planners and responders. In a complementary line of work, Giuliani et al. [3] advocate for a user-centric framework for disaster risk management (DRM), integrating EO data across the entire DRM cycle—from prevention and preparedness to response and recovery. Their review highlights how satellite-derived indicators of sensitivity (e.g., roof type, building density), exposure (e.g., land use, population), and hazard (e.g., sea-level rise, subsidence) can inform policy strategies. From a computational perspective, Mishra et al. [4] review how data mining techniques, including neural networks, decision trees, and text mining, have been applied to disaster detection and forecasting. They describe a two-phase architecture for an Indian disaster management system that fuses structured data (e.g., meteorological sensors) with unstructured data sources (e.g., social media and news feeds). Their findings highlight the growing importance of big data and real-time analytics in enhancing situational awareness and decision support.

[7] https://sentinels.copernicus.eu/web/success-stories/-/copernicus-sentinels-observe-earth-s-extreme-weather-events, as viewed July 2025

These studies illustrate a shift from hazard-centric approaches to impact-focused, human-aligned disaster forecasting. By integrating EO data with socio-economic indicators and machine learning techniques, researchers are opening new directions for more adaptive risk management frameworks. This evolution highlights the growing need for intelligent systems capable of integrating, interpreting, and communicating EO-derived knowledge, an area where recent advances in LLMs hold significant promise.

2.2. LLMs in EO Tasks

The integration of LLMs into EO has evolved rapidly, building upon a foundation of traditional AI approaches. Janga et al. [5] provide a comprehensive survey of classical machine learning techniques applied to EO tasks such as land cover mapping, change detection, object detection, and urban analysis. While these methods have led to significant advancements, they typically rely on modality-specific architectures and require extensive task-specific tuning. Persistent challenges, such as data availability, interpretability, and scalability have motivated the shift toward more unified systems.

To address these limitations, researchers have begun exploring multimodal LLMs as general-purpose interfaces for EO analysis. Early efforts include RemoteCLIP [6], which employs contrastive learning on remote sensing (RS) image–text pairs to enable zero-shot classification and retrieval. However, it lacks generative capabilities. RSGPT [7], which fine-tunes InstructBLIP for RS tasks, improves captioning and visual question answering (VQA) but underperforms on classification and visual grounding. GeoChat [8] introduces a region-aware, dialogue-based LLM built on the LLaVA framework. It supports interaction and spatial grounding but remains limited to optical imagery, restricting its generalisability. EarthGPT [9] aims to provide a unified interface for multimodal EO analysis by supporting a wide range of RS tasks, including scene classification, captioning, VQA, object detection, and visual grounding, across diverse sensor modalities (optical, SAR, and infrared). It combines a visual-enhanced perception module (fusing ViT and CNN features), a cross-modal mutual comprehension mechanism, and instruction tuning over the MMRS-1M dataset. These components enable EarthGPT to handle multi-sensor inputs and support dialogue-based interactions, addressing several of the limitations present in earlier models. Following this direction, GeoGPT[8], FrevaGPT [10] and DA4DTE [11] represent recent efforts to make geospatial and climate data analysis more accessible through conversational interfaces. Both systems enable users—regardless of technical background—to interact directly with EO datasets through natural language. Their deployment via web-based platforms and integration with tools like ChatGPT have contributed to their increasing adoption, supporting broader engagement with EO data and fostering interdisciplinary research.

While not based on LLMs, several systems enable semantic interaction with EO data via structured knowledge representations. WorldKG [12] structures OpenStreetMap data into a geographic KG linked to Wikidata and DBpedia. EarthQA [13] translates natural language queries into SPARQL using EO metadata and external knowledge bases like DBpedia. GeoQA2 [14] is a QA engine designed to answer geospatial questions—including those with quantities and aggregates—over the union of YAGO2 and YAGO2geo KGs. TerraQ [15] extends these efforts with a non-template-based QA engine over satellite image archives, offering rich spatiotemporal filtering and integration with a custom geospatial KG. These systems complement LLM-based models by enabling precise semantic querying and highlight the convergence of AI and knowledge-based reasoning in EO.

2.3. KG Generation Using LLMs

LLMs have become increasingly central to modern KG pipelines, enabling structured reasoning over unstructured or semi-structured data. Their contributions are in three core areas: (1) extracting `subject-predicate-object` triples from natural text, (2) assisting in ontology creation and extension, and (3) translating natural language queries into formal graph-based representations.

For triple extraction, prompt-based methods like AutoKG [16] demonstrate that LLMs can reliably

[8] https://geogpt.zero2x.org/, as viewed July 2025.

convert free text into ontology-aligned triples. Triplex[9] further refines this capability by introducing architectural optimisations for scalable and efficient KG construction. In addition, recent benchmarking studies [17] evaluate zero and few-shot KG generation, showcasing the robustness and flexibility of general-purpose LLMs across multiple domains. Beyond extraction, LLMs support ontology creation through techniques like prompt engineering, ontology reuse, and few-shot schema expansion [18]. Such systems facilitate domain-specific knowledge modelling by proposing consistent and reusable concept hierarchies. These tools are particularly valuable in emerging scientific domains. In semantic parsing and question answering, LLMs have shown promise in translating natural language into executable graph queries, such as SPARQL. Recent surveys [19, 20, 21] provide a systematic overview of this emerging field, highlighting the evolution from modular, rule-based NLP pipelines to unified, LLM-driven KG systems. These systems integrate tasks such as entity recognition, relation extraction, and graph population within a single generative or multitask framework.

