

The KnowWhereGraph Ontology

Cogan Shimizu^{*,*,a}, Shirley Stephen^{**,b}, Adrita Barua^c, Ling Cai^b, Antrea Christou^a, Kitty Currier^b, Abhilekha Dalal^c, Colby K. Fisher^d, Pascal Hitzler^c, Krzysztof Janowicz^{e,b}, Wenwen Li^f, Zilong Liu^{e,b}, Mohammad Saeid Mahdavejad^c, Gengchen Mai^g, Dean Rehberger^h, Mark Schildhauerⁱ, Meilin Shi^{e,b}, Sanaz Saki Norouzi^c, Yuanyuan Tian^f, Sizhe Wang^f, Zhangyu Wang^b, Joseph Zalewski^c, Lu Zhou^c, Rui Zhu^j

^aWright State University

^bUniversity of California, Santa Barbara

^cKansas State University

^dHydronos Labs

^eUniversity of Vienna

^fArizona State University

^gUniversity of Georgia

^hMichigan State University

ⁱNational Center for Ecological Analysis & Synthesis

^jUniversity of Bristol

Abstract

KnowWhereGraph is one of the largest fully publicly available geospatial knowledge graphs. It includes data from 30 layers on natural hazards (e.g., hurricanes, wildfires), climate variables (e.g., air temperature, precipitation), soil properties, crop and land-cover types, demographics, and human health, various place and region identifiers, among other themes. These have been leveraged through the graph by a variety of applications to address challenges in food security and agricultural supply chains; sustainability related to soil conservation practices and farm labor; and delivery of emergency humanitarian aid following a disaster. In this paper, we introduce the ontology that acts as the schema for KnowWhereGraph. This broad overview provides insight into the requirements and design specifications for the graph and its schema, including the development methodology (modular ontology modeling) and the resources utilized to implement, materialize, and deploy KnowWhereGraph with its end-user interfaces and public query SPARQL endpoint.

Key words: modular ontology modeling, ontology, spatially enabled knowledge graphs

1. Introduction

KnowWhereGraph¹ (KWG) is one of the largest publicly available geospatial knowledge graphs in the world [15]. It brings together over 30 data layers related to observations of natural hazards (e.g., hurricanes, wildfires, and smoke plumes), spatial characteristics related to climate (e.g., temperature, precipitation, and air quality), soil properties, crop and land-cover types, demographics, and human health (e.g., social vulnerability index), among others, resulting in a knowledge graph with over 16 billion triples.

KWG supports applications in the food, agriculture, and humanitarian relief, and their attendant supply chains, generally; environmental policy issues relative to interactions among agricultural sustainability, soil conservation practice, and farm labor; and

*Corresponding author

**Marked authors contributed equally, all others alphabetical

Email addresses: cogan.shimizu@wright.edu (Cogan Shimizu), shirlystephen@ucsb.edu (Shirly Stephen), adrita@ksu.edu (Adrita Barua), lingcai@ucsb.edu (Ling Cai), christou.2@wright.edu (Antrea Christou), kcurrier@ucsb.edu (Kitty Currier), adalal@ksu.edu (Abhilekha Dalal), colby@hydronoslabs.com (Colby K. Fisher), hitzler@ksu.edu (Pascal Hitzler), krzysztof.janowicz@univie.ac.at (Krzysztof Janowicz), wenwen@asu.edu (Wenwen Li), zilong.liu@univie.ac.at (Zilong Liu), saeid@ksu.edu (Mohammad Saeid Mahdavejad), gengchen.mai25@uga.edu (Gengchen Mai), rehberge@msu.edu (Dean Rehberger), schild@nceas.ucsb.edu (Mark Schildhauer), meilin.shi@univie.ac.at (Meilin Shi), sanazsn@ksu.edu (Sanaz Saki Norouzi), yuanyuantian@asu.edu (Yuanyuan Tian), wsizhe@asu.edu (Sizhe Wang), zhangyuwang@ucsb.edu (Zhangyu Wang), jzalewski@ksu.edu (Joseph Zalewski), lu.zhou@ksu.edu (Lu Zhou), rui.zhu@bristol.ac.uk (Rui Zhu)

¹<https://knowwheregraph.org/>

delivery of emergency humanitarian aid within the US and internationally.

KnowWhereGraph mission is to provide area briefings within seconds for any region on Earth (currently mostly in the US) to answer questions such as

- "What is here?"
- "What happened here before?"
- "Who [experts] knows more?"
- "How does it compare to other regions or previous events?"

In order to assist decision-makers and data scientists in enriching their own data with billions of connected, up-to-date contextual information about humans and their environment to rapidly gain the situational awareness required for proper decision-making.

The highly varied but critical importance of the use cases combined with a large number of data layers and triples created unique challenges for the project in terms of management and execution, challenges that required new approaches to structuring the knowledge graph for efficient querying and use.

To this end and to materialize KnowWhereGraph, we have added an additional layer — a schema — which is described formally as an ontology. This allows us to address the design and functionality requirements resulting from the nature of our mission and use cases [16].

Enable Spatial Integration The primary purpose of KnowWhereGraph is to provide a convenient method for integrating data along a spatiotemporal dimension. This is integral to the mission of the project and, subsequently, a core requirement for the graph and its schema.

Facilitate Data Integration KnowWhereGraph must be capable of providing an overarching framework for the semantic alignment of key terms and concepts.

Provide Rich Inferences Beyond a flat representation of the (integrated) datasets, KnowWhereGraph’s schema must be expressive enough to infer latent relationships between data layers, such as causality of events or the inheritance of spatial characteristics. Furthermore, it must provide a framework for additional inferences in the future.

Highly Maintainable To remain useful, KnowWhereGraph must be easily maintained by the community. This includes both the ease of data integration and how amenable the schema —

and thus the graph — are to modification: either through the incorporation of new or evolving use cases, rectifying conceptual errors in the graph, or adapting to changes in the data sources.

As a consequence of these stringent requirements, we identified the Modular Ontology Modeling (MOMo) methodology as a prime candidate for developing the KnowWhereGraph Ontology. MOMo is a *pattern-based* method, meaning that KWG leverages existing ontology design patterns (either developed by the community or extracted from existing standards) [6], modifies them to the appropriate context or use case [11], and assembles these modules into a modular “plug-and-play” schema, allowing for maximum maintainability [12, 23].

Such a pattern-based approach furthermore allows for relatively straightforward dataset integration. Such a pattern-based approach furthermore facilitates dataset integration. When integrating multiple datasets, it can be convenient to conceptualize them along the same (ontological) dimension. For example, understanding tabular data about a place as observations, similarly to how a hazard impacts a place (and the measurement thereof) is also an observation. This results in a predictable method for querying against the final, integrated schema of the knowledge graph.

Explicitly, the contributions that this manuscript describes are as follows:

- the KnowWhereGraph Ontology (2023), design principles, and implementation;
- the principled leveraging of a discrete global grid for spatial integration; and
- a representative example of answering (one aspect of) an important question that is driven by one of KWG’s use-case scenarios.

The structure of the paper is as follows. First, we provide in Section 2 some preliminary concepts to orient readers. Section 3 details the ontology modeling approach taken to develop the KnowWhereGraph Ontology, which is presented in Section 4, alongside discussion on its implementation. In Section 5, we provide context on how the ontology can be leveraged to use the KWG. Section 6 discusses related work and parallel efforts, and how the KWG connects or overlaps. Finally, in Section 7, as we conclude, we identify next steps.

2. Preliminaries

First, we introduce two important concepts: the Discrete Global Grid, which informs some of our ontological choices, and schema diagrams, which

use a simple graphical syntax to convey intuitively relationships between classes. Finally, we provide a full directory of availability and usage information can be found in Section 2.3.

