Received: 7 May 2021 | Accepted: 16 November 2021

DOI: 10.1002/aaai.12043

SPECIAL TOPIC ARTICLE

Know, Know Where, KnowWhereGraph: A densely connected, cross-domain knowledge graph and geo-enrichment service stack for applications in environmental intelligence

Krzysztof Janowicz[1] | **Pascal Hitzler**[2] | **Wenwen Li**[3] | **Dean Rehberger**[4] | **Mark Schildhauer**[1,5] | **Rui Zhu**[1] | **Cogan Shimizu**[2] | **Colby K. Fisher**[6] | **Ling Cai**[1] | **Gengchen Mai**[7] | **Joseph Zalewski**[2] | **Lu Zhou**[2] | **Shirly Stephen**[1] | **Seila Gonzalez**[4] | **Bryce Mecum**[5] | **Anna Lopez-Carr**[8] | **Andrew Schroeder**[8] | **David Smith**[1] | **Dawn Wright**[9] | **Sizhe Wang**[3] | **Yuanyuan Tian**[3] | **Zilong Liu**[1] | **Meilin Shi**[1] | **Anthony D'Onofrio**[4] | **Zhining Gu**[3] | **Kitty Currier**[1]

[1]University of California, Santa Barbara, California, USA

[2]Kansas State University, Manhattan, Kansas, USA

[3]Arizona State University, Tempe, Arizona, USA

[4]Matrix, Michigan State University, East Lansing, Michigan, USA

[5]National Center for Ecological Analysis and Synthesis, Santa Barbara, California, USA

[6]Hydronos Labs, Princeton, New Jersey, USA

[7]Stanford University, Stanford, California, USA

[8]Direct Relief, Santa Barbara, California, USA

[9]Esri, Redlands, California, USA

Correspondence
Krzysztof Janowicz, University of California, Santa Barbara, CA, USA.
Email: janowicz@ucsb.edu

Abstract

Knowledge graphs (KGs) are a novel paradigm for the representation, retrieval, and integration of data from highly heterogeneous sources. Within just a few years, KGs and their supporting technologies have become a core component of modern search engines, intelligent personal assistants, business intelligence, and so on. Interestingly, despite large-scale data availability, they have yet to be as successful in the realm of environmental data and environmental intelligence. In this paper, we will explain why spatial data require special treatment, and how and when to semantically lift environmental data to a KG. We will present our KnowWhereGraph that contains a wide range of integrated datasets at the human–environment interface, introduce our application areas, and discuss geospatial enrichment services on top of our graph. Jointly, the graph and services will provide answers to questions such as "what is here," "what happened here before," and "how does this region compare to …" for any region on earth within seconds.

INTRODUCTION AND MOTIVATION

Successful decision-makers have strong situational awareness. They have a comprehensive understanding of the context in which their actions will play out. In our global, fast-paced, and densely interconnected world, this context stems from a wide range of heterogeneous resources that span the physical and social sciences. For instance, decision-makers at humanitarian relief organizations need an immediate understanding of physical perils and the regions they affect. When a hurricane causes a disaster, getting supplies to the local population at the right time and location is key. Relief coordinators also need information about previous events such as cholera outbreaks that may have affected the region before the hurricane makes landfall and experts on the ground who can coordinate relief.

Similarly, the agricultural sector, including government agencies, FMIs, individual farmers, and retailers, requires immediate access to data about food safety, wildfires, floods, air pollution, worker health, supply chain disruptions, and transportation networks. For instance, our partners at the Food Industry Association (FMI) want to understand how a wildfire at one place may impact leafy greens, grapes, and the health of workers at another place 100 miles away due to heavy smoke and ashes. Making decisions based on such data is called *Environmental Intelligence* and is gaining traction due to increased environmental stress, correlated shocks, just-in-time supply chains, and a growing interest in *Environmental*, *Social*, and *Corporate governance* (ESG).

Unfortunately, for practical data-driven decision-making and data science, the first stages of gaining situational awareness consume 80% of a project's resources, be it funds, time, or person power. This leaves merely 20% of the resources for the actual analysis that determines the quality of the decisions. More concretely, most resources are spent on data retrieval, cleaning, and integration rather than on deriving insights from data. This puts data-driven decision-making out of range for many tasks. Several solutions to this well-known *data acquisition bottleneck* have been proposed, both in industry and academia. Most either target the retrieval problem by envisioning one-stop data portals or aim at cloud-based access and processing of data.

In the realm of Geographic Information Systems (GIS), one partial solution is geo-enrichment service. For instance, Esri's GeoEnrichment service enables analysts to enrich their local data on-demand with a range of up-to-date demographic variables apportioned to their area of concern and need. This has a number of advantages: (1) in theory, data are always up-to-date and does not age on the analyst's hard disk; (2) in times of misinformation and information overload, the data come from a trusted resource; (3) the data are tailored (apportioned) to the analyst's study area; and finally, (4) the data are GIS-ready in the sense that it can be directly processed, analyzed, and displayed. While current geo-enrichment services are valuable, they also face four key limitations: (1) they only serve data for a small set of predefined categories, such as demographic data. (2) They are closed data silos that encode just one domain/cultural perspective. (3) Because they are centrally maintained, scalability and timely updates become bottlenecks when those services try to incorporate more (diverse) data. (4) They do not have an integration layer that enables follow-up queries over the enriched data. Consequently, a new approach is needed that combines the strength of geo-enrichment services, that is, seamless access to contextual information for an analyst's areas of concern, with a technology that provides open, densely integrated, cross-domain data across a wide range of perspectives (Janowicz 2021).