These advances in LLM-driven KG construction offer a compelling foundation for interpretable and verifiable AI applications. In the context of EO, such KGs can be constructed directly from multimodal or metadata-rich datasets, providing structured representations that enable traceable reasoning. Instead of relying solely on dense embeddings or black-box models, we propose to leverage these interpretable KGs as intermediate, human-aligned objects for understanding and forecasting extreme events.

3. Methodology

Our methodology is grounded in a core principle: to make disaster-related data more explainable and actionable, we must first represent it in a structured and semantically rich format that supports human-aligned interpretation. We therefore propose a flexible framework for integrating heterogeneous data sources into a unified representation using ontologies and KGs. This approach enables us to compare events across multiple axes and supports more robust, interpretable decision-making.

3.1. From Multi-source Data to Unified Semantic Structures

Disaster data is inherently multimodal. For a single event, we might collect descriptive text from humanitarian agencies, visual signals from satellites, geographic features from open maps, and numerical metadata from sensors. Each modality offers a different view, but without integration, these insights remain hard to compare across events or regions. To address this, we employ an ontology-guided unification process. The main advantage of an ontology is that it can encode all the key concepts and relationships relevant to a specific task, in our case extreme events. Using this structured vocabulary, we can map raw inputs into consistent triples. The resulting KGs allow each disaster to be modelled as an interconnected entity whose attributes are logically organised and semantically comparable.

3.2. KG Construction from Multimodal Sources

We employ a pipeline that automatically constructs KGs from multi-source inputs. For each disaster event, we can collect and organise a diverse set of data types, including textual information (such as situation reports or narrative descriptions), satellite images and derived visual indicators, geospatial metadata capturing environmental or infrastructural context, and structured statistics that quantify the temporal dynamics and human or economic impacts of the event. These heterogeneous inputs provide complementary perspectives on the same phenomenon, which we unify through semantic alignment.

We then use a general-purpose multimodal LLM to extract triples from each input source. It is informed with the ontology, by giving the necessary relationships via the input prompt in order to extract structured triples from these sources. These triples describe the attributes of each event and are stored in a graph database. This enables efficient querying and visualisation of the graph structure. An overview of this end-to-end process is shown in Figure 1. Once constructed, these KGs serve as the basis for multiple tasks: identifying similar past events, computing descriptive statistics, generating interpretable summaries, and supporting response planning. The unified representation ensures that

[9] https://huggingface.co/SciPhi/Triplex, as viewed July 2025.

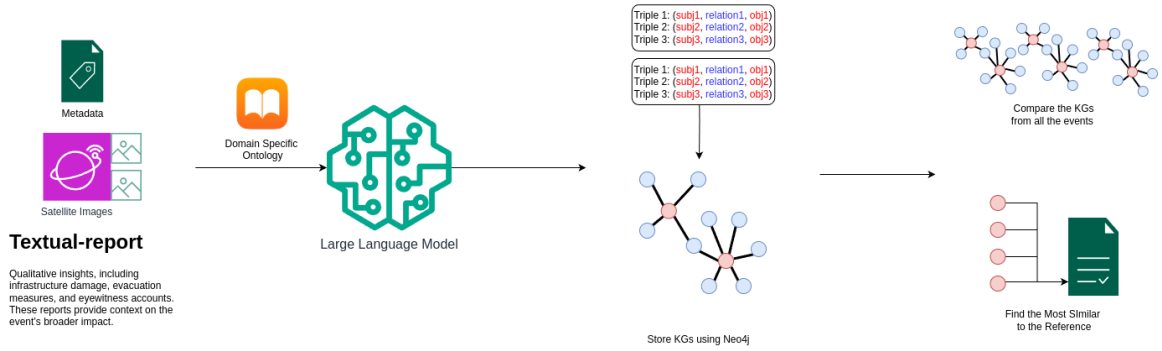


Figure 1: Overview of the end-to-end pipeline for transforming heterogeneous disaster-related data into structured KGs using domain-specific ontologies and multimodal LLMs. The resulting KGs enable interpretable analysis, event comparison, and decision support in disaster contexts.

both technical systems and human analysts can interpret, reason and act upon the collected data in a transparent way.

This representation supports interpretable and flexible comparisons. Analysts can assess event similarity globally or selectively, focusing on specific dimensions such as socio-economic impact, geographic setting, or environmental results. In contrast to opaque vector similarity, these comparisons are traceable and explainable, which is essential in high-risk applications such as disaster preparedness or policy planning.

4. Case Study: Structured Analysis of Flood Events

4.1. Data Collection and Enrichment for Flood Events

Our primary dataset is sourced from ReliefWeb [10], a leading humanitarian information service provided by the United Nations Office for the Coordination of Humanitarian Affairs (OCHA) [11]. Operated by the Digital Services Section of OCHA's Information Management Branch, ReliefWeb continuously monitors and curates content from over 4,000 sources, including humanitarian agencies, governments, research institutions, and media outlets. Its editorial team classifies and delivers high-quality, up-to-date information to support informed decision-making. Each disaster event entry on ReliefWeb includes structured metadata (e.g., disaster type, location, date), rich textual descriptions, situation reports, statistics on humanitarian impact, and supplementary content. These comprehensive entries provide critical context on the scale, effects, and emergency measures associated with each event. This makes ReliefWeb an invaluable source for constructing structured representations of real-world flood scenarios across diverse geographic regions.