2.1. The Discrete Global Grid

One of the primary motivations for KWG is to act as a geospatial informational backbone for any dataset describing physical phenomena and their impacts or places and their characteristics. All such datasets are mapped onto a *discrete global grid* (DGG) [8]. In our case, it is a hierarchical DGG, which then acts as the common spatial data frame to achieve a compromise between data precision (i.e., the fidelity of the geometries underlying geospatial phenomena, including regions), access speed, and ease of data integration among multiple vector and raster data sets, while supporting globally unique identifiers for the constituent cells.

In particular, the hierarchical DGG is a partitioning of the surface of the Earth into some number of “top-level” parent cells, which are then further partitioned into child cells, and so on until the desired spatial resolution is achieved [8, 27]. KnowWhereGraph utilizes the S2 Grid System² implementation, but other DGGs were also considered (e.g., Uber’s H3³).⁴

A DGG allows for a common underlying spatial reference system for both geospatial vector and raster datasets, for which we can pre-compute spatial relations between different features or regions for vector data and pre-compute summary statistics for a given cell at a certain level from different raster datasets. By emphasizing the notion of a cell, we can examine exact cell characteristics, predict or infer contents of its parent and child cells, and quickly generate an overview of spatially co-located features and regions of interest without having to compute directly spatial relations.

Furthermore, we can ontologically encode how geospatial information and other phenomena may affect (e.g., transitively) places across the hierarchy of cells within the DGG. This means that we can represent geospatial information and other phenomena at different levels of granularity. This method is discussed more deeply in Section 4, with an example in Section 5. Figure 1 shows an interface that uses the S2 cells to examine quickly regions of the surface of the Earth. By selecting specific cells, we

can use them both as a common integrator and as a way to examine what other datasets integrate with (or map onto) specific S2 cells.

2.2. Schema Diagrams

Following the MOMo methodology [23], we use a (visual) graphical structure called a schema diagram as our primary method for communicating ontological structure. This diagram carries a reduced semantics; it is meant to be intuitive and easily understood, rather than explicitly and visually conveying the exact, underlying logical axioms. We do, however, aim for a consistent visual syntax across all diagrams.

Boxes of any non-gray color indicate a class. Goldenrod boxes are atomic classes. Purple boxes indicate a class that strictly consists of a controlled set of individuals. Blue boxes (with dashed borders) indicate *hidden complexity* — i.e., that there are additional relations, but which have been removed from view for clarity. Frequently, this means that it represents a class that is drawn from outside of the KWG namespaces (e.g., OWL Time [5]). Large gray boxes that encapsulate many arrows and boxes indicate a module, meaning that they are conceptually related. Yellow ellipses indicate a datatype. These are generally prefixed with the appropriate namespace, for clarity. Occasionally, we will use duplicate boxes (i.e., identical labels), which are a convenience, so as to unclutter the diagram of intersecting edges. Filled arrows indicate a binary relation. If one points to a box, then it is an object property. If one points to an ellipse, it is a data property. Open-face arrows indicate a subclass relation. Occasionally, use of an open arrow, stylized with an additional mark, indicates that the subject or source of the arrow is an individual of the type that is represented by the box at the target.

2.3. Resource Availability

We provide multiple types of documentation for the KnowWhereGraph Ontology: visual representations (the schema diagrams), living documentation (generated using [7]), the formalization of the ontology, and a static, technical report (generated using [25]) can all be found in [30]. The KnowWhereGraph, its ontologies, and tools are maintained by the KnowWhereGraph team; more information can be found online.⁵

²<https://s2geometry.io/>

³<https://eng.uber.com/h3/>

⁴A thorough treatment of the different DGG systems is out of scope for this article; we refer the reader to a recent survey [2].

⁵<https://knowwheregraph.org/>

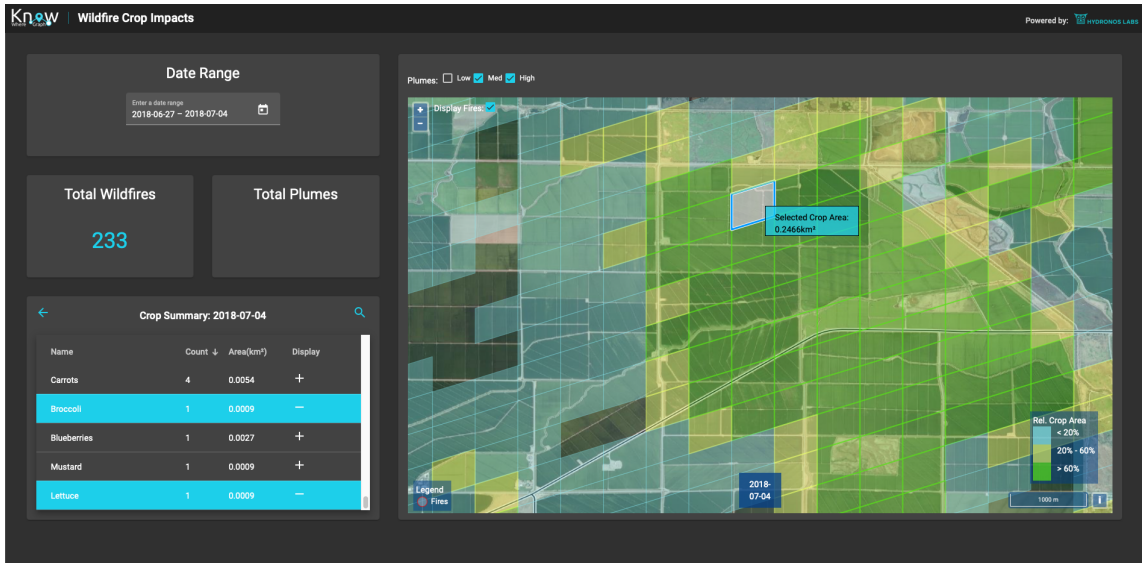


Figure 1: This interface displays a grid-based view of land-use/land-cover data, which are integrated with data from nearby wildfires and their smoke plumes. The grid tessellates the Earth and is called a *discrete global grid*.

3. The Ontology Modeling Approach

For KWG to be both a success as a project, and a usable resource, the ontology, that would act as its schema, would need to be sustained easily beyond the project’s lifetime. Furthermore, it would need to be easily adaptable to new data sources or unexpected changes in data sources to the extent possible both during the initial development and beyond the project’s funding period as the pilot (vertical) use cases are transformed. As such, we identified quickly the Modular Ontology Modeling methodology (MOMo; [23]) as an ideal candidate for performing the development work, since the pattern-based approach both allowed for iterative development, and mitigated long-term maintainability concerns with its resulting modular structure.

The MOMo methodology is a combination of top-down and bottom-up approaches. *Top-down* refers to the fact that MOMo generally assumes that the use cases for the ontology (or knowledge graph) are known at the outset. Then, interactions (competency questions) and important concepts (key notions) are identified and further developed, axiomatized, and serialized. Furthermore, MOMo is a pattern-centric methodology, whereby the use of patterns provides a super-structure for an ontology to follow. *Bottom-up* refers to the emphasis on the data and datasets that are available and can be acquired. Then, by identifying the appearance of important concepts (key notions), additional patterns can be identified.

However, because this fast-paced application-

oriented effort involved a significant number of people (about 50 students, researchers, software engineers, and domain experts, including non-governmental organization [NGO] and industry representatives), a pure MOMo approach was not entirely feasible. A pure application would have required a more detailed up-front understanding of the major data sources and the application scenarios and needs; these were not realizable within the project timelines, as well as both the rapidly changing data landscape and availability, and the initial knowledge of end-user needs. In summary, we did not have a stable understanding of the requirements (inhibiting the top-down approach) and we did not have a stable understanding of the available data or what exactly those data comprised (inhibiting the bottom-up approach).

