ENABLING KNOWLEDGE SHARING BY MANAGING DEPENDENCIES AND INTEROPERABILITY BETWEEN INTERLINKED SPATIAL KNOWLEDGE GRAPHS

N. McEachen[*], J. Lewis

TerraFrame, Inc. Lafayette, Colorado, USA – nmceachen@terraframe.com

KEY WORDS: Spatial Knowledge Graph, Data Mesh, Spatial Knowledge Mesh, Interoperability, Graph Repository, Ontology, Common Geo-Registry.

ABSTRACT:

Knowledge sharing is increasingly being recognized as necessary to address societal, economic, environmental, and public health challenges. This often requires collaboration between federal, local, and tribal governments along with the private sector, nonprofit organizations, and institutions of higher education. To achieve this, there needs to be a move away from data-centric to knowledge sharing architectures, such as a Geospatial Knowledge Infrastructure (GKI). Data from multiple organizations need to be properly contextualized in both space and time to support geographically based planning, decision making, cooperation and coordination. A spatial knowledge graph (SKG) is a useful paradigm for facilitating knowledge sharing and collaboration. However, interoperability between independently developed SKGs from different organizations that reference the same geographies is often not automated in a machine-readable way due to a lack of standardization. This paper outlines an architecture that automates interoperability and dependency management between SKGs as they are formally published by version and period of validity. We are calling this approach a spatial knowledge mesh (SKM), as it is a specialization of the data mesh architecture along with the concept of a common geo-registry to facilitate knowledge sharing more easily. The initial implementation, called GeoPrism Registry, is being developed as an open-source spatial knowledge infrastructure as a platform to help countries meet their NSDI and GKI objectives. It was fist funded and deployed to support ministries of health and is more recently being utilized in GeoPlatform.gov.

1. INTRODUCTION

1.1 The Transition from Data to Knowledge

There needs to be more institutional collaboration, coordination, interoperability, and integration across the various national data information systems and platforms (UN-GGIM, 2018). Geospatial data has become so ubiquitous that the ecosystem of stakeholders is no longer limited to the geospatial community. Static metadata records and data catalogs are not adequate and do not cater to the needs of non-geospatial experts, and thus limit the expansion of a geospatial ecosystem. Rather than providing just data with an intent to solve the strategic priorities of governments, future national geospatial information ecosystems must meet the needs of a new generation of geospatial users and enable use cases based on the demand of knowledge from citizens and businesses (UN-GGIM, 2022). These needs have evolved beyond simple data or digital data libraries to advanced knowledge-based solutions (KBS) and services, which can deduce new information from existing data. Cross-sectoral collaboration using machine-to-machine interfaces will help to share knowledge at scale. Sharing knowledge requires interoperability. Interoperability requires semantic standards.

Building a KBS requires integrating data silos, facilitating collaboration, simplifying spatial data use, and implementing semantic interoperability standards. A Geospatial Knowledge Infrastructure (GKI) can accomplish this when implemented using spatial knowledge graphs (SKG), which are ideal for knowledge representation and reasoning, because they model large networks of features, their semantic types, properties, and relationships between other features.

One of the key challenges is that existing geospatial data infrastructures are data-centric rather than knowledge-centric.

The world is transitioning from data to insights and knowledge sharing and future national geospatial ecosystems should help facilitate this transition. The United Nations Committee of Experts on Global Geospatial Information Management (UN-GGIM) recently recommended that governments progress beyond National Strategic Data Initiatives (NSDI) that are data-centric and move toward a GKI to meet strategic priorities of governments, citizens, and businesses: "Data is critical even today, however, it is no longer valuable as a stand-alone entity. The changing user expectations and the widening geospatial ecosystem requires more automation, analysis and intelligence, i.e., knowledge than just 'data'" (UN-GGIM, 2022).

Location and time are dimensions that bind information together. Data from multiple organizations need to be properly contextualized in both space and time to support geographically based planning, decision making, cooperation and coordination among agencies and bureaus. Some of the key objectives of the U.S. Geospatial Data Act of 2018 (GDA) are to improve collaboration, reduce waste, and minimize duplication of efforts by harmonizing sources and data standards associated with geospatial data, promoting the integration of geospatial data from all sources, and disseminating information or products that can be readily shared between U.S. federal agencies and other entities (FGDC, 2018). To help obtain the objectives of the GDA and to increase the sharing of knowledge for federal decision-making and operational needs, the U.S. Federal Data Strategy plan identifies as a best practice "linking data to the original source, and then publishing the data as a knowledge graph that integrates feedback from subject matter experts." (Federal Data Strategy, 2021)

A SKG is a useful paradigm for facilitating knowledge sharing and collaboration in a machine-readable way. A SKG can model how locations are related to each other for a given subject domain

---

* Corresponding author

(e.g., public health, economic development, ecological, climate, hydrography). Each node in the graph represents a location and edges represent a semantic relationship (e.g., flows into, adjacent to, is a part of). However, existing spatial data catalogs are dataset-centric and do not model all relevant information associated with an individual location across multiple organizations for a domain. Rather, they are typically tabular datasets containing information about locations of a given type.