To complement the disaster event data with geographic and environmental context, we incorporate information from OpenStreetMap (OSM)[12], a collaborative open-data project that provides freely accessible, high-resolution geographic information. Developed by a diverse global community of volunteers, including cartographers, GIS professionals, humanitarians, and local contributors, OSM emphasises local knowledge and accuracy. Contributors use aerial imagery, GPS traces, and field surveys to map features such as roads, rivers, forests, elevation peaks, and infrastructure across the world. In our study, we utilise OSM to extract natural features of the affected regions, such as water bodies, wooded areas, and topographical landmarks, based on the coordinates of each event. This geo-enrichment supports spatial reasoning and risk assessment by providing fine-grained insight into the environmental characteristics of the impacted locations.

In addition to textual and geographic data, we collect satellite images related to flood events from the

[10] https://reliefweb.int/disasters?search, as viewed July 2025
[11] https://www.unocha.org/?_gl=1*9vzczb*_ga*MTc1OTIzODM2NC4xNzQzMTAyNjYw*_ga_E60ZNX2F68*czE3NDkwNDU5MzAkbzE0JGcxJHQxNzQ5MDQ1OTQ4JGo0MiRsMCRoMA.., as viewed July 2025
[12] https://www.openstreetmap.org, as viewed July 2025

Table 1
Multimodal data for the Sri Lanka flood event (December 2019), combining metadata, geographic features, and satellite imagery for KG construction.

ReliefWeb	
Field (Type – Name)	Content
string – Name	Sri Lanka: Floods and Landslides - Dec 2019
date – Date	1 Dec 2019
string – Country	Sri Lanka
float – Latitude	7.61
float – Longitude	80.7
string – Disaster Type	Flood
string – GLIDE ID	FL-2019-000171-LKA
text – Description	According to media, as of 2 December, at least three people died following a landslide in Walapane Village. At least 570 people have been displaced in Batticaloa District and more than 4,100 people have been affected in Eastern and Northern Provinces [...]
OpenStreetMap	
Field (Type – Name)	Content
list – Natural Features	Peaks: 2 total; Woods: 2 total; Waters: 1 total
Microsoft Planetary Computer	
Field (Type – Name)	Content
image – RGB (visual.png)	True-colour composite. Date: 11–31 Dec
image – NIR (nir.png)	Near-Infrared. Date: 11–31 Dec
image – NDWI (ndwi.png)	Normalized Difference Water Index. Date: 11–31 Dec

Microsoft Planetary Computer[13], focusing primarily on events occurring after 2015. For each event, we download three types of visual products: (i) *visual.png*, a true-colour RGB composite (Bands B04, B03, B02), which approximates human vision and is suitable for general inspection and identifying major flooding or cloud cover; (ii) *nir.png*, a Near Infrared (NIR) reflectance image (Band B08), where bright areas indicate healthy vegetation and dark regions indicate water or built surfaces—useful for detecting flood impact on vegetation; and (iii) *ndwi.png*, a Normalized Difference Water Index (NDWI) image computed as $(B03 - B08)/(B03 + B08)$, where bright pixels correspond to water bodies or saturated land. For each of the three image types, we retrieve imagery for three time windows relative to the reported disaster date: *before* (30 to 10 days prior), *during* (5 days before to 10 days after), and *after* (10 to 30 days following the event). These durations capture both short and mid-term flood dynamics, supporting analysis of immediate impact and post-disaster recovery. In the current KG construction, we focus on the *after*-window, as post-event images are more likely to be cloud-free and reveal clearer evidence of flooding impact. In future work, we plan to use all time windows to model temporal changes in the affected regions. Due to cloud cover and limited satellite availability, usable imagery was not collected for every event. We therefore began with a few hundred well-covered cases and will expand the database in the next steps.

Table 1 presents a representative example combining both sources for a specific flood event in Sri Lanka. This includes a brief geo-summary extracted from OSM and an excerpt from the detailed disaster report on ReliefWeb.

4.2. Ontology Design, Structuring, and Multimodal Integration

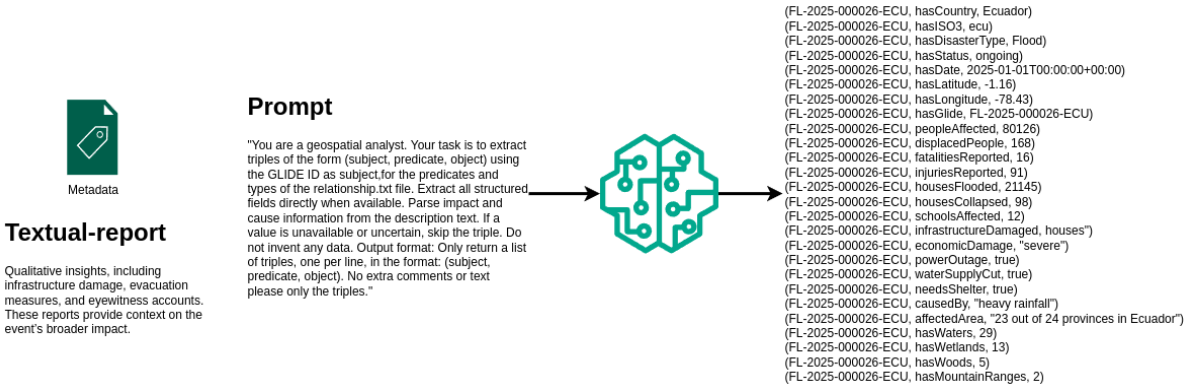
4.2.1. Ontology Design and Schema Construction