As a result, we agreed on two critical aspects: we would use the DGG as our geospatial backbone and utilize a kernel of the SOSA/SSN ontology (the kernel is discussed in Section 4.2 and SOSA/SSN and our implementation in Section 4.4). Furthermore, it necessitated that we conduct steps 3-5 of the methodology (as listed in the next section) in parallel, by separate teams, and with repeated iterations to ensure convergence.⁶ Indeed, in the beginning stages, we did multiple drafts of both which pieces of the SOSA/SSN ontology we expected to reuse and how they would be used as methods of

⁶In this case, we take convergence to mean either that our conceptualizations of important terms (i.e., phenomena) were consistent, or that we were using the same implementations for quantities and time, for example.

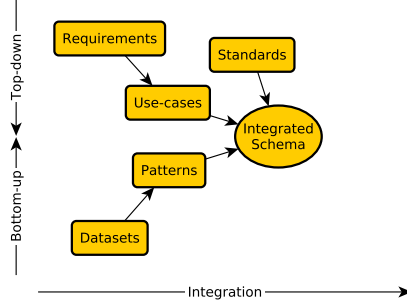


Figure 2: This shows a graphical view to the bottom-up and top-down components of the ontology development methodology utilized for KWG.

integration. Key to the eventual convergence were overlapping members between teams, which helped propagate a shared conceptualization, willingness to retrace and redo previous work, and a commitment to our initial minimal requirements.

3.1. The Steps

In brief, we provide this adapted MOMo approach in nine steps, listed as follows with accompanying commentary. Figure 2 provides a complementary, graphical view of the approach.

3.1.1. Define the use case.

The first step of MOMo has, in general, three stages, 1) to understand the available data, datasets, and provisioning institutions, as well as the originating schemas (if any); 2) to identify possible vocabularies, ontologies, ontology design pattern repositories, and standards; and 3) to work with domain experts, knowledge engineers, and stakeholders to gain an understanding of the myriad — possibly competing — perspectives on the use case. A particular knowledge elicitation methodology is not recommended here, merely ascribing what the outcome of such methodology should be. An additional complication results from the fact that KnowWhereGraph has multiple use cases, which fall under its overarching role as a geoinformational integration service (i.e., the area briefings). Understanding how the use cases both intersected and could support each other was a continuous process. We provide an example use case in Section 5.1.

3.1.2. Develop competency questions.

This step focuses on the interaction aspects of a traditional knowledge elicitation step. Essentially, this step aims to narrow the scope of the use case by describing, explicitly, the question-answering needs,

and thus the expected responses one should receive from a fully materialized knowledge graph. A competency question takes the form “How many wildfires had smoke plumes that impacted leafy greens in California in 2019?” This demonstrates which datasets should be related and thus accessible from queries against the graph, thus guiding both design and validation. Additional examples are provided in Section 5.2.

3.1.3. Identify key notions.

Generally, *key notions* are the concepts that are central to the knowledge graph. For example, this might mean that either they have high connectivity between other concepts, or they may appear in a high proportion of the competency questions. However, in the case of KWG, the key notions appear at two different levels of abstraction, resulting in the integration of the top-down and bottom-up approaches. At the outset, we knew key notions would exist for the entire KnowWhereGraph — which we call the KWG Core (i.e., Hazards — events that happen and have the potential to impact people, Regions — places that are important to people, and S2Cell — which abstracts our geospatial backbone). We provide more details in Section 4.1. Key notions also exist on the level of datasets, and, furthermore, we wanted to use SOSA/SSN [13], to unify how the datasets that describe Phenomena would be ontologized.⁷ Essentially, this meant that these first three steps would need to happen in conjunction, while the rest occurred in sequence (although iteratively). New patterns — and modules — needed to be developed to describe the KWG Core concepts, a reusable pattern needed to be extracted from SOSA/SSN, and the datasets needed to be ontologized. This resulted in a clear process for rapidly integrating newly encountered data, leveraging a powerful and robust W3C recommendation, and reusing a well understood pattern.

3.1.4. Match ODPs to key notions.

Ontology design patterns [6] can be considered to be tiny ontologies that solve domain-invariant modeling problems in (purposefully) abstract ways — these are sometimes called *conceptual components*; their purpose is to be adapted to specific use cases. This step matches the conceptual component of a pattern to the key notion in the use case. These patterns can be sourced from resources such as the ODP portal [19] or MODL [24]. In some cases these pattern repositories may not be sufficient, and

⁷By ontologized, we mean the act of integrating a dataset by providing an appropriate ontological description.

it may be necessary to develop new patterns. For example, we needed to develop novel patterns for representing our DGG, causal relations between events, and the alignment of terms between taxonomies. We discuss these in more detail in Section 4.3.

3.1.5. *Instantiate the patterns to create modules.*

This step is the process by which a general or abstract pattern is taken and modified to fit the use case. The result of these modifications, we call a module. This process is called *template-based instantiation* [11], as it differs from other pattern-adaptation strategies insofar that the pattern is not treated as a top-level ontology and its classes subsequently subclassed. Instead, the structure of the pattern is reused. Intuitively, we can consider the pattern to be a sort of “fill in the blanks.” This strategy is extremely effective in the case of KWG, as we have multiple datasets that describe physical phenomena and their relation to space and time. By reusing the same pattern for these concepts, we already achieve a baseline conceptual compatibility and interoperability.

3.1.6. *Systematically axiomatize each module.*

This is the step where most ontological analysis occurs. In general, the MOMo approach emphasizes the use of *schema diagrams* to communicate the connectivity of data (i.e., it purposefully obscures the *exact* underlying formalization) until this step, where domain experts and knowledge engineers both dig into the specifics of the relations. However, in many cases, since we have chosen a pattern-centric methodology, we could reuse the axiomatization from the pattern. We provide an excerpted example of the axiomatization of a dataset in Section 5.3.

3.1.7. *Assemble the modules.*

This is a quick step that involves connecting the different modules together. Additionally, we sought to add any additional axioms or relations that would span multiple modules. For example, in KWG, we have explored including causal relations between the events and phenomena of the different datasets.

3.1.8. *Review the final product.*

This step generally pertains to ensuring the naming consistency, catching any possible URI minting errors, and checking the consistency (in a logical sense) of the assembled modules.

3.1.9. *Produce artifacts.*

This step is the serialization of the ontology (i.e., the assembled modules), generation of the docu-

mentation, and polishing of the schema diagrams. Please see Section 2.3 for the locations of these artifacts.

4. The KnowWhereGraph Ontology

The KnowWhereGraph (KWG) Ontology integrates over thirty datasets across multiple spatial and temporal resolutions and thematic dimensions, which are shown in the Appendix (Tables 1 and 2). By *thematic*, we mean datasets that describe physical phenomena and their relation to a place and time, as well as the measured or observed characteristics of a place or the populations within them (i.e., a county’s adult obesity rate). This may be a natural hazard, forecasts, or man made features (e.g., highways). On the other hand, *place-centric* datasets describe human-meaningful places, such as the boundaries of sociopolitical regions. Due to the size of the ontology (there are 312 classes and 2980 axioms), it is less convenient to discuss exhaustively each class and their respective axioms. The purpose of this paper is to provide a brief, but thorough holistic overview of the KWG ontology.

To achieve this, we **(a)** do not explicitly examine our formalization, instead providing pointers to the formalization of each constituent piece; **(b)** make use of schema diagrams to convey generalized relationships between classes in a simplified visual format⁸; and **(c)** differentiate between our *conceptualization* and *implementation* of the ontology, as a way to specifically address design decisions independent of any reused ontologies.