Collaboration using SKGs for knowledge sharing involves integrating graphs with nodes and edges by common geographies and other semantic properties, including ontologies, that are sourced from a single source of truth curated by entities with an authoritative mandate. This can be used to derive new and meaningful insights through knowledge inferencing by location or a network of related locations.

1.2 Challenge

The U.S. Office of the Inspector General has identified that some U.S. federal agencies are facing challenges meeting the objectives of the GDA for managing inventory of geospatial assets and promoting the integration of geospatial data (OCIO, 2022). Agencies and bureaus within the U.S. Federal Government have a data silo problem. Public and private organizations do not have a way to easily find data that are mandated to be publicly available. Additionally, data integration is time consuming due to a lack of interoperability standards, including location identity, geometric association, and semantic models. This results in a large portion of data integration and analysis initiatives spent on data cleansing, which is recognized as an obstacle to scaling machine learning efforts (Crowdflower, 2016).

Existing integration difficulties will also be encountered with the increasing adoption of SKGs. Schemas are often not machine interoperable, as inconsistent names and datatypes are used even within the same organization. Having the same (i.e., common) geographies being represented differently across multiple information systems is a challenge for integrating SKGs in an automated way by location, as multiple copies of the same geographic data can exist without a machine-readable way to establish authoritativeness or identity. When there are multiple nodes representing the same location, a graph cannot model all knowledge associated with that location. Rather, each duplicated node can have different semantic properties and associations that, in effect, represent silos within the same graph.

Metadata often do not capture the timeliness of geographic data, such as when boundaries change, split, or merge over time. This can result in erroneous statistical outputs from geospatial analysis performed using data that are non-authoritative, stale and for the wrong time periods. It is common practice to develop analytics or data integration approaches that reference files or database tables rather than machine-readable metadata covering authoritativeness, versioning, and timeliness[1]. Without a mechanism for easily identifying authoritative geographic data of the correct version and for the right time period, agencies and bureaus also end up storing multiple copies of each other's data, resulting in additional storage costs. SKGs should be Findable, Accessible, Interoperable, and Reusable (FAIR) to support a Geospatial Knowledge Infrastructure. However, there are

additional challenges that commonly exist within a knowledge sharing ecosystem:

Findable: Spatial knowledge graphs are a network of interconnected geo-objects (i.e., features). There is typically no mechanism for finding individual geo-objects and their history as they have evolved over time from the organizations with the respective authoritative mandate. Additionally, one cannot query to discover the relationships that a geo-object has with other objects without querying across dozens or more data layers across multiple data sources.

Accessible: Many important semantic relationships between geo-objects are not accessible in a geospatial data ecosystem as graph edges or RDF triples. Rather, many of these relationships exist in the form of analytical outputs that may not have been published as a data product for access outside of the organization that created them. Historical versions of geographic objects are also not always published.

Interoperable: A mechanism for establishing equivalence for the identity of geo-objects across multiple organizations commonly does not exist for the purpose of knowledge sharing. As a result, graphs with geographic data from multiple sources can end up having duplicate geo-objects for the same locations rather than referencing common geographies.

Reusable: Analytical outputs should be published in a format that can be utilized by the analytics efforts of other organizations to facilitate collaboration through knowledge sharing. For example, the ability to combine a hydrographic model and a transportation model using machine-to-machine interfaces that reference common geographies would be beneficial for understanding how flooding could affect transportation for a given area.

2. SPATIAL KNOWLEDGE MESH

When a relationship exists between two geo-objects from graphs curated by different organizations, then this interlinking of graphs represents a dependency between them. A GKI that uses SKGs for knowledge sharing should manage the history of and the relationships between individual geo-objects as they change over time. The architecture we describe in this paper to automate dependency management between interlinked SKGs is the result of learnings from developing open-source geospatial semantic data integration and analysis solutions in the health sector and for U.S. federal agencies for over ten years. We are calling it a spatial knowledge mesh (SKM), as it is a specialization of the data mesh architecture with additional components to provide a framework for automating interoperability for SKGs. The success criteria for this approach are whether it lowers the time to create a sharable knowledge product that can leverage knowledge products from other organizations.

2.1 Requirements

The Geospatial Data Act of 2018 states that one of the roles for GeoPlatform.gov is to "harmonize sources and data standards associated with geospatial data" (FGDC, 2018). As a member of the GeoPlatform.gov ideation team, we identified the following characteristics to enable knowledge interoperability for a GKI within a knowledge sharing ecosystem:

---

[1] Utilizing metadata instead of files to automate collaborative data efforts was a recurring theme of many presentations supporting cloud native approaches at FOSS4G 2022.

Authoritative: Often data need to be copied from a remote source to a local database to ensure adequate performance for analytics. Even as data are copied and linked with other sources, the authoritativeness should always be explicit and unambiguous in a machine-readable way. Therefore, the architecture needs to support the notion of an authoritative copy that always preserves the identity of its source.