To extract structured information from multimodal sources, we design tailored prompts for both textual and image-based inputs. These prompts guide the LLM to produce triples aligned with a controlled

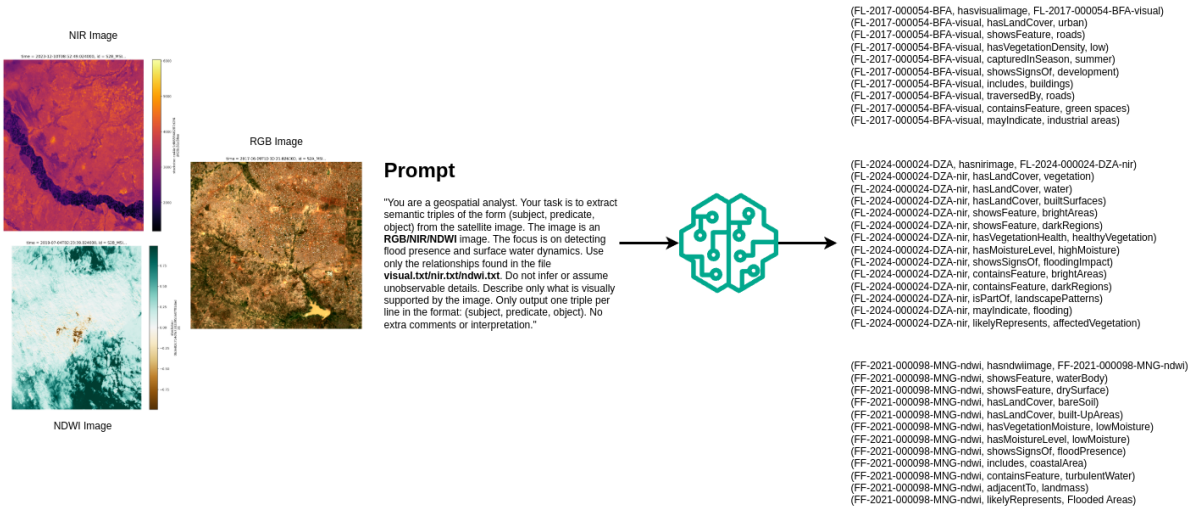
[13] https://planetarycomputer.microsoft.com/, as viewed July 2025

vocabulary of relationships, reflecting the observable attributes of each disaster. Examples of the prompts used for both modalities are shown in Figure 2.

To integrate and reason over the heterogeneous information described above we require a structured representation of each flood event. As outlined in Section 3, we adopt an ontology-based approach to unify the various data modalities into a stable schema. The ontology captures key characteristics of each event, supporting downstream tasks such as querying, analysis, and comparison over the events. Our design is based on established ontologies such as YAGO2geo [14] and GeoSPARQL[15]. However, the structure of our final ontology was also directly influenced by the empirical organisation of our dataset. We intentionally designed the schema to reflect the actual content and granularity of the collected data, ensuring that all relevant dimensions, ranging from humanitarian impact to environmental features and RS outputs, are fully captured and semantically linked. Figure 3 summarises the core ontology relationships used to describe each disaster event in a unified format.



(a) Prompt for triple extraction from text modality.



(b) Prompt for RGB, NDWI, and NIR image triple extraction.

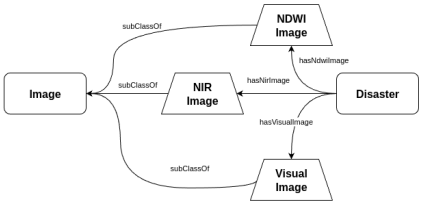
Figure 2: LLM prompting strategy for extracting triples from different EO data modalities.

4.2.2. Triple Quality Evaluation Across Modalities

To evaluate the quality of the multimodal data extracted by our system, we performed a two-part evaluation covering both the textual and visual modalities. Starting with the text, we employed the CLIP encoder [22] to extract text embeddings. Specifically, we used the encoder to process both the natural language descriptions sourced from ReliefWeb and OSM, as well as the triples generated by the LLM.

[14] https://yago2geo.di.uoa.gr/, as viewed July 2025
[15] https://opengeospatial.github.io/ogc-geosparql/geosparql11/, as viewed July 2025

Image Properties	
hasLandCover	includes
showsFeature	traversedBy
hasVegetationDensity	containsFeature
hasVegetationHealth	adjacentTo
hasVegetationMoisture	isPartOf
hasMoistureLevel	mayIndicate
capturedInSeason	likelyRepresents
showsSignsOf	experiencing



Disaster Properties				
hasCountry	hasGlide	schoolsAffected	agriculturalLandAffected	hasWetlands
hasISO3	peopleAffected	businessesDamaged	needsShelter	hasWoods
hasDisasterType	displacedPeople	roadsDamaged	humanitarianNeeds	hasScrubs
hasStatus	fatalitiesReported	infrastructureDamaged	causedBy	hasGrasslands
hasDate	injuriesReported	economicDamage	weatherCause	hasSprings
hasLatitude	housesFlooded	powerOutage	affectedArea	hasTrees
hasLongitude	housesCollapsed	waterSupplyCut	hasWaters	hasMountainRanges

Figure 3: Ontology properties used to describe multi-source disaster event data.