The conceptual layer is discussed in sections 4.1 and 4.2: the KWG Core, a set of three tightly related classes, and a reusable kernel whose structure was extracted from (and still eventually implemented using) the SOSA/SSN ontology. These are shown in the top and bottom, respectively, in Figure 3. The implementation of the KWG ontology (i.e., how we materialized these core components) is discussed in Section 4.4. In Section 4.3, we briefly refer to the ontology design patterns utilized within the KWG Ontology. Section 4.4 discusses at a high level the implementations of different components of the KWG Ontology, including metadata, observations, quantities, and time. Finally, Section 4.5 provides pointers to tangential, yet important, standalone ontologies, which were developed to aid KWG in achieving its objectives.

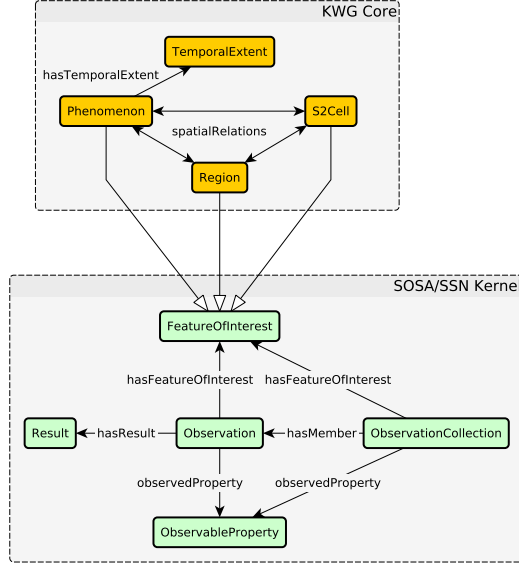


Figure 3: This figure shows the KWG kernel, which consists of three classes and their spatial relations (taken from the KWG Ontology), combined with the SOSA/SSN kernel. The boxes indicate classes. The grey dashed boxes are for visual partitioning of the kernels. Filled arrows indicate object properties. Open arrows indicate subclass relations. The *spatialRelations* label is a placeholder for the KWG Ontology’s implementation of the Simple Functions relation ontology.

4.1. The KnowWhereGraph Core Structure

The core of the KWG ontology is actually quite simple, as it serves a straightforward purpose: “hazards need to be linked to the places that they impact.” These links are served in two ways: direct integration and alignment to the underlying DGG. Thus arise our three main classes: Hazard, Region, and Cell.

The Hazard Class is a specific form of physical phenomena that primarily interests (at the time of this writing) KWG. We consider a hazard to be a physical phenomenon that can (generally negatively) impact civilization in some way and that is anchored in space and time. That is, it will have its own underlying geometry, which is integrated with both the individual cells of the DGG and the underlying geometries (when available) for human-meaningful regions. The Hazard is then temporally scoped in a manner appropriate to the originating dataset. Hazards and their effects are described via observations; more detail is provided in the next section. The instances of this class (and its subclasses) are generally populated from the thematic datasets, as shown in Table 1.

The Region Class is the conceptualization of a

human-meaningful place for the KWG Ontology. Crucially, they may, or may not, have an explicit geometry. The instances of this class are generally populated from the place-centric Datasets, as shown in Table 2. A Region might have socio-political boundaries (e.g., a US census tract or county) or have natural boundaries (e.g., a river). Regions are aligned to the underlying DGG, providing a backbone to see how arbitrary Hazards and Regions interact without explicitly performing the spatial integration between the two datasets.

The Cell class is purposefully generalized and intended to be organized into a DGG. We specifically use the S2Cell class, which is axiomatically specified as *hierarchical*, via different spatial relations (e.g., RCC-8 [4]). This general class is included in case we wish to support additional DGGs in the future. The purpose of the DGG is to provide a level of interoperability beyond our own knowledge graph. Any phenomenological dataset with a geospatial component can thus be aligned to the S2Cell DGG and immediately achieve interoperability and spatial integration with KWG and its ontology.

As a Cell itself does not rightly exist (e.g., the spatial geometry of the Cell is not human-meaningful), but is merely an arbitrary human convenience, any Cell points to a serialized geometry (more detail in Section 4.4.1). Regions and Hazards then overlap, touch, or contain (among the other RCC-8 relations), with the S2Cells.

4.2. A Reusable SOSA/SSN Kernel

There are two types of integration that KWG supports. The KWG Core classes (from the previous section) integrate across broader notions of space, time, phenomena, and their alignment to some DGG. The second method is through a consistent representation of *how* the different hazards and other geospatial phenomena impact these places (e.g., through storm tracks) or what characteristics can be measured or observed about these places (e.g., demographic statistics). To accomplish this, we extracted a kernel structure from the SOSA/SSN ontology [13, 14, 10]. Eventually this is implemented directly from the classes of the SOSA/SSN ontology (Sec. 4.4.3), GeoSPARQL (Sec. 4.4.1), and QUDT (Sec. 4.4.5).

We chose this route, in large part, to simplify the process of data integration, as well as eventually leverage W3C and Open Geospatial Consortium (OGC) recommendations and standards, but not utilize every aspect of the more comprehensive ontologies. Furthermore, by using this structure, and

⁸Note our syntax and motivation in Section 2.2.

in particular the tripartite purposeful disconnect between an observation, the target of the observation (i.e., the feature of interest) and the observed property, allows for quick semantic harmonization between terms. For example, KWG has multiple datasets that describe wildfires and their properties. Yet, not all properties are present across all datasets. However, it is still easy and straightforward to receive all observations about a particular fire without knowing exactly from which dataset the fire originated.

This structure thus forms the underlying semantic backbone of the KWG and for the ontology as a whole; it is shown in the bottom of Figure 3 (the green boxes).

4.3. Utilized Ontology Design Patterns

During the development of the KWG ontology, it was necessary both to create new patterns and to adapt existing patterns. So far, four new patterns have been developed: the hierarchical features ODP [27], the causal relations ODP [28], the taxonomy alignment ODP [29], and the computational observation ODP [26]. Existing patterns were selected from MODL [24]: entity with provenance ODP and the agent role ODP.

4.4. Implementation

We have developed several standalone ontologies and resources, and reuse quite a few well-known, standardized, or W3C recommended vocabularies, taxonomies, and ontologies. This helps to enhance interoperability with other knowledge graphs, to maintain consistency in our data model, and to leverage existing tools that support these vocabularies.

4.4.1. Modeling and Representing Space

We used GeoSPARQL[17, 1], an OGC standard, to represent our geospatial data in RDF. It, in turn, reuses the OGC Simple Features (SF) standard, which defines a set of geometric primitives (e.g., points or polygons) and their spatial relationships. Within the KWG Ontology, we represent any discrete geographic feature type (Hazard, Region and their subclasses) that has a spatial extent as a subclass of GeoSPARQL's `geo:SpatialFeature` class. We also use the spatial relationships from GeoSPARQL (based on DE-9IM spatial relations) to establish pre-computed spatial relationships between any two spatial features. Several graph databases support GeoSPARQL, including GraphDB [9], which we use for KWG.

4.4.2. Modeling Time

The Time Ontology is (re)used for all representations of time within the KWG Ontology. The super-property for most time-related conceptualizations is `kwg-ont:hasTemporalScope`, which effectively has a range of any `TemporalEntity` from `OWL:Time` and is most often used with `Hazard`. However, we also specify `TemporalEntity` to be the range of `resultTime` and `phenomenonTime` for `Observations`. For serialization of the actual time-related data, we reuse the XML schema datatypes (e.g., `xsd:dateTime`).