For these challenges, knowledge graphs (KGs) promise to provide a solution (Hogan et al. 2021; Noy et al. 2019). They are a combination of technologies, specifications, and data cultures for densely interconnecting (Web-scale) data across domains in a human and machine readable and reasonable way. They are a novel approach to publishing, representing, integrating, and interlinking individual data (not merely datasets) by concentrating on connections among places, people, events, and entities instead of their properties. More formally, a KG (as a set of node-edge-node statements called triples) can be thought of as a node and edge-labeled directed multigraph. While the term KG itself does not prescribe any particular technology stack, the largest publicly available KG is the Linked Data cloud based on the RDF/Semantic Web technology stack (Bizer, Heath, and Berners-Lee 2011). Interconnected statements can be of the form ThomasFire → affected → SantaBarbara and SantaBarbara → partOf → California. Together with schemata (ontologies) specified in knowledge representation (KR) languages, these triples would entail a third triple, namely that the Thomas Fire happened in California. As these ontologies encode the semantics of the used terminology, they foster interoperability without restricting semantic heterogeneity (Hitzler 2021; Janowicz et al. 2015).

Inspired by open KGs such as DBpedia (Lehmann et al. 2015) and Wikidata (Vrandečić and Krötzsch 2014) and services such as GeoEnrichment, our KnowWhereGraph provides a densely connected, cross-domain KG and geo-enrichment services for a wide range of applications in environmental intelligence by giving decision-makers and data analysts on-demand access to area briefings at a high spatial and temporal resolution for any location on the surface of the earth. To do so, we translate data about extreme

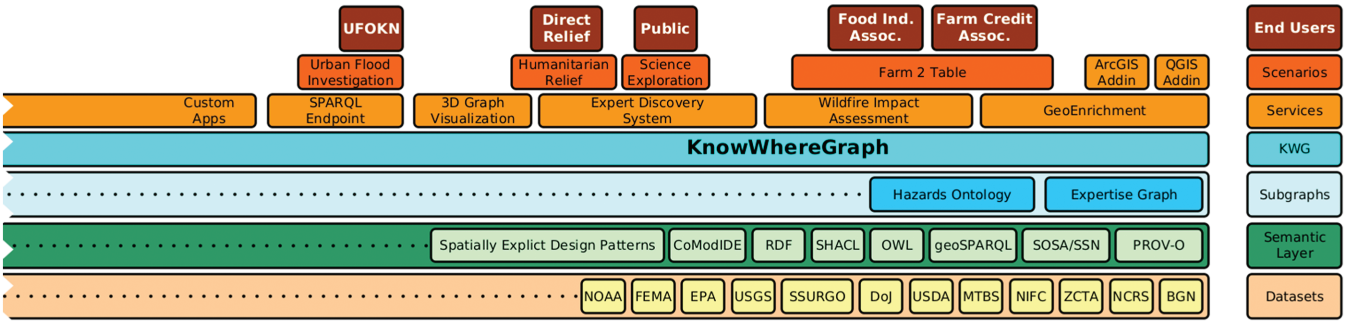


FIGURE 1 A layer-wise depiction of the architecture of KnowWhereGraph and the services and use-cases that it supports (as of August 2021)

events, administrative boundaries, soils, crops, climate, transportation, and so on, into a KG and pre-integrate them to provide answers to questions such as "what is here," "what happened here before," "how does this region compare to ...." While DBpedia and Wikidata contain only rudimentary information about places/regions, such as their populations, we give rapid access to information such as the wildfires that have affected an area, the major transportation axis crossing a certain region, and the type of crops and soils present in a given region.

TECHNOLOGICAL APPROACH

KnowWhereGraph is quickly and continuously growing as new data silos are identified, and subsequently integrated into our graph, based on the needs of our users and application scenarios. We have developed a number of techniques and ontologies to aid in growing and maintaining KnowWhereGraph. Figure 1 shows a layer-wise view of KnowWhereGraph, as well as the services and use-cases it supports, which directly correspond to many of our techniques.