Temporal: The period for which data values are valid needs to be explicit and unambiguous in a machine-readable way. This will ensure the correct values from multiple sources for the desired time period are being used. Otherwise, a temporal mismatch can occur where data are integrated from multiple sources but for incongruent periods of time, which will likely result in incorrect analysis. For example, when data reference geo-objects that have split or merged, but the geographic data is from a period prior to these changes, then values will potentially be assigned to the wrong boundaries or to boundaries for which there is no representation in the geographic data. The period of validity should be specified in metadata as a moment in time (such as a date), a frequency (e.g., annually, quarterly), or an interval (year 2000 to 2005) in which data have not changed.

Distributed: One of the significant weaknesses of the data lake or geospatial data warehouse approaches is that IT managers of centralized data stores cannot be domain experts across all the data they manage. Information often gets lost in translation when data processed through extract, transform, and load conform to a centralized model (Majchzak et al., 2023). In his book, *Domain-Driven Design*, Eric Evans outlines how clearly defining software models for a domain and associating them within a bounded context makes it easier to manage complexity when integrating software from different organizations. Otherwise, the inherent complexity of a monolithic data model increases with the addition of domains to the point where it becomes unmanageable (Eric Evans, 2004). For these reasons, a centralized spatial knowledge graph for all geo-objects referenced by every federal, state, local, and private organization and for every context in which they could relate to each other would become an unmanageable behemoth. Rather, A GKI should make knowledge findable across an ecosystem by giving organizations the ability to publish locally hosted graph assets that are domain specific by the experts who curated them. Other organizations can then build a graph with only the data needed for its intended purpose.

Transitive: As changes are made to data dependencies by the respective organizations with the authoritative mandate, they should automatically and transitively propagate to the objects that reference them. For example, in **Figure 1**: graph C depends on graph B which depends on graph A. As changes are made to A, graph C is updated via its dependency on graph B. Dependencies also need to be maintained over different versions of geo-objects and relationships and for the correct period of time.

Versioned: Metadata should explicitly capture the version of data published in a machine-readable way. This is distinct from the period of validity, as data valid for the year 2020 can have multiple published versions, such as 2020 v1, 2020 v2, etc.

Interoperable: The semantic identity of data types, attributes, and relationships should be defined such that equivalency and identity can be established in a machine-readable way across an ecosystem. This would include the use of namespaces, controlled vocabularies, taxonomies, ontologies, and SKGs. **Figure 2** depicts SKGs developed by different organizations that are

interoperable by location because they reference common geographies defined by organization A. Organization D is able to see all relationships associated with these geographies across the ecosystem from organizations B and C.
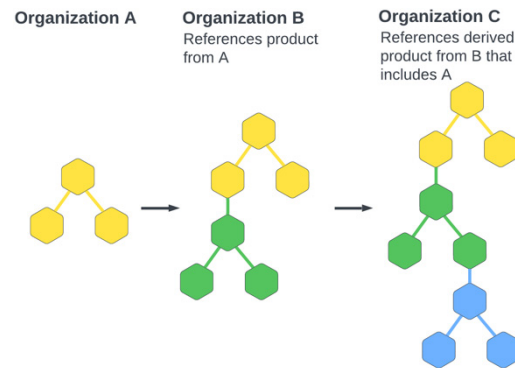

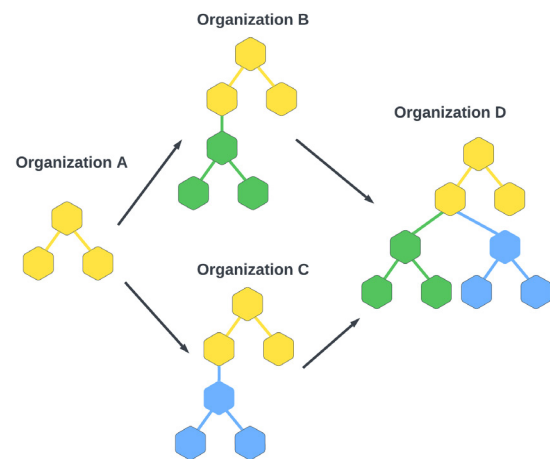**Figure 1.** Transitive management of graph dependencies.

Figure 2. Disparate and interoperable SKGs by common geographies.

2.2 Components

There are some innovative concepts, open-source innovations, and architecture patterns that can be leveraged to implement a spatial knowledge mesh to support a GKI.

2.2.1 Common Geo-Registry (CGR): A common geo-registry provides a single source of truth for managing geographic data over time across multiple organizations and information systems. This enables data to be contextualized from different sources in both space and time to facilitate trend analysis, aggregate data according to different hierarchies, use geographic objects as the common link between data sources, and support GIS-based AI and ML efforts based on common geography (HGLC, 2022). A CGR can be used to implement the following properties of a GKI:

Authoritative. Authoritative entity of data with the curation mandate are explicitly defined.

Temporal. Periods of validity are explicitly defined for changes to geometries, attributes, relationships, and historical events including splits and merges.

Versioned. Versions of data are explicitly defined.

Interoperable. Semantic standards for geo-object types and relationship types are explicitly defined so that they are globally identifiable.