We then computed the cosine similarity between the embedding of the input text and the embedding of its corresponding set of triples. Additionally, we computed similarities between the input text and all other non-corresponding triples in the dataset, capturing statistics such as the mean, minimum, maximum, and standard deviation. As shown in Table 2 (left), the average cosine similarity for matching pairs is substantially higher than the mean non-matching similarity, indicating that the generated triples effectively capture the semantic content of the original descriptions. Although the non-matching similarities are lower, they remain relatively high, with low standard deviation, suggesting consistent structure across events. This can be attributed to the fact that all events are of the same type (floods), often involve similar consequences (e.g., displacement, infrastructure damage), and occur in overlapping regions. The uniformity in description style across the dataset also contributes to higher similarities, even among unrelated cases.

Table 2
Evaluation metrics for the generated triples: text-based extraction (left) and three image-based extractions (right).

Text-based Triples

Metric	Value
Matching	**0.8915**
Mean non-matching	0.8285
Min non-matching	0.7759
Max non-matching	0.8821
Std. non-matching	0.0153

Image-based Triples

Metric	Visual	NDWI	NIR
Precision	0.5491	0.5962	0.7832
Recall	0.5602	0.6446	0.8303
F1 score	0.5537	0.6091	0.8018
True Positive	5.4	9.3	11.3
False Positive	4.4	5.2	3.3
False Negative	4.3	4.4	2.5
Ground Truth Triples	9.78	14.6	13.8
Predicted Triples	9.84	14.5	14.6

As for the image modality, we manually created a set of ground truth triples for each image, based on expected content, and evaluated the predicted triples using Precision, Recall, and F1 score. This assessed both semantic similarity and structural correctness of the extracted KGs. The evaluation covered RGB, NDWI, and NIR image types. The results in Table 2 (right) show varied performance. NIR-based triples achieved the highest scores (F1: 0.8018), followed by NDWI (F1: 0.6091), and Visual (F1: 0.5537), indicating that in this phase RGB images provide fewer cues for structured extraction. True positives increased from Visual (5.4) to NDWI (9.3) to NIR (11.3), with a consistent number of predicted triples across modalities.

Despite lower performance in Visual and NDWI, all modalities offer complementary information. The NIR images include a colour bar that guides interpretation, whereas Visual images present raw satellite views with limited contextual support. NDWI highlights water presence but lacks clarity without domain knowledge. A potential improvement is using LLMs fine-tuned for EO tasks, such as those in Section 2.2, to better understand such imagery.

4.3. Demonstration Scenarios: Querying the Knowledge Graph

We stored all the triples in Neo4j[16], a widely adopted database that supports efficient storage, querying, and visualisation of the KGs. To demonstrate the practical utility of our system, we provide representative queries over the constructed KG. These queries illustrate how decision-makers, researchers, or humanitarian organisations can interact with the structured knowledge to extract actionable insights.

4.3.1. Queries on Human, Economic, and Temporal Dimensions of Disasters

To illustrate the versatility and analytical power of our KGs, we present several example queries addressing core dimensions of disaster impact, based on structured information extracted from textual reports. This form of retrieval is useful for identifying disasters with similar socio-economic consequences, which may inform planning or response efforts for current events.

Top 10 floods (peopleAffected)

Region	Date	PeopleAffected
China	1989-07-01T...	100,000,000
China	1994-06-01T...	73,360,000
China	1999-06-20T...	60,000,000
China	2010-05-06T...	29,000,000
China	1990-06-01T...	26,000,000
Bangladesh	1998-07-17T...	24,000,000
India	2018-11-16T...	23,000,000
Pakistan	2010-07-22T...	20,000,000
India	2015-06-09T...	13,700,000

Top 10 floods (fatalities)

Region	Date	Fatalities
Libya	2023-09-10T...	3,922
Philippines	1991-11-05T...	2,155
India	1994-06-24T...	2,001
Sudan	2011-08-05T...	2,000
China	1989-07-01T...	2,000
Pakistan	2010-07-22T...	1,985
China	2012-07-30T...	1,250
China	2011-09-25T...	1,250
China	2004-02-20T...	1,250

Top 10 floods (BusinessesDamaged)

Region	Date	BusinessesDamaged
China	2013-04-30T00:0...	1,400,000
Bangladesh	2012-06-27T00:0...	230,000
Niger	2017-06-16T00:0...	16,000
Niger	2017-06-16T00:0...	9,800
Bangladesh	1998-07-17T00:0...	6,500
Bangladesh	2023-08-07T00:0...	1,430
China	1997-05-01T00:0...	1,000
Iran (Islamic Repu...	2020-01-11T00:0...	719
Viet Nam	2017-06-05T00:0...	425

Top 10 floods (AgriculturalLandAffected)

Region	Date	AgricultureAffected
Syrian Arab Republic	2019-03-31T00:00:...	86,000,000
China	1989-07-01T00:00:...	8,500,000
India	1998-08-20T00:00:...	7,000,000
China	1994-06-01T00:00:...	4,500,000
Pakistan	2011-07-30T00:00:...	2,100,000
Pakistan	2010-07-22T00:00:...	2,000,000
India	2010-12-06T00:00:...	1,500,000
Pakistan	2013-08-04T00:00:...	1,500,000
Thailand	1995-10-01T00:00:...	1,336,000

Figure 4: Top 10 flood disasters ranked by (from left to right) people affected, fatalities, businesses damaged, and agricultural land affected.