First and foremost, many of our data sources naturally overlap in space and time and we need to manage a vast amount of heterogeneous spatial data. To do so, we partially depart from traditional linked data approaches that often represent spatial regions as points or polygons on the earth's surface. Instead, we utilize a Discrete Global Grid (Bondaruk, Roberts, and Robertson 2020) called the "S2 Grid System." This lays a hierarchical grid over the earth's surface; each grid cell in a level is comprised of four subcells of increasing spatial resolution. KnowWhereGraph serves data at least at S2 Level 11 (about 20 km^2 per cell) for the United States. However, some regions may have a substantially higher resolution based on data availability, rates of change, and application needs. This approach provides a compromise between data precision and access speed in such a way that it does not preempt downstream, finer-grained topological investigations of the original geometries. Figure 2 depicts selected triples from KnowWhereGraph about regions affected by a hurricane, the impacts, and experts on storm-related topics. In addition to grid cells, we serve many other region identifiers with globally unique IDs so that users can request information about them or interlink and thereby enrich their own data. Examples include, FIPS codes, ZIP codes, media market areas, national weather zones, administrative areas, gazetteer features, and so on.

Using the S2 grid system as a base, we developed a design pattern[1] for easily relating how features and regions may interact throughout the hierarchy. Additionally, we have adopted a number of open standards such as *GeoSPARQL*[2] and the *Sensors, Observations, Sample, Actuator (SOSA)* ontology[3] and its extension (Zhu et al. 2021), as well as other frequently used ontologies such as *QUDT*.[4] Modeling all data from a sensor and observation perspective eases querying, connecting data to the geographic features they describe, and also enables us to link data about events with human experts and research results. Finally, we also worked on connectivity and coverage of our graph. In particular, we provide enriched representations of regions, such as climate divisions or counties, and link them to entities from Wikidata or the Geographic Names Information System, where possible, giving instant access to a wide range of broad contextual information such as population density, previous extreme events, soil health, and so on; topological relations (e.g., RCC8) among regions for flexible inference and triple compression; and link together events and places through causal relationships and provenance (Shimizu et al. 2021). For instance, we model where a fire took place, which events it triggered, and which regions have been affected, for example, by heavy smoke. Altogether, this allows domain scientists to represent geospatial objects, which are traditionally represented as vector geometries, as a collection of S2 cells at various hierarchical grid levels and instantly have tight integration with any other dataset in KnowWhereGraph.

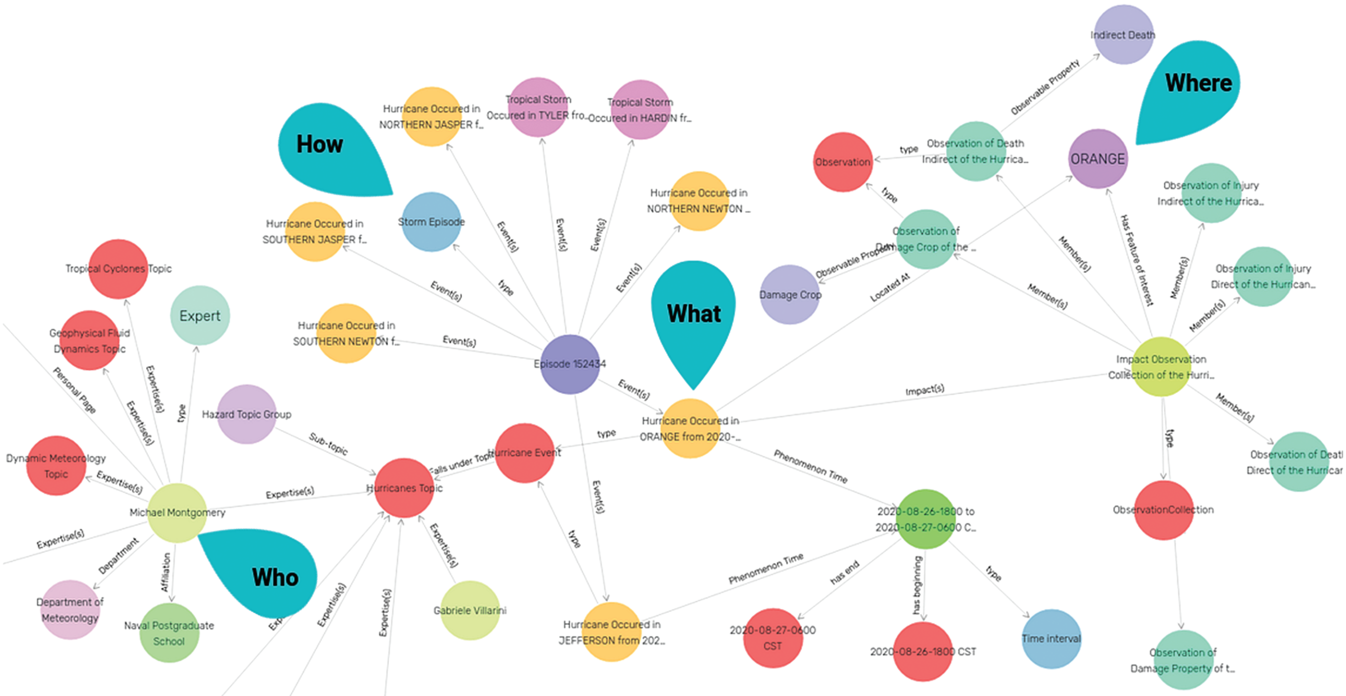


FIGURE 2 Triples from KnowWhereGraph about a hurricane, impacted areas, impacts, and experts on relevant topics

In Figure 2, for example, we focus on named places such as counties, but users may request storm damage for any collection of S2 grid cells. The level of S2 cells is not uniform across regions but depends on data layers and (in the future) also on the underlying variation within these layers across space.