2.2.2   Data Mesh: A data mesh provides a unified framework of interoperability to an ecosystem of largely independent and autonomous data products. It is a domain-oriented decentralized approach to data ownership and architecture (Majchzak et al., 2023). Organizations publish their geographic data in a way that is self-described, discoverable, and interoperable. This will ensure that spatial knowledge graphs will reference common geographies for the correct periods of time, thus enabling the automation of data integration by location from different organizations and across domains. Additionally, storage costs will be reduced by eliminating redundant and non-authoritative copies of geographic data. A data mesh can be used to implement the following properties of a GKI:

Distributed. Authoritative organizations publish curated data that are findable.

Transitive. Updates are automatically propagated across nested data dependencies.

2.2.3   Terminology Services: Standardized vocabularies, taxonomies, and ontologies provide semantic standards for interoperability. Additionally, terms can change over time by splitting, merging and have periods of validity, which will need to be formally managed. Terminology Services enable a GKI to be:

Temporal: Periods of validity are explicitly defined as terms change, split, and merge over time.

Interoperable: Implements semantic standards for data values.

2.2.4   Spatial Knowledge Graph Repository (Graph Repository): In software development, manually coping and merging source code files has long been relegated to the practices of the past, yet copying and linking geospatial data from multiple sources in this manner is common. A spatial knowledge graph repository is an IT concept that we conceived, partially implemented, and deployed to production that manages version dependencies between interlinked geospatial data within a spatial knowledge graph as they change over time. The approach was inspired by how a source code repository manages dependencies and versions of artifacts and libraries. This partial implementation is called GeoPrism Registry (GPR)[2], which is also an implementation of the CGR concept. A graph repository can be used to implement the following properties of a GKI:

Temporal: The graph repository models all changes to data in the graph over time.

Distributed. Updates to graphs, hierarchies, and geospatial data from other organizations can be pulled and merged with local graphs in an automated way.

Versioned. Provides the means to publish versioned knowledge products including labeled property graphs, tabular datasets, and RDF triples.

2.2.5   Interlinked Graph Index: A GKI needs to provide a way to discover not only which published datasets and graphs are available, but also discover everything that is related to an individual geo-object for a domain of interest. It is not practical to build a spatial knowledge graph containing all relationships between geo-objects from every organization within a geospatial ecosystem. Rather, we envision a mechanism for discovering the types of relationships a geo-object participates in and where the are hosted. This would allow in a machine-readable way for organizations to find the information they are looking for that pertains to a given geo-object or a set of objects.
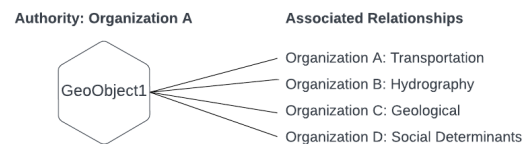

Figure 3. Geo-object discovery.

For example, **Figure 3** shows a geo-object curated by organization A. The interlinked graph index records that it participates in transportation relationships curated by organization A, but also hydrography, geological, and social determinants of health relationships curated by other (and ideally authoritative) organizations.

2.3 Approach

An important characteristic of the data mesh architecture is that it promotes distributed domain-oriented ownership. In the context of a spatial knowledge mesh, SKGs are published as knowledge products that can represent the single source of truth for geo-objects as they relate to each other for different periods of time. They can also model the results of analytics efforts as semantic properties on a graph so that outputs from multiple organizations can automatically be integrated by location (**Figure 2**). These graphs are not necessarily the internal knowledge representation from their respective organizations. Rather, they could come from one or more internal sources that are published in a format to be compatible with the broader ecosystem.
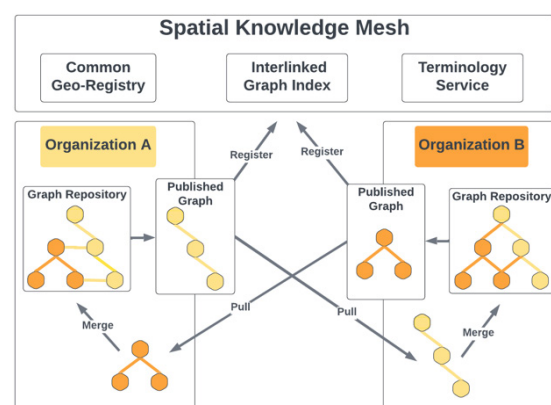

Figure 4. Spatial Knowledge Mesh Overview.

Figure 4 illustrates that individual organizations publish their versioned curated graphs as knowledge products for their respective periods of validity. Dependencies between graphs curated by other organizations are managed within the graph

---

[2] https://geoprismregistry.com

repository. Distributed systems require interoperability standards that can be governed globally (Zhamak Dehghani, 2020). In the case of a spatial knowledge mesh, the interoperability between polyglot domains is achieved through standardized representations of common geographies and relationship types. Building a domain agnostic infrastructure is implemented using domain agnostic data structures such as geo-objects (i.e., nodes) and relationships (i.e., edges), but with domain specific properties captured in the metadata that define them. The governance of this metadata is managed using a common geo-registry and a terminology service can ensure classifications are consistent. The published information is registered in the interlinked graph index so that knowledge associated with individual locations is discoverable across the ecosystem.