4.4.3. Modeling Observations

As mentioned previously, SOSA/SSN was reused to implement the `Observation` and related classes within the KWG ontology. In particular, we model observations made by sensors that detect physical properties (e.g., climate, weather, hazard properties) and chemical properties (e.g., air pollutants), and information about demographics and public health (e.g., disease statistics, socioeconomic indicators such as poverty and social vulnerability). We have accordingly refined the interpretation of two concepts: `sosa:FeatureOfInterest` and `sosa:Observation`.

The first, `sosa:FeatureOfInterest`, is the main class for representing FOIs (objects and phenomena that are being observed or measured) in the ontology. FOIs can be physical entities or conceptual entities, such as a hazard event, an administrative region boundary, or an abstract concept like S2 cell geometry. The second, `sosa:Observation`, is a class to represent any measurement of a property of a FOI that is made by a sensor. `ObservationCollection` denotes a group of related observations that are made by one or more sensors over a period of time, and share common characteristics such as the observed property or time of observation, and so on. The `sosa:hasMember` property denotes the set of observations that belong to the collection. An observation or observation collection has a number of properties that describe its characteristics, such as `sosa:observedProperty`, which indicates the property or phenomenon being observed; `sosa:madeBySensor` indicates the sensor that made the observation; `sosa:resultTime` indicates the time at which the observation was made; `sosa:hasResult` indicates the value of the observation; and `hasFeatureOfInterest` indicates the entity being observed.

In KWG, a `sosa:FeatureOfInterest` represents both (i) *the thing whose property can be estimated or calculated* and (ii) *anything that can have a spatial representation and an associated geometry* [17, 1].

In this sense, even though upper ontologies classify events and things/objects/features as distinct classes based on the *endurant-perdurant* categorization [18], the fact that a hazard satisfies both (i) and (ii) allows us to identify the *HazardEvent* class both as a *sosa:FeatureOfInterest* and a *geo:Feature*.

Observations (and their collections) in SOSA are defined as *estimates or calculations of the value of a property made using a sensor (device, agent, or software)*, while a feature of interest is a *thing whose property is being estimated or calculated in the course of an observation to arrive at a result*. Although measurements such as census data from community surveys and soil-type maps constructed from surveys conducted over the course of a century do not precisely satisfy this definition, it seems plausible for us to model them using the *sosa:Observation|ObservationCollection* classes. In this case the features of interest are either some spatial region (*Region*) such as an administrative region or a region that shares a common property (e.g., soil type), or a spatial grid (*S2Cell*).

4.4.4. Modeling the KWG Metadata

As KWG and its ontology grew, it became very difficult to understand the contents of the graph because its size engendered project management gaps (i.e., overlapping graph development efforts due to similarity in phenomena or diverging sources of datasets as new, better, or less stale sources were identified). Furthermore, we wished to ascribe provenance to the constituent data, but avoiding so-called triple explosion.⁹ This strategy links observable properties to the datasets from which they were derived — and the organizations that originally provided them — as opposed to a per assertional statement ascription of provenance. While this slightly increased query complexity (i.e., the number of statements between the original provenance and any particular observation is increased), this slight increase in complexity actually created a much clearer top-level view of the graph and what it contains. This strategy is distinct from creating named graph, as we are not directly encapsulating any triples into a distinct graph. Instead, we are explicitly grouping the *ObservableProperty*s and *FeatureOfInterests*, which can be then be used to extract a subgraph, as necessary.

⁹Triple explosion is what we call the process of compact (tabular) data becoming a disproportionate number of triples (i.e., subject-predicate-object statements). In this case, we wished to avoid every *Observation* in our graph carrying detailed provenance, when identical data would be repeated for every observation from a dataset.

The metadata model distinguishes between two different types of subgraphs that appear in the KWG: *dataset* and *thematic*. A dataset subgraph is straightforward: it is merely an encapsulation of a particular dataset as materialized against the KWG Ontology. A thematic subgraph is a set of dataset subgraphs that are unified according to some theme — which for now is merely modeled as a *xsd:string*. Our thematic examples largely pertain to our pilot use cases: land valuation (which incorporates the dataset subgraphs for smoke plumes, smoke plume forecasts, wildfires, and logistics) and disaster response (which incorporates subgraphs for hurricane impact predictions, FEMA disaster declarations, and public health indicators). Every dataset subgraph is spatially integrated against our different regional representations.

Additionally, this metadata model shored up project management of large graphs by allowing for a representation of the KnowWhereGraph team itself, and it indicates who was in charge of developing particular subgraphs.

We reused several vocabularies to annotate the KWG Ontology. For instance we used annotation properties from Dublin Core Metadata Initiative (DCMI) Metadata Terms [32] to describe the title, description, rights, license, date created and creator. We used Friend of a Friend (FOAF¹⁰) to describe the development team and their roles. The Simple Knowledge Organization System (SKOS) was used to annotate definitions, examples, and the taxonomic structure between domain concepts. Finally, we used PROV-O [22] to connect the subgraphs to the datasets from which they were derived.

4.4.5. Modeling Quantities

The QUDT (Quantities, Units, Dimensions and Data Types) ontology [21] is used for representing climate measurement data (such as temperature, Palmer drought severity index, cooling degree days) and their corresponding units of measure. Specific climate quantity types (such as mean or value) are denoted using the *kwg-ont:Quantity* class, a subclass of *kwg-ont:QuantityValue*. Measured values and corresponding data properties are then captured using data properties *qudt-unit:unit* and *qudt-unit:numericValue*.

4.5. The KnowWhereGraph Ontology Network

While developing the KWG Ontology, we identified the need for integrating, harmonizing, and reasoning over specific domain data in the graph, and

¹⁰<http://xmlns.com/foaf/0.1/>

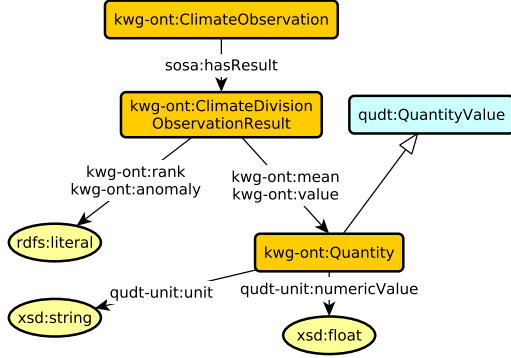


Figure 4: This schema diagram shows how QUDT is used for representing measured quantities, their values and units of measurement within KWG. It uses the graphical syntax described in Section 2.2.

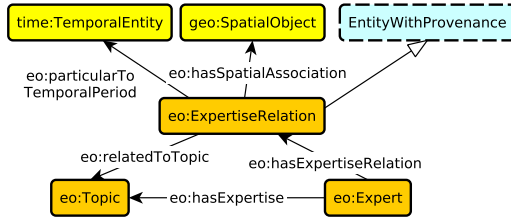


Figure 5: Simplified schema diagram of the Expertise Ontology

also for consistently representing these data following standard definitions and best practices. Thus, two standalone ontologies were developed that are imported by the KWG Ontology, yet are not interdependent (i.e., other ontologies may play the role¹¹). We provide brief pointers for reference and documentation on how they assist the KWG Ontology in meeting the requirements of the overall graph.

4.5.1. The Expertise Ontology

KWG contains information on agents who are experts on topics related to specific disaster types, disaster management activities, named disasters, and public health. The Expertise Ontology¹² (EO) was developed to represent all varied expertise-related information consistently. At a high level, EO consists of a core set of classes and properties to 1) model experts (people and groups), topics and their relations, 2) represent hierarchical relations between topics of different levels of granularity, and 3) connect topics with relevant content in a KG. EO facilitates representing not only research-based and

¹¹These ontologies were developed since we were not able to find sufficient other resources to use without doing so.