CHALLENGES AND RELATION TO ARTIFICIAL INTELLIGENCE

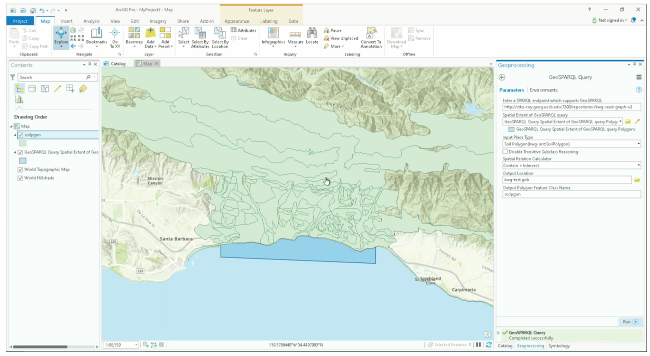
KG technology is to a substantial part based on KR methods and thus on the corresponding subfield of Artificial Intelligence (Hitzler 2021). In particular, the central W3C standards RDF (Resource Description Framework; Cyganiak, Wood, and Lanthaler (2014)) and OWL (Web Ontology Language (Hitzler et al. 2012)) for representing graphs and their schemas (known as ontologies), are formal logics in the tradition of the KR field (Hitzler, Krötzsch, and Rudolph 2010).

However, in contrast to traditional lines of KR research, recent developments in KG data management shift the focus to pragmatics, in particular how to make KR work in practice—at industrial scale, functionality, and stability levels—for data management. While traditional academic literature on KR has a heavy focus on developing KR languages and provably correct and theoretically analyzed algorithms, pragmatic aspects such as the question which KR approach works the best in which situation, or how to apply a KR framework or representation language to an industry scale problem, have played a minor role in academic outlets. Similar questions such as how to lift individual data to the graph, when to do so, at which resolution (e.g., level of granularity), and how to balance schema complexity between optimizing for use versus reuse remain largely unanswered.

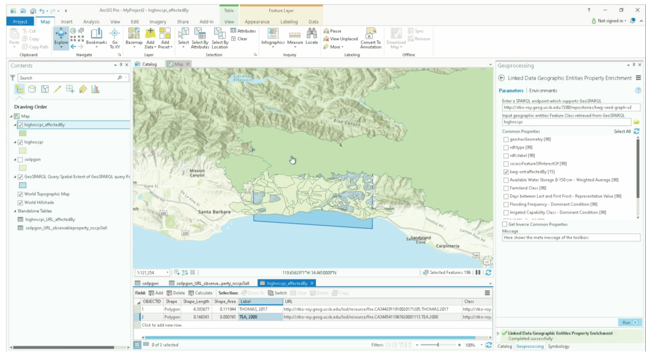
KnowWhereGraph focuses on this transition gap between theoretical results and applicability in practice. In particular, it is about the general question of how to achieve practically relevant levels of scale and speed based on real high-volume heterogeneous data from diverse sources, and how to do this without an undue compromise of the quality in representation and solutions that come out of the KR field. In other words, KnowWhereGraph is about finding the right trade-off between principled approaches and rapid, scalable development. It is about finding the sweet spot between theory and practice. In case of KnowWhereGraph, this happens in the context of a multidisciplinary setting that requires rapid convergence across topics such as climate forecasts, extreme events, health, supply chains, and even the spatio-temporal bounds of human expertise for our pilot in disaster relief. Integrating these datasets also requires solutions that can handle noisy and missing (and contradictory) data, as well as changes in perspective as they relate to different schema, and services that enable data exploration using similarity-based search. To handle real-world and noisy data, our work combines symbolic and sub-symbolic methods for representation and reasoning.

Particular challenges related to Artificial Intelligence that we address are (1) bringing principled KG methods to a level of maturity sufficient for transfer to industrial practice, (2) scaling up of methods and processes for our applications for which we currently project a required KG containing about 10 billion triples, and (3) KG methods and tools development that is aimed at maximum flexibility for future growth, extension, and reuse.

Specific innovations within the KnowWhereGraph work that are relevant for Artificial Intelligence include: In terms of representation of spatial knowledge, we have combined hierarchical grids with standard region boundaries and Region Connection Calculus methods (Zalewski, Hitzler, and Janowicz 2021), in what we believe is a novel approach for KGs to meet scale and uniformity requirements. In terms of access to large-scale spatial data, we are integrating KG and GIS technology by offering graph-based geo-enrichment and *n*-degree property path queries from within a GIS. With respect to KR methods, we are combining top-down and bottom-up ontology engineering processes with a principled modular approach to KG schema development to balance between quality of the graph model and speed of development and integration.