3. IMPLEMENTATION

The common geo-registry and graph repository concepts have been implemented in an open-source project called GeoPrism Registry (GPR)[3] that is seeing adoption with some U.S. federal agencies and GeoPlatform.gov. The technology stack has evolved through several projects that needed support for managing location hierarchies and ontologies to facilitate interoperability and perform spatial analysis. The model-driven engineering (MDE) methodology has been utilized to ensure it remains domain agnostic. This was accomplished by implementing a metamodel (i.e., a model for defining models) so that domain specific concepts can be defined declaratively rather than programmatically. We are building upon this foundation to develop an open-source spatial knowledge infrastructure as a platform for implementing a GKI using a spatial knowledge mesh.

3.1 Technical Evolution from Previous Work

3.1.1 Disease Intervention: The foundation of GeoPrism Registry can trace its origins to a Java-based MDE application development framework called Runway SDK[4] to allow for the dynamic creation of models including classes, attributes, enumerations, and associations at runtime. This was partially based on an approach for composing disparate models using configurable semantic rules (Reddy YR et al., 2006). Runway SDK was the foundation for the development of the Disease Data Management decision support tool for vector-borne disease intervention efforts. Support for multiple geographic hierarchies was developed using a geospatial ontological model to normalize multiple datasets by common geographies (Eisen et al., 2011) using a relational database, which saw performance degradation as the length of paths and hierarchy depths grew.

3.1.2 Common Geo-Registry: GPR is the first implementation of the CGR concept. It is open-source and utilizes spatial knowledge graphs to provide a single source of truth for managing geographic data over time across multiple organizations and information systems. It is used to host, manage, regularly update, and share lists, associated hierarchies, and geospatial data through time for geographic objects core to spatial data infrastructure, sustainable development, and public health (e.g., administrative divisions, settlements, health facilities, schools, and other relevant physical and non-physical geographic features).

Initial funding came from the Digital Solutions for Malaria Elimination initiative (DSME) and saw several significant technical enhancements. The relational tables for managing graphs were replaced in the technology stack with a graph database, which allowed hierarchy depths and path lengths to increase by at least an order of magnitude. Through the introduction of a multitenant architecture, dependency management between graphs curated by different organizations was added. Support for tracking change over time in the graph was implemented, as well as the ability to publish snapshots of the graph in tabular form for different moments in time. GPR is currently deployed by the Ministry of Health (MOH) in the Lao People's Democratic Republic to manage the location of health facilities, medical warehouses, and lists of villages for health operational planning relative to multiple and interlinked administrative and health hierarchies. It was recently deployed by ADE in Mozambique, which is the national statistics division, for ensuring consistency of geographic data across ministries.

3.1.3 Climate: Additional capabilities were added through a project with the U.S. Army Corps of Engineers to integrate data from multiple sources to support climate change analysis for U.S. civil infrastructure. Directed acyclic graphs and undirected graphs were added to the metamodel allowing for the definition of non-hierarchical relationships between geo-objects.

3.2 Architecture

GPR manages dependencies between interlinked graphs as they change over time using a labeled property graph, which has advantages over a triple store for managing object state. OrientDB was chosen for its flexible open-source license and for its ability to store documents on nodes, which proved useful for managing the time component. GPR models locations as *geo-objects* whose types are defined by *geo-object types*. A *hierarchy* defines the type of edge that connects geo-objects in a graph tree structure. GPR provides an API for defining geo-object types, hierarchies, directed acyclic graphs, and undirected graph types at runtime[5]. **Figure 5** shows how these concepts are related in the metamodel façade used by the API.
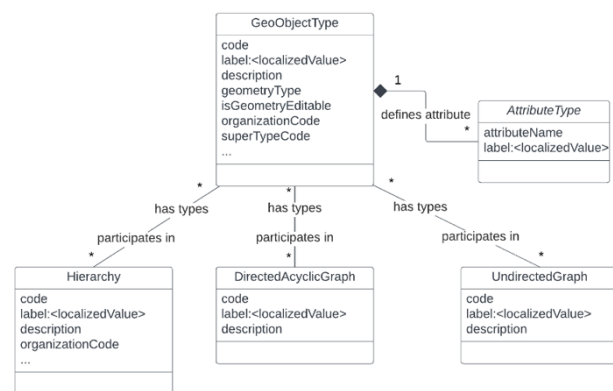

Figure 5. Metamodel façade abstraction.

Figure 6 illustrates the types of location hierarchies used in the field of public health. Dependencies between them are color coded. In this example, which uses fictious geo-object types inspired by J.R.R. Tolkien novels, the Ministry of Home Affairs (MOHA) manages the country administrative hierarchy that is commonly referenced by many organizations. The Ministry of