¹²<https://github.com/KnowWhereGraph/expertise-ontology>

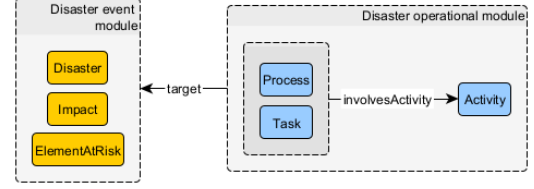


Figure 6: High-level schema diagram of the DMDO. Orange boxes indicate classes from the *Hazard Event* module. Blue boxes indicate classes from the *Hazard Operational* module.

theory-based expertise, but also experience-based expertise. To do so, EO models the activities that an expert may have engaged in, or their role and affiliation within an organization, and scopes these spatially and temporally.

4.5.2. The Disaster Management Domain Ontology

KWG contains at least 11 hazard datasets and at least one hazard (Fire) characterized by data from four different sources (see Table 1). To model their semantics and enable integration using existing ontologies, we are developing the Disaster Management Domain Ontology (DMDO), which will provide a framework both to align diverse hazard types, formats of data, and domain vocabularies consistently within KWG, and to clarify better situational awareness of the spatiotemporal interactions of similar events. The ontology disambiguates hazards from disasters and their impacts, but also distinguishes spatiotemporal events from their observations. DMDO also formalizes the United Nations Office for Disaster Risk Reduction (UNDRR) hazard classification [31] using the taxonomy alignment ODP [29]. Fig 6 shows some of the core concepts in each module and their connections.

5. Intended Usage & Example

In this section, we introduce one of our core use-case scenarios. In brief, a use-case scenario is a narrative coupled with a set of competency questions. Together, these describe an expected set of interactions between an end user and the KWG. Then, we provide a glimpse at what data — materialized against the KWG Ontology — look like, and how they can be served to answer those competency questions.

5.1. Use Case Scenario

One context in which we envision KWG to be used is humanitarian response. To explore this, we part-

nered with Direct Relief,¹³ a non-profit organization that responds to disasters worldwide by delivering urgently needed medical supplies. The organization’s effectiveness depends largely on their ability to quickly and efficiently provide resources that are appropriate for the situation, often in the face of logistical challenges like interrupted communication channels and supply chains. Among their primary tasks following a disaster are to (a) identify local authorities and experts who can advise on immediate medical needs; and (b) assess which supplies are likely in demand — and where — based on the nature of the disaster and demographic and health characteristics of the affected populations. As often unpredictable and complex events, disasters challenge humanitarian responders to comprehend and address many issues simultaneously and immediately. From the operational state of a region’s hospitals and health clinics to the prevalence of poverty and chronic health disorders within a community, information on a variety of themes can contribute to a responder’s situational awareness and ability make informed decisions.

5.2. Selected Competency Questions

Typical questions following a hurricane’s landfall might include,

- CQ1. *Who is most vulnerable?*
- CQ2. *What are the critical health risks faced by these populations?*
- CQ3. *Who has expertise relevant to this situation and place, and how do we contact them?*
- CQ4. *How can we tailor our response to respect the historical and cultural sensitivities of affected populations?*

5.3. Worked Example

For a brief example, we will focus on CQ1 from the previous section. Figure 7 shows how this question would be answered via a geo-enrichment service built on top of KWG. At the basic level, we see the S2 DGG (rendered at level 13, which means about 1.27 square kilometers per grid cell) combined with Administrative Region boundaries at Level 3 (corresponding to counties, parishes, or boroughs) in the Baton Rouge, Louisiana, USA region. The yellow-to-red gradient indicates the values of the Social Vulnerability Index, provided by the CDC¹⁶ and integrated into KWG. The grey shaded area and green line indicate the areal extent of impact

and path of Hurricane Ida, respectively. Altogether, these provide a quick, intuitive overview of the *a priori* state of the disaster. That is, when available data are scarce known shortly after landfall, can educated guesses be made regarding where aid will be needed first? Thus, this gives a holistic answer to CQ1.

Figure 8 gives a fragment of the abbreviated query that would replicate this view.

6. Related Work

At the time of this writing, we are not aware of an ontology that performs hazard (or geospatial phenomenological) data and regional identifier integration over a common geospatial backbone (i.e., the S2 DGG) while maintaining a rich description of the phenomena, their impacts, provenance, and lineage. Of course, we are aware of several ontologies that can be leveraged to accomplish this (and we have described many of them in Section 4.4). Briefly, we discuss other ontologies we encountered but did not use, as well as graphs that fill a similar niche to KWG.

6.1. Environmental Concepts

The Environment Ontology (ENVO) [3] was determined to have limitations that impacted its usability and effectiveness within the context of KWG’s use cases. For instance, while ENVO covers a wide range of environmental concepts, it has limited coverage of human-related environmental concepts, such as environmental pollution. Some definitions in ENVO are ambiguous or imprecise, which led to consistent misinterpretation. For example, the term ‘flood’ as a ‘continuant’ versus ‘flooding’ as an ‘occurrent’ made it somewhat ambiguous for us to categorize a flood as reported by NOAA. Overall, the usage of an upper ontology and its principles made it difficult to adequately perform the necessary integration over significantly heterogeneous datasets and sources. Due to this strong ontological commitment, it was determined that trying to extend ENVO in a new direction was outside of the scope effort for both KWG and ENVO.

SWEET (Semantic Web for Earth and Environmental Terminology) [20] has several overlapping definitions and inconsistent use of relations. For example, ‘phenomenon’ and ‘observable property’ refer to measurable or observable characteristics of the natural world in their definitions. Yet another example is where ‘part of’ and ‘has part’ are used

¹³<https://www.directrelief.org/>

¹⁴This visualization is powered by <https://felt.com/>.

¹⁶<https://www.atsdr.cdc.gov/placeandhealth/svi/index.html>

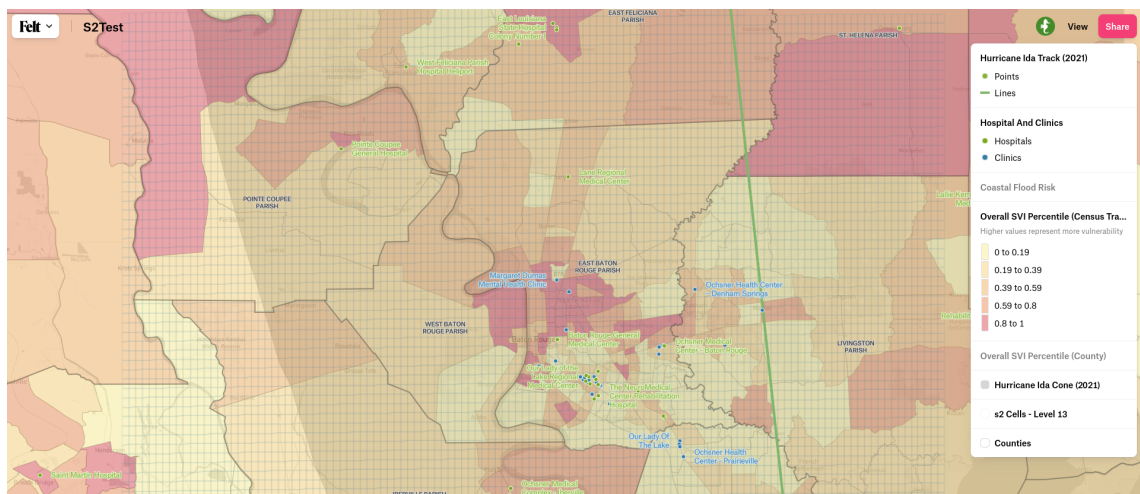


Figure 7: This figure displays a visualization of the geo-enriched data that KWG provides. The yellow-to-red gradient depicts Social Vulnerability Index scores by census tract. Superimposed on this are S2 DGG cells and boundaries of the counties (parishes) of the Baton Rouge, LA, USA area. Finally, the areal extent of Hurricane Ida's impacts is projected around the track across the landscape (the grey shading and green line, respectively). This provides a quick and intuitive grasp of where attention might be most valuable when no other information is available (e.g., early in the disaster-response scenario).¹⁴

```
SELECT * WHERE {
  ?cell a kwg-ont:S2Cell .
  ?county a kwg-ont:AdminRegion_3 ;
    geo:sfWithin
      kwr:Earth.NA.US.USA.19_1 .
  ?cell kwg-ont:sfWithin ?county .
  ?county sosa:isFeatureOfInterestOf ?obs .
  ?obs a kwg-ont:VulnerabilityObservation .
  ?obs sosa:hasSimpleResult ?result .
}
```