(A) Retrieving soil polygons



(B) Retrieve wildfires that affected soil polygons

FIGURE 3 Our Knowledge Graph-based geo-enrichment toolbox collections for ArcGIS Pro. (A) The GeoSPARQL Query toolbox, (B) The Property Enrichment toolbox

CURRENT STATUS

While KnowWhereGraph can serve a wide range of domains and use cases that require spatial data and spatial question answering, we have three initial pilots: Humanitarian relief: Together with Direct Relief we demonstrate how our technologies can inform humanitarian supply chains and help identify and match domain experts to the needs of an emerging crisis. Farm to table supply chain and sustainability: In collaboration with the FMI, we demonstrate how KGs can enhance the sustainability, efficiency, and safety of consumer food supply with a focus on the impact of wildfires on agriculture and food security. Land valuation and risk of default: This new pilot is a joint research with farm credit associations concerned with driver-based land potential assessment for model-based valuation and risk assessment for agricultural credit applications and loan portfolio monitoring.

To date we have included 27 different data layers from 16 major data sources that extensively cover the topics discussed in the domain application areas (e.g., climate hazard, wildfire, and air quality). At the time of writing, KnowWhereGraph already consists of about 4.9B triples, and we expect it to grow to as many as 10–20B triples over the next years as we ingest additional data. Put differently, in contrast to data portals, we do not merely provide access to datasets for download but make every single data record within these sets web-available and query-able. We will also provide area briefings at even higher S2 cell resolution, achieve global coverage (beyond our mostly US-centric data), as well as mine new and more complex relationships across the described places and events.

Built upon the KnowWhereGraph, our geo-enrichment services provide a set of toolboxes that support domain scientists to explore environment-related knowledge from within a GIS in various ways such as region-based spatial data retrieval (e.g., soil polygons can be retrieved based on a user-defined study area as shown in Figure 3(A)), property enrichment for geographic entities (e.g., querying a crop productivity index for each of millions of soil polygons), direct relation exploration among geographic entities (e.g., querying for landslides on soils previously affected by wildfires as shown in Figure 3(B)), and *n*-degree relation identification (e.g., SoilPolygonA → affectedBy → ThomasFire → causedEvent DebirsFlowX).

We have also developed a range of additional services tailored to our vertical applications. For example, the

FIGURE 4 Left: Similarity interface for experts. Right: Follow-your-nose interface for previous disasters

KnowWhereGraph enables disaster relief specialists to explore knowledge about experts and their areas of expertise, as related to specific disasters. To achieve this, we provide a similarity search interface and a follow-your-nose interface, which are shown in Figure 4. In case of the similarity interface, users can type in an expert name into the search box and the system will return the top 15 experts who are most similar. The similarity score is computed using a combination of Doc2Vec and KG embedding techniques (Le and Mikolov 2014; Mai, Janowicz, and Yan 2018), which are computed based on the particular expert's three most cited papers, three most recent papers, and their relation to other experts in the graph. Figure 4 (left) shows an example of the similarity search. From there, users can directly search information about the experts, their area of expertise, and events that they have worked on. Conversely, users can start by selecting a certain event or a geographic region, learn about previous events, their impacts, and the relevant experts that could be contacted. In fact, this ability to seamlessly navigate between physical events, areas of expertise, affected regions, and people is one of the key strengths of our KG.

In terms of our food safety work, KnowWhereGraph is used to enhance assessment and strategic planning during near real-time hazard events affecting the food supply chain by providing online analysis, forecasting, and alerts that are enriched with location and context-specific intelligence, to ensure that key stakeholders throughout the supply chain are ready with backup strategies to keep products moving. It also allows farmers and growers to identify how they can be better prepared to mitigate and build resilience in the face of such events. Currently, our graph serves pre-integrated data about wildfires, smoke plumes, and crop locations, together with topological information

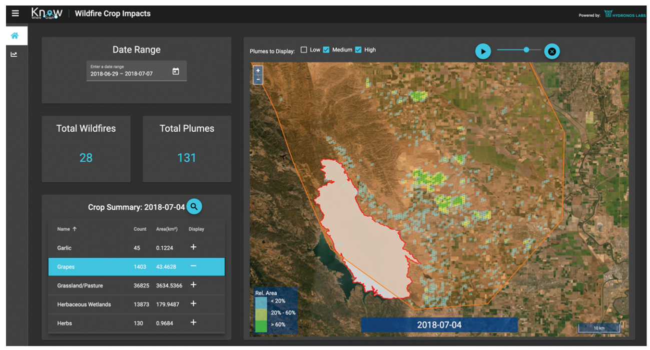


FIGURE 5 Wildfire crop impacts interface displaying a smoke plume (yellow shape) from July 4, 2018 associated with the County Fire (red outlined shape). Within this plume, we have queried for areas with high densities of grapes to identify areas where the crop may be affected by smoke taint

about the affected areas. In one implementation for FMI, a custom front-end web interface (Figure 5) and API enable decision-makers to process a series of queries important to assessing the impact of ongoing wildfires, smoke plumes, and ashes on key food (crop) supply chains. Users can progress through these queries without any experience in using complex GIS software or the specific data and analysis techniques necessary, seeing visualizations of the results at each step. Despite the simplicity of this system, the interface is dynamically generating SPARQL queries based on the user inputs (e.g., defining a region of interest, selecting multiple crop types), sending these queries to the graph via an API and receiving/displaying the results, all within a matter of seconds. This system highlights the ease with which new bespoke end-user applications can be developed from the core resources of the KnowWhereGraph, enabling a multitude of use cases at the human–environment interface.