---

Health (MOH) has hierarchies to manage where health facilities reside, how they are grouped together operationally, and how patients are referred to different types of facilities. The MOH Referral hierarchy is the authoritative source for health facility information, leverages the administrative hierarchy as the authoritative source for villages, and maintains its own link between villages and health posts to indicate where people within a village seek care. The MOH Geographic leverages both the Administrative and Referral hierarchies to model where health facilities reside geographically. The Operational hierarchy models where operational boundaries reside and the facilities they are responsible for.
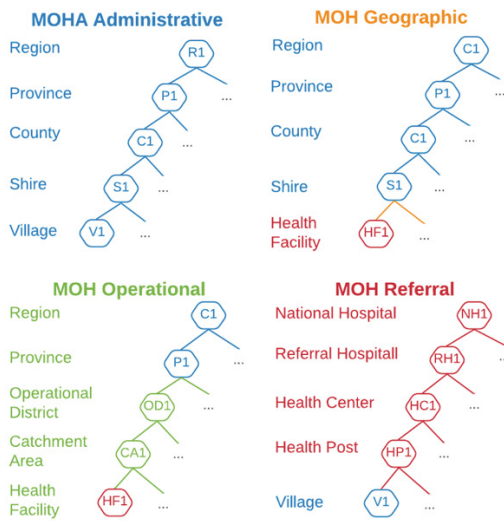

Figure 6. Multiple interlinked hierarchies.

Updating such dependencies across information systems with corrections and changes over time is often a manual and error-prone process (HGLC, 2022). When a province splits, MOHA reassigns child counties within their hierarchy. However, if this information is not propagated to MOH, then their health facility registry will incorrectly state the county in which some facilities reside, which can cause problems for planning and produce errors for aggregate reporting.
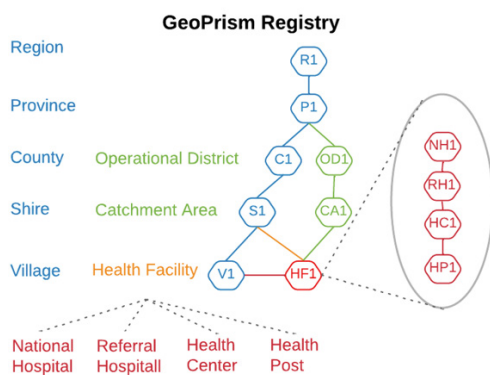

Figure 7. Managing graph dependencies.

Figure 7 shows how GPR manages the kinds of dependencies illustrated in **Figure 6** in a labeled property graph. Organizations are responsible for defining their respective geo-object types and hierarchies. A geo-object is represented as a single node for all references and all moments of time to facilitate propagation of change. Also, note that health facilities in the Referral hierarchy

are represented as four types, but in the Operational and Geographic hierarchies as a single type. GPR supports the ability to group related types into a single abstraction using type inheritance to allow relationships to be defined against a set of related geo-object types.

The metamodel in **Figure 5** is façade that provides a simple abstraction at the API level and for the development of an administrative user interface to define geo-object types, hierarchies and the dependencies between them (**Figure 8**).
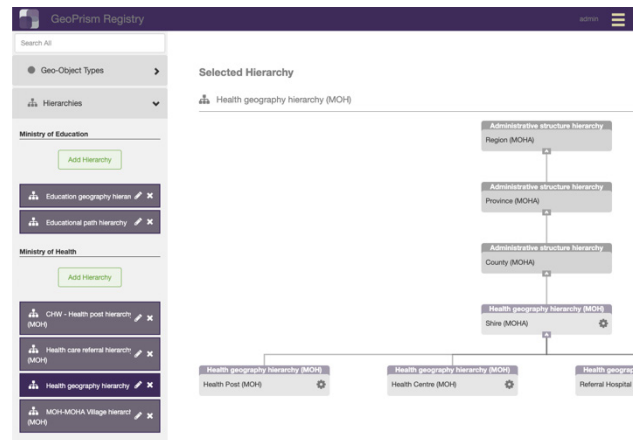

Figure 8. User interface for defining graph hierarchies.

The internal metamodel implementation in GPR is more complex, as it is an evolutionary extension of our prior work from the development of Runway SDK for dynamic type creation (**Figure 9**). Additionally, to implement the requirements of a CGR, it leverages the existing capability for defining tabular datasets that are used to represent the state of the graph (or a subset thereof) for different periods of validity. An *MdClass* is a root-level abstraction for graph classifiers (*MdGraphClass*) and entity tables (*MdEntity*). An *MdAttribute* is the abstraction for defining attributes and is subclassed for different attribute types, including primitives, terms that reference ontologies, and localized attributes. *MdVertex* defines node types and *MdEge* defines edge types between nodes. An *MdGeoVertex* is the abstraction used to define geo-object types.
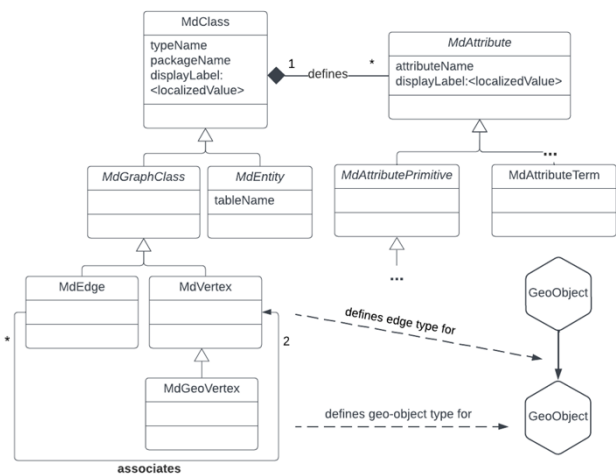

Figure 9. GPR metamodel abstraction.