Figure 8: This query represents a portion of Figure 7. The exact query can be found online,¹⁵ and we note that the full replication would be best served by combining multiple queries and overlaying them in an external tool. This particular component retrieves all of the Social Vulnerability Index values for all counties in Louisiana, USA, as well as all S2 cells that compose the counties.

interchangeably to describe hierarchical relationships between concepts. Moreover these vocabularies are large and complex, with many concepts and relationships, which made it challenging to understand and use them effectively.

6.2. Similar Resources

The Google Data Commons¹⁷ integrates many public datasets and, indeed, draws inspiration from the Google Knowledge Panels and (humbly) our own interfaces. The data integration occurs through Schema.org annotations. Thus, the composite schema is quite shallow. Additionally the graph itself is not made public, but can be accessed through a variety of rate-limited APIs.

¹⁷<https://www.datacommons.org/>

Esri's ArcGIS Knowledge is an enterprise knowledge graph specifically built to be compatible with their GeoEnrichment platform, ArcGIS Knowledge Server. It integrates and makes accessible through the platform a set of curated data, as well as common publicly available or governmental datasets. We are not aware of an associated ontology with their graph structure. Additionally, this resource was released during the development of KWG and its ontology.

7. Conclusion

The KnowWhereGraph is a complex project with multiple evolving use cases, a large team, and an ambitious goal. To create the schema for the underlying knowledge graph — the KnowWhereGraph Ontology — it was necessary to choose a methodology that produced a schema where emphasizing sustainability and adaptability were prioritized. To this effect, we chose — and adapted — the Modular Ontology Modeling methodology, to produce a pattern-based (modular) ontology that could quickly incorporate new data sources, and be easily maintained beyond the project lifetime. Additionally, it was executed in a distributed fashion, resulting in a quickly changing schema, but has since converged.

As a result, we have developed the KnowWhereGraph Ontology and reported on its kernel structure (i.e., the top-level, repeated structure) of our knowledge graph, the ontology design patterns we have utilized, and a description of the metadata model

used to manage and record the provenance of our graph.

Future Work

Over the course of developing the KnowWhere-Graph Ontology, we have identified many opportunities for next steps, as follows.

- Develop tighter integration with our developed patterns, in particular for the support of identifying causality between events.
- Extract additional patterns that would be useful in the development of other spatially enabled knowledge graphs, such as a simple pattern for representing entities with geometries or a pattern for representing observations.

Acknowledgement. This work was funded 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.

References

- [1] R. Battle and D. Kolas. Enabling the geospatial semantic web with parliament and geosparql. *Semantic Web*, 3(4):355–370, 2012.
- [2] B. Bondaruk, S. A. Roberts, and C. Robertson. Assessing the state of the art in discrete global grid systems: Ogc criteria and present functionality. *Geomatica*, 74(1):9–30, 2020.
- [3] P. L. Buttigieg, N. Morrison, B. Smith, C. J. Mungall, S. E. Lewis, and E. Consortium. The environment ontology: contextualising biological and biomedical entities. *Journal of biomedical semantics*, 4:1–9, 2013.
- [4] A. G. Cohn, B. Bennett, J. Gooday, and N. M. Gotts. Qualitative spatial representation and reasoning with the region connection calculus. *GeoInformatica*, 1(3):275–316, 1997.
- [5] S. Cox and C. Little, editors. *Time Ontology in OWL*. W3C Recommendation 19 October 2017, 2017. Available from <https://www.w3.org/TR/2017/REC-owl-time-20171019/>.
- [6] A. Gangemi and V. Presutti. Ontology design patterns. In S. Staab and R. Studer, editors, *Handbook on Ontologies*, International Handbooks on Information Systems, pages 221–243. Springer, 2009.
- [7] D. Garijo. WIDOCO: A wizard for documenting ontologies. In C. d’Amato, M. Fernández, V. A. M. Tamma, F. Lécué, P. Cudré-Mauroux, J. F. Sequeda, C. Lange, and J. Heflin, editors, *The Semantic Web - ISWC 2017 - 16th International Semantic Web Conference, Vienna, Austria, October 21-25, 2017, Proceedings, Part II*, volume 10588 of *Lecture Notes in Computer Science*, pages 94–102. Springer, 2017.
- [8] M. F. Goodchild. Discrete global grids for digital earth. In *International Conference on Discrete Global Grids*, pages 26–28. Citeseer, 2000.
- [9] GraphDB. <http://graphdb.ontotext.com/>.
- [10] A. Haller, K. Janowicz, S. J. D. Cox, M. Lefrançois, K. Taylor, D. L. Phuoc, J. Lieberman, R. Garcia-Castro, R. Atkinson, and C. Stadler. The modular SSN ontology: A joint W3C and OGC standard specifying the semantics of sensors, observations, sampling, and actuation. *Semantic Web*, 10(1):9–32, 2019.
- [11] K. Hammar and V. Presutti. Template-based content ODP instantiation. In K. Hammar, P. Hitzler, A. Krisnadhi, A. Lawrynowicz, A. G. Nuzzolese, and M. Solanki, editors, *Advances in Ontology Design and Patterns [revised and extended versions of the papers presented at the 7th edition of the Workshop on Ontology and Semantic Web Patterns, WOP@ISWC 2016, Kobe, Japan, 18th October 2016]*, volume 32 of *Studies on the Semantic Web*, pages 1–13. IOS Press, 2016.
- [12] P. Hitzler and C. Shimizu. Modular ontologies as a bridge between human conceptualization and data. In P. Chapman, D. Endres, and N. Pernelle, editors, *Graph-Based Representation and Reasoning - 23rd International Conference on Conceptual Structures, ICCS 2018, Edinburgh, UK, June 20-22, 2018, Proceedings*, volume 10872 of *Lecture Notes in Computer Science*, pages 3–6. Springer, 2018.
- [13] K. Janowicz, A. Haller, S. Cox, M. Lefrançois, D. L. Phuoc, and K. Taylor. Semantic sensor network ontology. W3C recommendation, W3C, Oct. 2017. <https://www.w3.org/TR/2017/REC-vocab-ssn-20171019/>.
- [14] K. Janowicz, A. Haller, S. J. D. Cox, D. L. Phuoc, and M. Lefrançois. SOSA: A lightweight ontology for sensors, observations, samples, and actuators. *J. Web Semant.*, 56:1–10, 2019.
- [15] K. Janowicz, 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. G. Estrecha, B. D. Mecum, A. Lopez-Carr, A. Schroeder, D. Smith, D. J. Wright, S. Wang, Y. Tian, Z. Liu, M. Shi, A. D’Onofrio, Z. Gu, and K. Currier. Know, know where, knowwhere-graph: A densely connected, cross-domain knowledge graph and geo-enrichment service stack for applications in environmental intelligence. *AI Mag.*, 43(1):30–39, 2022.
- [16] K. Janowicz, C. Shimizu, P. Hitzler, G. Mai, S. Stephen, R. Zhu, L. Cai, L. Zhou, M. Schildhauer, Z. Liu, Z. Wang, and M. Shi. Diverse data! diverse schemata? *Semantic Web*, 13(1):1–3, 2022.
- [17] F. Knibbe, J. Herring, J. Abhayaratna, M. Bonduel, M. Perry, N. J. Car, S. J. D. Cox, and T. Homburg. Geosparql ontology. W3C recommendation, W3C, Oct. 2021. <https://opengeospatial.github.io/ogc-geosparql/geosparql11/index.html>.
- [18] V. Mascardi, V. Cordì, and P. Rosso. A comparison of upper ontologies. In *Woa*, volume 2007, pages 55–64, 2007.
- [19] Ontology design patterns. [org \(odp\). https://ontologydesignpatterns.org/](https://ontologydesignpatterns.org/).
- [20] R. G. Raskin and M. J. Pan. Knowledge representation in the semantic web for earth and environmental terminology (sweet). *Computers & geosciences*, 31(9):1119–1125, 2005.
- [21] H. Rijgersberg, M. Van Assem, and J. Top. Ontology of units of measure and related concepts. *Semantic Web*, 4(1):3–13, 2013.
- [22] S. Sahoo, D. McGuinness, and T. Lebo. PROV-o: The PROV ontology. W3C recommendation, W3C, Apr. 2013. <http://www.w3.org/TR/2013/REC-prov-o-20130430/>.
- [23] C. Shimizu, K. Hammar, and P. Hitzler. Modular ontology modeling. *Semantic Web*, 14(3):459–489, 2023.
- [24] C. Shimizu, Q. Hirt, and P. Hitzler. MODL: A modular ontology design library. In K. Janowicz, A. A. Krisnadhi, M. P. Villalón, K. Hammar, and C. Shimizu, editors, *Proceedings of the 10th Workshop on Ontology Design and Patterns (WOP 2019) co-located with 18th International*