FUTURE PLANS

In this work, we have introduced the KnowWhereGraph, a densely connected, cross-domain KG together with geo-enrichment services to support a variety of application areas that benefit from environmental intelligence. Our graph delivers area briefings for any place on earth within seconds to answer questions such as “what is here” or “what happened here before.” For instance, decision-makers and data scientists can easily retrieve all extreme events (e.g., previous storms, fires, cholera outbreaks) that have impacted an area that is predicted to be in the path of an approaching hurricane. Most importantly, we do not only serve individual data (observations) across many layers, but also connections across them. For instance, graph hubs such as Wikidata or DBpedia contain information about Santa Barbara, the Thomas Fire, highway 101, and the 2018 debris flow in Southern California. However, they do not locate the fire nor the debris flow and most importantly do not contain facts such that the fire affected Santa Barbara and that the fire and a massive storm caused a debris flow that killed 23 people and disrupted transportation for weeks as it blocked highway 101. This is exactly the type of relationships that we are most interested in exploring in the future. We also do not just serve data at pre-defined levels, for example, counties, but deliver a variety of regions identifiers thereby making KWG a gazetteer of gazetteers. In addition, we also serve data registered to fine-grained global grid cells. So far, the KnowWhereGraph largely contains information about the United States due to easy access to high-quality, well-documented governmental data, as many of our use cases revolve around the United States, and to keep the graph size at bay. In the future, we will increase global coverage, add more data layers, enable geo-enrichment for open source GIS and spatial statistics packages in general, and mine more (complex) relationship across our entities with the ultimate goal of creating a global KG of environmental and geographic information.

ACKNOWLEDGMENTS

The authors acknowledge support by the National Science Foundation under Grant 2033521 A1: KnowWhereGraph: Enriching and Linking Cross-Domain Knowledge Graphs using Spatially-Explicit AI Technologies. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

CONFLICT OF INTEREST

No conflict of interest has been declared by the author(s).

ENDNOTES

[1] https://github.com/KnowWhereGraph/hierarchical-cell-features
[2] https://www.ogc.org/standards/geosparql
[3] https://www.w3.org/TR/vocab-ssn/
[4] http://www.qudt.org/

REFERENCES

Bizer, C., T. Heath, and T. Berners-Lee. 2011. “Linked Data: The Story So Far.” In *Semantic Services, Interoperability and Web Applications: Emerging Concepts*, 205–27. IGI Global.

Bondaruk, B., S. A. Roberts, and C. Robertson. 2020. “Assessing the State of the Art in Discrete Global Grid Systems: OGC Criteria and Present Functionality.” *Geomatica* 74 (1): 9–30.

Cyganiak, R., D. Wood, and M. Lanthaler, eds. 2014. “RDF 1.1 Concepts and Abstract Syntax.” W3C Recommendation. February 25, 2014. http://www.w3.org/TR/rdf11-concepts/

Hitzler, P. 2021. “A Review of the Semantic Web Field.” *Commun ications of the ACM* 64 (2): 76–83.

Hitzler, P., M. Krötzsch, B. Parsia, P. F. Patel-Schneider, and S. Rudolpheds. 2012. “OWL 2 Web Ontology Language: Primer.” 2nd ed. W3C Recommendation. December 11, 2012. http://www.w3.org/TR/owl2-primer/

Hitzler, P., M. Krötzsch, and S. Rudolph. 2010. *Foundations of Semantic Web Technologies*. Chapman and Hall/CRC Press.

Hogan, A., E. Blomqvist, M. Cochez, C. d'Amato, G. de Melo, C. Gutierrez, S. Kirrane, et al. 2021. “Knowledge Graphs.” *Synthesis Lectures on Data, Semantics, and Knowledge* 12(2): 1–257.

Janowicz, K. 2021. “Knowwheregraph Drives Analytics and Cross-Domain Knowledge.” *ArcUser*, 16–9.

Janowicz, K., F. Van Harmelen, J. A. Hendler, and P. Hitzler. 2015. “Why the Data Train Needs Semantic Rails.” *AI Magazine* 36 (1): 5–14.

Le, Q. and T. Mikolov. 2014. “Distributed Representations of Sentences and Documents.” In *International Conference on Machine Learning*, 1188–96, PMLR.

Lehmann, J., R. Isele, M. Jakob, A. Jentzsch, D. Kontokostas, P. N. Mendes, S. Hellmann, et al. 2015. “Dbpedia—A Large-Scale, Multilingual Knowledge Base Extracted from Wikipedia.” *Semantic Web* 6 (2): 167–95.