We begin by ranking disasters by the number of people affected, targeting displaced or injured ones. This highlights large-scale humanitarian events and supports emergency response prioritisation. Next, we identify the most lethal disasters by retrieving events with high reported fatalities. This ranking helps characterise high-risk scenarios and quantify human loss. We then examine economic disruption by comparing the number of businesses destroyed. This serves as an indicator of urban resilience and informs recovery planning. Another query assesses the extent of agricultural land affected, helping to identify events with potential impacts on food systems and rural livelihoods. The results of these queries are also visualised in Figure 4 Finally, we explore temporal trends by filtering disasters that occurred in the 21st century. We group them by decade—2000s, 2010s, and 2020s—to analyse changes in disaster frequency over time. This supports the investigation of trends linked to climate factors, vulnerability shifts, or improved reporting. The results of this categorisation are summarised in Table 3.

Table 3
Number of recorded disasters by century and by decade for the 21st century.

Century	Disasters	Decade (21st)	Disasters
20th Century (before 2000)	305	2000s (2000–2009)	513
21st Century (2000 and after)	1293	2010s (2010–2019)	508
		2020s (2020–2029)	272

4.3.2. Geospatial and Environmental Comparison

One of the key strengths of our KG framework is its support for rich, context-aware querying that goes beyond traditional numerical analysis. By incorporating geographic coordinates, regional names, and environmental attributes, our KG enables users to explore disaster events through both semantic and spatial perspectives.

As a first use case, we retrieve all disaster events that occurred in a specific country—Afghanistan. This type of region-based filtering allows localised analysis and supports decision-making for national

[16] https://neo4j.com/, as viewed July 2025.

agencies. By focusing on a country, analysts can identify recurring patterns, assess preparedness levels, and compare the frequency and impact of disasters over time. To further exploit spatial data, we leverage geographic coordinates (latitude and longitude) associated with each disaster event. This enables bounding-box queries, where users can specify a rectangular geographic region, such as an area covering Europe and retrieve all disasters that occurred within its spatial extent. Such functionality is particularly valuable for regional comparison.

In a third experiment, we demonstrate semantic similarity search using environmental features extracted from OSM, such as rivers, forests, wetlands, and mountain ranges. Given a reference disaster, specifically a flood in Thailand (FL-2021-000147-THA), we query the KG to find other disasters that occurred in regions with similar natural characteristics. This facilitates retrieval of events influenced by comparable terrain and ecosystem structures, enabling responders and analysts to anticipate risk factors and intervention challenges tied to specific environmental contexts.

These querying capabilities illustrate the expressiveness and flexibility of our KGs. By integrating geospatial and semantic filters into the event retrieval pipeline, we support a wide range of analytical tasks, from focused regional exploration to generalisable scenario identification, grounded in both structured knowledge and physical geography.

4.3.3. Image-Based Flood Event Comparison

In our KG-database, each flood event is connected to its corresponding satellite image representations, specifically visual, NIR, and NDWI images, as illustrated on the left side of Figure 5. This design allows us to capture complementary perspectives of the same event: visual images provide general scene context, NIR images highlight vegetation health and moisture stress, and NDWI images emphasize water bodies and potential inundation areas. By explicitly linking each event to its diverse image modalities, we enable richer semantic descriptions and enhance the system's ability to interpret and compare events based on multiple features.