The ability to store documents on nodes in OrientDB was useful for implementing the time component. Default attributes such as

ID and code, which is used for unique human readable IDs like postcodes, are immutable. **Figure 10** illustrates how values can be specified for different periods of time on dynamically defined attributes in the graph repository. The same approach is used for geometries as well. Additionally, the period for when a geo-object exists can be specified, such as when a health facility closed and physically no longer exists. When a product is published for a period of validity, only the value which is considered authoritative for that period is included in the product. This typically is the last value within that period. For example, although **Figure 10** shows the name of a location having changed three times within a year, the product only contains the last value for that same time span.
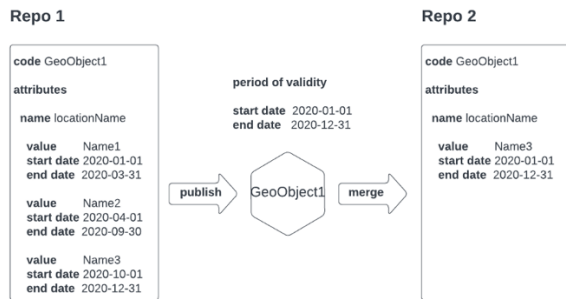

Figure 10. Products and change over time.

Currently GPR supports the ability to publish data tables as products for a geo-object type for a single date, frequency (e.g., annual, quarterly), or for intervals (e.g., 2017 to 2022). Users can specify which hierarchy relationships to include. The graph is then traversed to populate tables with data for the correct time. This is currently being extended to support the publication of labeled property graph products. **Figure 11** shows the UML class model in development that allows for the definition of a *Labeled Property Graph Type*, which includes which geo-object types (via. The MdGeoVertex abstraction) and edge types will be included in the product. The *LPGEntry* class represents a definition of the time component (single date, frequency, etc.) and the *LPGVersion* class represents an instance of the graph for the defined time period.
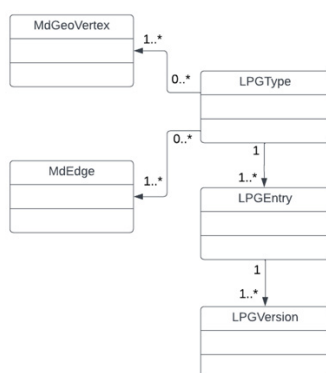

Figure 11. Abstraction for defining LPG products.

This approach is being adapted for an application on GeoPlatform.gov. The Imagery Data Manager (IDM) is an open-source and cloud-based application to provide a storage, management, and processing collaborative solution for sensor outputs collected by unmanned aircraft systems (UAS)[6]. Spatial

knowledge graphs built using GeoPrism libraries organize data by location and how sensors, imagery and processed outputs are related to each other. The United States Forest Service (USFS) wants to organize data according to their own hierarchies, yet the U.S. Department of Interior (DOI), which is the entity that funded the development of the application, would like an architecture that can accommodate the diverse needs of multiple agencies and bureaus. GPR was chosen as the solution for managing multiple hierarchies for IDM with the intent to explore leveraging these same hierarchies in other systems to ensure consistency as they change over time.
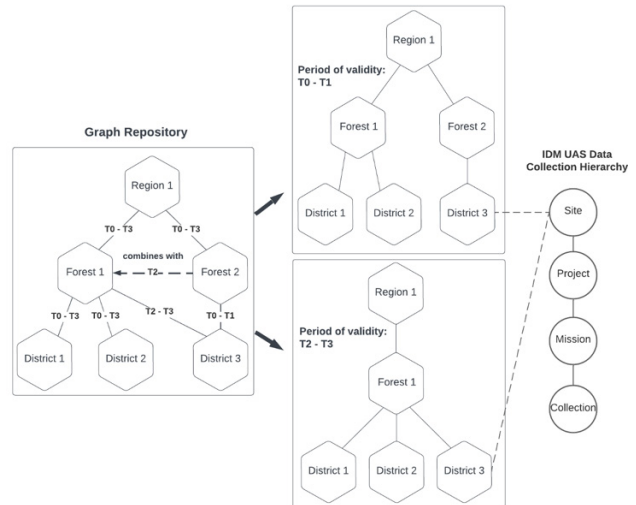

Figure 12. Publishing products for different periods of validity.

Figure 12 illustrates how the graph repository keeps track of changes to the USFS hierarchy. At time T2, Forest 2 merged with Forest 1. District 3 was a child of Forest 2 from T0 to T1, and then from T2 to T3 became a child of Forest 1 as a result of the merge. Graph products can be published showing the state of the hierarchy before and after the combining of forests. The data collection hierarchy in IDM will reference the published USFS hierarchy for the appropriate period of validity. This approach will allow UAS data collection sites to reference any number of hierarchies from different U.S. agencies, allowing each entity that adopts the IDM to find sites, raw sensor data, and processed sensor outputs (including orthorectified images and point clouds) according to their own organizational structure or other geographic context.