Mai, G., K. Janowicz, and B. Yan. 2018. “Combining Text Embedding and Knowledge Graph Embedding Techniques for Academic Search Engines.” In *Proceedings of the Semdeep/NLIWoD@ ISWC*, 77–88.

Noy, N. F., Y. Gao, A. Jain, A. Narayanan, A. Patterson, and J. Taylor. 2019. “Industry-Scale Knowledge Graphs: Lessons and Challenges.” *Communications of the ACM* 62 (8): 36–43.

Shimizu, C., R. Zhu, M. Schildhauer, K. Janowicz, and P. Hitzler. 2021. “A Pattern for Modeling Causal Relations between Events.” In *Proceedings of the 12th Workshop on Ontology Design and Patterns (WOP 2021), co-located with the 20th International Semantic Web Conference (ISWC 2021)*, Volume 3011, 38–50. October 24, 2021.

Vrandečić, D. and M. Krötzsch. 2014. “Wikidata: A Free Collaborative Knowledgebase.” *Communications of the ACM* 57 (10): 78–85.

Zalewski, J., P. Hitzler, and K. Janowicz. 2021. “Semantic Compression with Region Calculi in Nested Hierarchical Grids.” In *Proceedings of the SIGSPATIAL'21: 29th International Conference on Advances in Geographic Information Systems, Virtual Event*, ed. X.

Meng, F. Wang, C. Lu, Y. Huang, S. Shekhar, and X. Xie, 305–8, Beijing, ACM. November 2–5, 2021.

Zhu, R., S. Ambrose, L. Zhou, C. Shimizu, L. Cai, G. Mai, K. Janowicz, P. Hitzler, and M. Schildhauer. 2021. "Environmental Observations in Knowledge Graphs." In *Proceedings of the 2nd Workshop on Data and Research Objects Management for Linked Open Science*.

AUTHOR BIOGRAPHIES

Krzysztof Janowicz is a Professor for Geoinformatics and director of the Center for Spatial Studies at the University of California, Santa Barbara. His research interests include knowledge graphs, GeoAI, geo-semantics, and geographic information retrieval. Janowicz is studying how humans conceptualize the space around them based on their behavior, focusing particularly on regional and cultural differences with the ultimate goal of assisting machines to better understand the information needs of an increasingly diverse user base.

Pascal Hitzler is a Professor and endowed Lloyd T. Smith Creativity in Engineering Chair and Director of the Center for Artificial Intelligence and Data Science (CAIDS) at the Department of Computer Science at Kansas State University. His research interests include Semantic Web, Knowledge Representation and Reasoning, and Neuro-Symbolic Artificial Intelligence.

Wenwen Li is a Professor in GIScience in the School of Geographical Sciences and Urban Planning, Arizona State University. She also directs the Cyberinfrastructure and Computational Intelligence Lab (http://cici.lab.asu.edu/). Li's research interests include cyberinfrastructure, big data, geospatial artificial intelligence (GeoAI), and their applications in data-intensive environmental and social sciences, including global warming and Arctic change, terrain analysis, disaster relief, and water insecurity in underserved communities.

Dean Rehberger is the Director of Matrix and Faculty of History at MSU. Dean specializes in developing digital technologies for history and cultural heritage. He also oversees Matrix project planning, research, and development, coordinating many of the projects for the Center.

Mark Schildhauer is a Research Associate at the National Center for Ecological Analysis and Synthesis, after being NCEAS' Director of Computing from its opening in 1995 until 2017. His technology research interests are in the areas of environmental informatics, data semantics, the Semantic Web, Knowledge Graph technologies, and Open Science, especially in the context of facilitating integrative environmental and conservation science.

Rui Zhu is a postdoctoral scholar at the Center for Spatial Studies, University of California, Santa Barbara. His research focuses on geospatial semantics, geospatial knowledge graphs, spatial statistics, as well as their broader interactions in geospatial artificial intelligence (GeoAI).

Cogan Shimizu is a Postdoctoral Researcher at Kansas State University and co-leads the Data Semantics Laboratory. He focuses primarily on the methodological and pedagogical aspects of knowledge engineering, in particular the use of pattern-based methods for improved outcomes in both manual and automated learning tasks.

Colby K. Fisher is the Director of R&D and Managing Partner of Hydronos Labs, an independent environmental software consulting and research firm based in Princeton, NJ, as well as an advisor to Oliver Wyman and the Center for Spatial Studies at the University of California, Santa Barbara. His research involves global scale hydrologic modeling, remote sensing data assimilation, and high-performance computing with a focus on applications in hydrologic extremes and their impact on society. He holds a Ph.D. degree in water resources, hydrology, and remote sensing from Princeton University.

Ling Cai is a PhD Student at the Center for Spatial Studies, University of California, Santa Barbara. Her research interests include geospatial semantics, geospatial knowledge graphs, qualitative spatial and temporal reasoning, as well as geospatial artificial intelligence.

Gengchen Mai is a Postdoctoral Fellow at Stanford Artificial Intelligence Laboratory, Department of Computer Science, Stanford University. He got his Ph.D. in Geographic Information Science from the University of California, Santa Barbara. His research interests include spatially explicit machine learning, GeoAI, geospatial knowledge graph, geographic question answering, and computational sustainability.