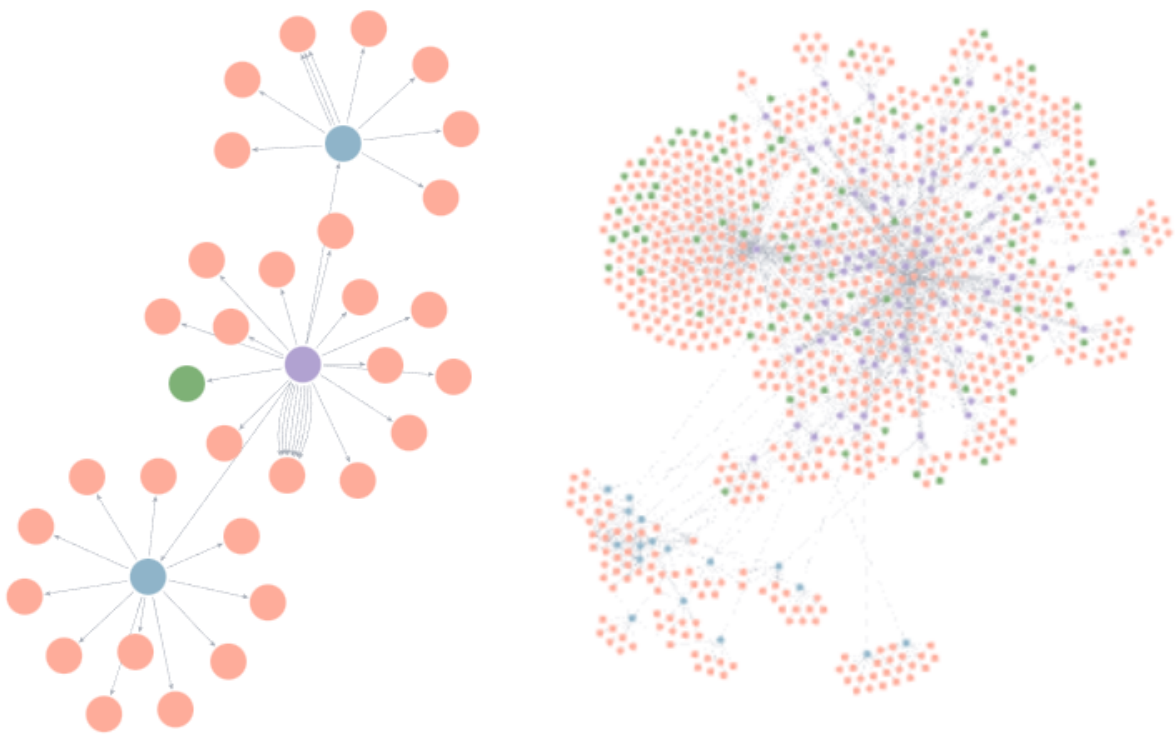


Figure 5: Graph view of a single flood event (left) and the full dataset (right). Purple node indicates the event, blue represent associated images, the green shows the country, and orange denote extracted attributes.

For each image, we retrieve the top similar events by comparing their properties, such as land cover,

vegetation and signs of flooding. These queries reveal comparable past events and also make the similarity interpretable, as users can examine which attributes contribute to each match. This capability supports various applications, including event monitoring, rapid flood assessment, and understanding patterns across different flood scenarios by leveraging multi-modal image-derived knowledge.

The entire KG construction pipeline, including data preprocessing, triple extraction scripts, and all query examples used and referred above, is publicly available in our GitHub repository. [17]

5. Conclusions and Future Work

In this paper, we present a novel framework that constructs structured, ontology-guided KGs from multimodal disaster-related data. By combining textual reports, satellite imagery, and geospatial metadata with multimodal LLMs, we enable transparent event representations and interpretable similarity comparisons. Our system supports a range of applications, from early warning systems to post-disaster analysis, and demonstrates how KGs can serve as an effective medium for aligning machine reasoning with human understanding.

Looking ahead, we plan to expand our pipeline in several directions. First, we aim to integrate additional data sources, such as social media content, sensor networks, and governmental databases, to enrich the information represented in the graph and improve its contextual depth. Second, we will explore the use of multimodal LLMs finetuned specifically on EO tasks, with the goal of improving the accuracy and relevance of triples extracted from satellite imagery. Finally, we intend to engage with domain experts and end-users through the development of interactive tools, enabling qualitative evaluation of our system's usefulness and impact in real-world disaster management scenarios. Through these efforts, we hope to further enhance the interpretability, robustness, and practical value of structured AI in Earth and space science domains.

Acknowledgment

This work has received funding from the European Union's Digital Europe Programme (DIGITAL) under grant agreement No 101146490. This work was supported by the Intelligent Information Systems division of the Institute of Informatics and Telecommunications at National Scientific Research Centre "Demokritos".

Declaration on Generative AI

During the preparation of this work, the authors used ChatGPT and Grammarly to: Grammar and spelling check. After using these tools, the authors reviewed and edited the content as needed and take full responsibility for the publication's content.

References

[1] S. Li, S. Dragicevic, F. A. Castro, M. Sester, S. Winter, A. Coltekin, C. Pettit, B. Jiang, J. Haworth, A. Stein, T. Cheng, Geospatial big data handling theory and methods: A review and research challenges, ISPRS Journal of Photogrammetry and Remote Sensing 115 (2016) 119–133. doi:https://doi.org/10.1016/j.isprsjprs.2015.10.012, theme issue 'State-of-the-art in photogrammetry, remote sensing and spatial information science'.

[2] B. Merz, C. Kuhlicke, M. Kunz, M. Pittore, A. Babeyko, D. Bresch, D. I. V. Domeisen, S. Fuchs, M. Garschagen, K. Goda, M. Kohler, N. Komendantova, E. Lorenz, R. Ludwig, J. M. Monteiro, A. Oliver-Smith, T. Plattner, M. Pelling, C. Riggelsen, J. Schanze, K. Schröter, A. H. Thieken, Impact forecasting to support emergency management of natural hazards, Reviews of Geophysics 58 (2020) e2020RG000704. URL: https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2020RG000704.

[3] G. Giuliani, B. Chatenoux, A. De Bono, D. Rodila, J.-P. Richard, B. Allenbach, H. Dao, N. Ray, Building an earth observations data cube: lessons learned and technical progress, International Journal of Digital Earth 13 (2020) 70–92. doi:10.1080/17538947.2019.1655148.

[4] D. Mishra, A. Kalra, D. Sharma, Data mining techniques for decision support system in disaster management: A review, International Journal of Computer Applications 133 (2016) 7–12. doi:10.5120/ijca2016908016.

[17] https://github.com/teoaivalis/XtremeKG

[5] B. Janga, G. Asamani, Z. Sun, N. Cristea, A review of practical ai for remote sensing in earth sciences, Remote Sensing 15 (2023) 4112. doi:10.3390/rs15164112.