- Semantic Web Conference (ISWC 2019), Auckland, New Zealand, October 27, 2019*, volume 2459 of *CEUR Workshop Proceedings*, pages 47–58. CEUR-WS.org, 2019.
- [25] C. Shimizu and P. Hitzler. Automatically generating human readable documentation for ontology design patterns. In O. Seneviratne, C. Pesquita, J. Sequeda, and L. Etcheverry, editors, *Proceedings of the ISWC 2021 Posters, Demos and Industry Tracks: From Novel Ideas to Industrial Practice co-located with 20th International Semantic Web Conference (ISWC 2021), Virtual Conference, October 24-28, 2021*, volume 2980 of *CEUR Workshop Proceedings*. CEUR-WS.org, 2021.
 - [26] C. Shimizu, P. Hitzler, and C. F. V. II. A pattern for modeling computational observations. In V. Svátek, V. A. Carriero, M. Poveda-Villalón, C. Kindermann, and L. Zhou, editors, *Proceedings of the 13th Workshop on Ontology Design and Patterns (WOP 2022) co-located with the 21th International Semantic Web Conference (ISWC 2022), Online, October 24, 2022*, volume 3352 of *CEUR Workshop Proceedings*. CEUR-WS.org, 2022.
 - [27] C. Shimizu, R. Zhu, G. Mai, C. K. Fisher, L. Cai, M. Schildhauer, K. Janowicz, P. Hitzler, L. Zhou, and S. Stephen. A pattern for features on a hierarchical spatial grid. In *IJCKG'21: The 10th International Joint Conference on Knowledge Graphs, Virtual Event, Thailand, December 6 - 8, 2021*, pages 108–114. ACM, 2021.
 - [28] C. Shimizu, R. Zhu, M. Schildhauer, K. Janowicz, and P. Hitzler. 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) : online, October 24, 2021*, volume 3011, pages 38–50, 2021.
 - [29] S. Stephen, C. Shimizu, M. Schildhauer, R. Zhu, K. Janowicz, and P. Hitzler. A pattern for representing scientific taxonomies. In V. Svátek, V. A. Carriero, M. Poveda-Villalón, C. Kindermann, and L. Zhou, editors, *Proceedings of the 13th Workshop on Ontology Design and Patterns (WOP 2022) co-located with the 21th International Semantic Web Conference (ISWC 2022), Online, October 24, 2022*, volume 3352 of *CEUR Workshop Proceedings*. CEUR-WS.org, 2022.
 - [30] The KnowWhereGraph schema. <https://stko-kwg.geog.ucsb.edu/lod/ontology>.
 - [31] Hazard definition and classification review. <https://www.undrr.org/publication/hazard-definition-and-classification-review>, 2021.
 - [32] S. Weibel, J. Kunze, C. Lagoze, and M. Wolf. Dublin core metadata for resource discovery. Technical report, OCLC, 1998.

A. Tables

Thematic Dataset	Source Agency	Example Attributes
Soil Properties	USDA	soil type, farmland class
Wildfires	USGS, USDA, USFS, NIFC	wildfire type, num acres burned
Earthquakes	USGS	magnitude
Climate Hazards	NOAA	casualties, property damage
Experts (Covid-19 Mobility)	DR	name, affiliation, expertise
Expert (General)	KWG, UCS, DR, SS	name, affiliation, expertise
Cropland Types	USDA	crop types (raster data)
Air Quality	EPA	air quality index
Smoke Plumes Forecasts	NOAA	daily smoke plume forecast
Climate	NOAA	temperature, precipitation
Disaster Declarations	FEMA	area, program, amount approved
Smoke Plume Extents	NOAA	smoke plume extent
BlueSky Forecasts	BlueSky	PM10, PM5
Highway Networks	DOT	road type, road length, signage
Public Health Observations	CDC, USCB, UW:PHI	poverty, diabetes, obesity
Public Health Infrastructure	HIFLD	pharmacies, hospitals
Social Vulnerability	CDC/ATSDR	social vulnerability index
Hurricane Tracks	NOAA	max wind speed, min pressure

Table 1: Thematic Datasets

Place-Centric Dataset	Defining Authority	Spatial Coverage
S2 cells	Google	Lvl 9 (Global), Lvl 13 (US)
Global Administrative Regions	GADM.org	Global
US Federal Judicial Districts	DoJ, Esri	US
National Weather Zones	NOAA	US
FIPS Codes	USCB	US
Designated Market Areas	Nielsen	US
ZIP Codes	USPS	US
Climate Divisions	NOAA	US
Census Metropolitan Area	USCB	US
Drought Zone	NDMC	US
GNIS	USGS	US

Table 2: This table shows the Place-centric Datasets which are integrated in the KWG.

ATSDR	Agency for Toxic Substances and Disease Registry
CDC	United States Center for Disease Control
DHS	United States Department of Homeland Security
DoJ	United States Department of Justice
DoT	United States Department of Transportation
DR	Direct Relief
EPA	Environmental Protection Agency
FEMA	Federal Emergency Management Agency
HIFLD	Homeland Infrastructure Foundation-Level Data
KWG	KnowWhereGraph
NDMC	National Drought Mitigation Center
NIFC	National Interagency Fire Center
NOAA	National Oceanographic and Atmospheric Administration
SS	Semantic Scholar
UCS	University of California System
UCSB	University of California, Santa Barbara
USCB	United States Census Bureau
USDA	United States Department of Agriculture
USGS	United States Geological Survey
USPS	United States Postal Service
UW:PHI	University of Wisconsin: Public Health Institute

Table 3: This table displays the abbreviations that we use within the paper, especially the tables.

kwg-ont:	http: //stko-kwg.geog.ucsb.edu/lod/ontology/
kwgr:	http: //stko-kwg.geog.ucsb.edu/lod/resource/
sosa:	http://www.w3.org/ns/sosa/

Table 4: This table displays the prefixes that we use within Section 5.1, and more broadly within our graph.