Joseph Zalewski is a Ph.D. Student in Computer Science at Kansas State University, working with Dr. Pascal Hitzler. He is mainly interested in logic and theoretical computer science.

Lu Zhou is a Postdoctoral Researcher in Data Semantics Laboratory at Kansas State University. He focuses primarily on research on the topic of Knowledge Graphs Construction, Semantic Data Integration, and Deep Deductive Reasoning, in particular, large-scale Knowledge Graph Construction and Enrichment, Ontology Matching and Alignment, Entity Resolution, and Neural-Symbolic Integration.

Shirly Stephen is a Postdoctoral Scholar at the Center for Spatial Studies, University of California, Santa Barbara. Her research focuses on geospatial semantics and geo-ontologies.

Seila G. Estrecha manages and oversees the design and development of all software at Matrix, including all frontend and backend aspects of web applications, designing databases architecture, decision made for tools and technologies to be implemented, roadmap of software development of any Matrix products, identifying issues and common patterns, and developing standard operating procedures. She has experience implementing semantic web-based systems and standards for ontology-centered knowledge graphs, including work on knowledge graph modularization, ontology design patterns, interdisciplinary knowledge graph development, ontology alignment, data integration, and implementation of SPARQL queries.

Bryce Mecum is a Software Engineer at the National Center for Ecological Analysis and Synthesis. He holds an MS in Fisheries from the University of Alaska, Fairbanks and is interested in ecological forecasting and building reliable software systems for the sciences.

Anna Lopez-Carr is the Monitoring, Evaluation, and Learning specialist at Direct Relief, a humanitarian aid and disaster response organization based in Santa Barbara, CA. She holds a Ph.D. in Geography from the University of California, Santa Barbara; and an M.Sc. in Political Ecology from the University of London. Her research interests include population, health, and the environment, particularly in relation to natural disasters, health equity, and low-resourced communities.

Andrew Schroeder is Vice President of Research and Analysis at Direct Relief and Co-Founder and Board President of WeRobotics.org.

David Smith is a Soil Scientist and former federal Senior Executive with the U.S. Department of Agriculture now serving as an advisor to Oliver Wyman, Hydronos Labs, and the Center for Spatial Studies at the University of California, Santa Barbara. He brings expertise in soil science and related agricultural and natural resources management.

Dawn Wright is Chief Scientist of the Environmental Systems Research Institute (aka Esri), a privately held and world-leading geographic information system and spatial data science company, as well as a professor of Geography and Oceanography at Oregon State University. She holds a joint Ph.D. in Physical Geography and Marine Geology from the University of California, Santa Barbara; and an M.Sc. in Political Ecology from the University of London. Her research interests include global ecosystem characterization and mapping, seafloor mapping, and ocean informatics.

Sizhe Wang is a Ph.D. Student at Arizona State University, majoring in computer science. His research interests include terrain feature detection and recognition, multidimensional geospatial data visualization. He also uses visualization techniques to help information discovery in KnowWhereGraph.

Yuanyuan Tian is a PhD Student at Arizona State University. Her research interests include causality, GeoAI, as well as human mobility and coastal resource management, especially in megacities and megadeltas. She also uses natural language processing and ontology engineering to discover geographic knowledge.

Zilong Liu is an M.A./Ph.D. Student working at the Center for Spatial Studies in the University of California, Santa Barbara. His research interests include GIScience, geospatial semantics, and spatial-temporal knowledge representation and reasoning.

Meilin Shi is a PhD Student at University of California, Santa Barbara. Her research interests include geospatial semantics, geographic knowledge graphs, and spatial statistics.

Anthony D'Onofrio is the lead developer on the KORA Digital Repository platform. A graduate of Michigan State University, and former student developer intern of Matrix, Anthony returned in 2015 to head up development on the 3.0 release of KORA.

Zhining Gu is a PhD Student in GIScience in the School of Geographical Sciences and Urban Planning, Arizona State University. Her research interests include spatial analysis, machine learning, and geographic knowledge graph.

Kitty Currier is a postdoctoral scholar at the Center for Spatial Studies, University of California, Santa Barbara with an interest in participatory mapping and kite aerial photography.

How to cite this article: Janowicz, K., P. Hitzler, W. Li, D. Rehberger, M. Schildhauer, R. Zhu, C. Shimizu, C. K. Fisher, L. Cai, G. Mai, J. Zalewski, L. Zhou, S. Stephen, S. Gonzalez, B. Mecum, A. Lopez-Carr, A. Schroeder, D. Smith, D. Wright, S. Wang, Y. Tian, Z. Liu, M. Shi, A. D'Onofrio, Z. Gu, and K. Currier. 2022. "Know, Know Where, KnowWhereGraph: A densely connected, cross-domain knowledge graph and geo-enrichment service stack for applications in environmental intelligence." *AI Magazine* 43: 30–39. https://doi.org/10.1002/aaai.12043