[6] F. Liu, D. Chen, Z. Guan, X. Zhou, J. Zhu, Q. Ye, L. Fu, J. Zhou, Remoteclip: A vision language foundation model for remote sensing, 2024. URL: https://arxiv.org/abs/2306.11029. arXiv:2306.11029.

[7] Y. Hu, J. Yuan, C. Wen, X. Lu, X. Li, Rsgpt: A remote sensing vision language model and benchmark, 2023. URL: https://arxiv.org/abs/2307.15266. arXiv:2307.15266.

[8] K. Kuckreja, M. S. Danish, M. Naseer, A. Das, S. Khan, F. S. Khan, Geochat: Grounded large vision-language model for remote sensing, 2023. URL: https://arxiv.org/abs/2311.15826. arXiv:2311.15826.

[9] W. Zhang, M. Cai, T. Zhang, Y. Zhuang, X. Mao, Earthgpt: A universal multi-modal large language model for multi-sensor image comprehension in remote sensing domain, 2024. URL: https://arxiv.org/abs/2401.16822. arXiv:2401.16822.

[10] C. Kadow, J. Saynisch-Wagner, S. Willmann, S. Lentz, J. Baehr, K. Sieck, F. Oertel, B. Wentzel, T. Ludwig, M. Bergemann, FrevaGPT: A Large Language Model-Driven Scientific Assistant for Climate Research and Data Analysis, in: EGU General Assembly 2025, Vienna, Austria, 2025, pp. EGU25–15507. URL: https://doi.org/10.5194/egusphere-egu25-15507.

[11] M. Corsi, G. Pasquali, C. Pratola, S. Tilia, S. Kefalidis, K. Plas, M. Pollali, E. Tsalapati, M. Tsokanaridou, M. Koubarakis, K. N. Clasen, L. Hackel, J. Hackstein, G. Sumbul, B. Demir, N. Longépé, DA4DTE: Developing a Digital Assistant for Satellite Data Archives, in: Proceedings of the Big Data from Space (BiDS), European Space Agency, 2023.

[12] A. Dsouza, N. Tempelmeier, R. Yu, S. Gottschalk, E. Demidova, Worldkg: A world-scale geographic knowledge graph, in: Proceedings of the 30th ACM International Conference on Information & Knowledge Management, CIKM '21, Association for Computing Machinery, New York, NY, USA, 2021, p. 4475–4484. URL: https://doi.org/10.1145/3459637.3482023.

[13] D. Punjani, M. Koubarakis, E. Tsalapati, Earthqa: A question answering engine for earth observation data archives, in: Geographic Information Retrieval Workshop, collocated with ACM SIGSPATIAL, 2023. URL: https://ai4copernicus-project.eu/wp-content/uploads/2023/12/EarthQA.pdf.

[14] S.-A. Kefalidis, D. Punjani, E. Tsalapati, K. Plas, M.-A. Pollali, P. Maret, M. Koubarakis, The question answering system geoqa2 and a new benchmark for its evaluation, International Journal of Applied Earth Observation and Geoinformation 134 (2024) 104203. URL: https://www.sciencedirect.com/science/article/pii/S1569843224005594.

[15] S.-A. Kefalidis, K. Plas, M. Koubarakis, Terraq: Spatiotemporal question-answering on satellite image archives, 2025. URL: https://arxiv.org/abs/2502.04415. arXiv:2502.04415.

[16] Y. Zhu, X. Wang, J. Chen, S. Qiao, Y. Ou, Y. Yao, S. Deng, H. Chen, N. Zhang, Llms for knowledge graph construction and reasoning: Recent capabilities and future opportunities, 2024. URL: https://arxiv.org/abs/2305.13168.

[17] A. Papaluca, D. Krefl, S. M. Rodriguez, A. Lensky, H. Suominen, Zero- and few-shots knowledge graph triplet extraction with large language models, 2023. URL: https://arxiv.org/abs/2312.01954.

[18] N. Fathallah, S. Staab, A. Algergawy, Llms4life: Large language models for ontology learning in life sciences, 2024. URL: https://arxiv.org/abs/2412.02035. arXiv:2412.02035.

[19] L. Zhong, J. Wu, Q. Li, H. Peng, X. Wu, A comprehensive survey on automatic knowledge graph construction, 2023. URL: https://arxiv.org/abs/2302.05019. arXiv:2302.05019.

[20] X. Zhu, Z. Li, X. Wang, X. Jiang, P. Sun, X. Wang, Y. Xiao, N. J. Yuan, Multi-modal knowledge graph construction and application: A survey, IEEE Transactions on Knowledge and Data Engineering 36 (2024) 715–735. URL: http://dx.doi.org/10.1109/TKDE.2022.3224228.

[21] H. Ye, N. Zhang, H. Chen, H. Chen, Generative knowledge graph construction: A review, 2023. URL: https://arxiv.org/abs/2210.12714. arXiv:2210.12714.

[22] A. Radford, J. W. Kim, C. Hallacy, A. Ramesh, G. Goh, S. Agarwal, G. Sastry, A. Askell, P. Mishkin, J. Clark, G. Krueger, I. Sutskever, Learning transferable visual models from natural language supervision, 2021. URL: https://arxiv.org/abs/2103.00020.