Using Integrated and Enriched Linked Data for Ukraine Resilience

Manar Attar[1][0009−0000−0163−2534], Shuai Wang[1][0000−0002−1261−9930], Ronald Siebes[1][0000−0001−8772−7904], and Eirik Kultorp[2][0009−0005−5211−045X]

[1]Vrije Universiteit Amsterdam, Amsterdam, the Netherlands
m.m.attar@student.vu.nl, {shuai.wang | r.m.siebes} @vu.nl
[2]Amstelveen, the Netherlands
ekultorp@gmail.com

1 Introduction

The mission of resilience of Ukrainian cities calls for international collaboration with the scientific community to increase the quality of information by identifying and integrating information from various sources of news and social media. In response to the conflict, the Centre for Information Resilience launched the Eyes on Russia (EoR) [1] project in January 2022, aiming at gathering and verifying media content related to Russia's invasion of Ukraine. Another noteworthy project is the Civilian Harm in Ukraine TimeMap (CH) [2], which provides a descriptive record of incidents, including source links, precise location data, and descriptions based on visual evidence. In this paper, we demonstrate how linked data technology can be used to unify, enrich, and integrate data from multiple relevant resources. We present use cases using the resulting data[1].

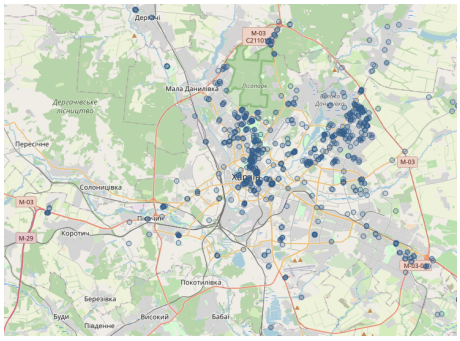
2 Data Processing

We convert both datasets to linked data and enrich them with additional geospatial information. Firstly, we address the issue of varying formats and fields in reported events by adopting entities and relations from well-established ontologies like schema.org [3], Dublin Core [4], Simple Event Ontology [5], and GeoNames [6]. This ensures a consistent and unique representation of geographic information. Following that, we present an algorithm for the detection of identical events from different datasets. Our pipeline makes it easy to convert and enrich datasets to integrated linked data [7]. Finally, we demonstrate in use cases how our dataset can be applied to different scenarios for resilience purposes. Finally, as for data enrichment, we use GeoNames' APIs to retrieve missing information, such as postal codes, for events. To handle multilingual cases and spelling errors, manual intervention is used for resolving difficulties in data enrichment. The process results in a more complete and standardized dataset that consists of 10K reported events covering damage to hospitals, schools, roads, residential buildings, etc., with enhanced overall quality for analysis and further research. This approach can be easily adapted to other data sources such as [8] and [9].

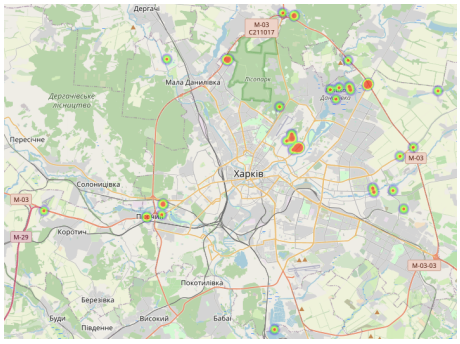
[1] The code and the original thesis are at https://github.com/LinkedData4Resilience/linked-data. The integrated dataset is available upon request while the other datasets are at https://triplydb.com/linked4resilience/-/datasets.

3 Use Cases

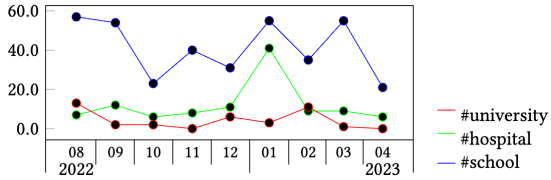
Use Case 1 and 2: Events visualization and shelter location suggestion As a demonstration[2] of the use of our integrated dataset, Figure 1a presents the result of a SPARQL query that retrieves events in Kharkiv in the integrated datasets. We retrieved shelter data in the city of Kharkiv [10], and measured the distance of events in Kharkiv to the nearest shelter. Figure 1b is a heatmap that shows the location of damaging events where there is no shelter within 1km distance. Thus, we suggest that shelters could be built to cover these areas.



(a) A representation of the damaging events in the city of Kharkiv using YASGUI



(b) A heatmap regarding the location of attacks in Kharkiv without any shelter within 1km



(c) Timelapse of events about public facilities

(d) Multilingual

Fig. 1: Usecases

Use Case 3: Timelapse of damaging events about public facilities Figure 1c illustrates dates and their corresponding number of events about schools, universities, and hospitals between 1st August 2022 and 30th April 2023.

Use Case 4: Multilingual representation of labels Incorporating multilingual information in a resilience project utilizing linked data is crucial for effective international collaboration, enhancing usability for many users, and improving interoperability. Figure 1d displays multilingual labels for the city of Kupyansk, showcasing the diverse linguistic representation in our enriched data. This holistic approach ensures that potential users of the data can actively participate, comprehend, and contribute to the project, ultimately fostering more resilient and inclusive communities.

[2] A video demo is included in the supplementary material for the use cases `https://youtu.be/E_fr1KzfsVs`.

References

1. Eyes on russia. https://eyesonrussia.org/, 2023. Accessed: 2023-04-30.
2. Civilian harm. https://ukraine.bellingcat.com/, 2023. Accessed: 2023-04-30.
3. Ramanathan V Guha, Dan Brickley, and Steve Macbeth. Schema. org: evolution of structured data on the web. *Communications of the ACM*, 59(2):44–51, 2016.
4. Stuart L Weibel and Traugott Koch. The dublin core metadata initiative. *D-lib magazine*, 6(12):1082–9873, 2000.
5. Willem Robert Van Hage, Véronique Malaisé, Roxane Segers, Laura Hollink, and Guus Schreiber. Design and use of the simple event model (sem). *Journal of Web Semantics*, 9(2):128–136, 2011.
6. Dirk Ahlers. Assessment of the accuracy of geonames gazetteer data. In *Proceedings of the 7th Workshop on Geographic Information Retrieval*, GIR '13, page 74–81, New York, NY, USA, 2013. Association for Computing Machinery.
7. Manar Attar, Shuai Wang, Ronald Siebes, and Eirik Kultorp. Converting and enriching geo-annotated event data: Integrating information for ukraine resilience. In *ACM International Conference on Advances in Geographic Information Systems (ACM SIGSPATIAL)*, 2023. Updated datasets, use cases and demos are available at https://linked4resilience.eu/ (accessed 2023-10-04).
8. Clionadh Raleigh, rew Linke, Håvard Hegre, and Joakim Karlsen. Introducing acled: An armed conflict location and event dataset. *Journal of peace research*, 47(5):651–660, 2010.
9. Vasilis Kopsachilis, Nikos Vachtsavanis, and Michail Vaitis. Wikievents - a novel resource for nlp downstream tasks. In *Proceedings of the 5th International Workshop on Semantic Methods for Events and Stories, SEMMES, 2023, Hersonissos, Greece, May 29th, 2023*, CEUR Workshop Proceedings, 2023.
10. Bomb shelters, shelters and defense structures in the regions of ukraine. https://visitukraine.today/blog/101/bomb-shelters-shelters-and-defense-structures-in-the-regions-of-ukraine, 2022. Accessed: 2023-08-01.