4. FUTURE WORK

The U.S. Department of Interior, the U.S. Department of Agriculture, members of the Federal Geographic Data Committee, and participating members of Earth Science Information Partners (ESIP)[7] Discovery Cluster have provided valuable feedback regarding this approach. One of the concerns raised is that some agencies have very large graphs and the merging of dependencies within a graph repository could result in the need for significant additional storage. We will explore a design to add support for a symbolic link between geo-objects curated in graphs from different entities to minimize the amount of data that would be duplicated. Rather than copying every geo-object from a graph, only enough data would be copied as necessary to model direct associations with geo-objects between different organizations. It was also suggested that the ability to limit the amount of data published in a graph product by a

---

[6] https://www.geoplatform.gov/apps

[7] https://www.esipfed.org

geographic context would be useful. This could be implemented by specifying a geo-object as the root of a published hierarchy or the starting point of a directed graph. Work needs to be done to explore how to define product outputs for triple stores to benefit environments using RDF and GeoSPARQL. One area of interest to us is to research how the concepts for dependency management translate into non-geospatial semantic structures, including taxonomies and ontologies. We are currently only addressing publicly available products, but consideration needs to be given to how this approach could support a permissions model.

5. CONCLUSION

In the recently released *Geo-enabled Microplanning Handbook* by the WHO-UNICAF COVAX GIS Working Group, the need for more efficient approaches for health services delivery planning was highlighted as critical due to the limited resources available for public health in many countries. A common geo-registry was identified as a fundamental requirement to facilitate needed interoperability at scale (WHO, 2023).

The explosive uptake of ChatGPT seems to indicate people will increasingly seek information and generate content using chatbots. Examples of AI-driven chatbot technology providing misleading, harmful, biased, or inaccurate answers due to a lack of access to information highlight the importance of making authoritative knowledge accessible, interoperable, and usable for machine-to-machine readable interfaces through GKIs to support AI efforts. Should indeed there be a transition underway from data-centric to knowledge-centric solutions, then tools and methodologies will need to become more knowledge-centric. Our aim is to reduce the effort to build GKIs by continuing to refine the spatial knowledge mesh architecture and developing the GeoPrism technology stack into a spatial knowledge infrastructure as a platform. Ultimately, we need more test cases to demonstrate our approach lowers the time to create a sharable knowledge product that leverages knowledge products from other organizations.

ACKNOWLEDGEMENTS

REFERENCES

Crowdflower, 2016. Data Science Report. https://visit.figure-eight.com/rs/416-ZBE-142/images/CrowdFlower_DataScienceReport_2016.pdf (20 April 2023)

Eisen L., Coleman M., Lozano-Fuentes S., McEachen N., Orlans M., Coleman M., 2011. Multi-disease data management system platform for vector-borne diseases. *PLoS Negl Trop Dis.*;5: e1016. pmid:21468310

Evans E, *Domain-Driven Design* (Boston: Addison-Wesley, 2004), p. 336

Federal Data Strategy Working Group, 2021. Federal Data Strategy 2021 Action Plan. https://strategy.data.gov/assets/docs/2021-Federal-Data-Strategy-Action-Plan.pdf (13 December 2022)

Federal Geospatial Data Committee, 2018. Geospatial Data Act of 2018. https://www.fgdc.gov/gda/43-usc-ch-46-geospatial-data-geospatial-data-act.pdf. (22 November 2022)

Health GeoLab Collaborative, 2022. HIS geo-enabling: Guidance on the establishment of a common geo-registry for the simultaneous hosting, maintenance, update, and sharing of master lists core to public health (Version 2.0). https://healthgeolab.net/DOCUMENTS/Guidance_Common_Geo-registry_Ve2.pdf. (4 December 2022)

Majchzak et al., *Data Mesh in Action* (Shelter Island: Manning Publications, 2023), p. 58

Office of the Chief Information Officer, 2022. USDA's Compliance with the Geospatial Data Act for Fiscal Year 2022 https://www.oversight.gov/report/USDAOIG/USDA's-Compliance-Geospatial-Data-Act-Fiscal-Year-2022. (20 April 2023)

Reddy YR et al., 2006. Directives for Composing Aspect-Oriented Design Class Models. *Transactions on Aspect-Oriented Software Development I, vol. 3880, Berlin, Heidelberg*, pp. 75–105 . doi: 10.1007/11687061_3.

UN-GGIM, 2018. Integrated Geospatial Information Framework Part 1: Overarching Strategic Framework. Version 1. https://ggim.un.org/meetings/GGIM-committee/8th-Session/documents/Part%201-IGIF-Overarching-Strategic-Framework-24July2018.pdf. (26 November 2022).

UN-GGIM, 2022. Development of Contextual Understanding, Information, and Analytics Towards Determining the National Geospatial Information Ecosystem. https://ggim.un.org/meetings/GGIM-committee/12th-Session/documents/Future_National_Geospatial_Information_%20Ecosystem_Jul2022.pdf. (8 September 2022)

World Health Organization GIS Working Group for COVAX Innovation, 2023. The Geo-enabled Microplanning Handbook. https://www.covid19giswg.net/geo-enabled-microplanning. (10 February 2023)

Zhamak Dehghani, 2020. Data Mesh Principles and Logical Architecture. https://martinfowler.com/articles/data-mesh-principles.html. (4 April 2020)