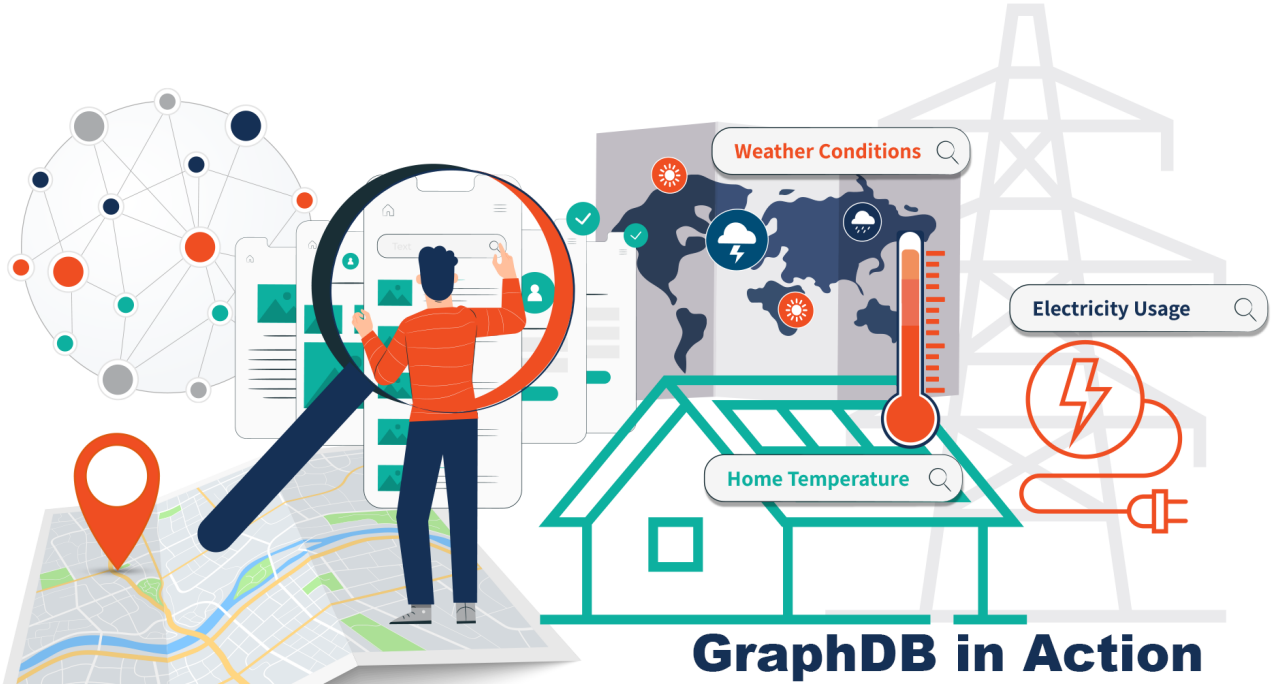
Authors: Teodora Petkova, Gergana Petkova



In today's world, we increasingly interact with the environment around us through data. This might include a range of tasks such as searching weather conditions on an app, looking at the temperature at home, reviewing electricity usage on a dedicated web platform, all the way to managing and utilizing data in a professional system at work.

For all these data operations to flow smoothly, data needs to be interoperable, of good quality and easy to integrate. In other words, data coming from different sources need to be interlinked, contextualized and normalized in a graph that allows for its consistent and unambiguous interpretation. And this is what semantic data management is about. In this part of our series *GraphDB in action*, we highlight cutting-edge research where GraphDB has been used to power solutions built with semantic data.

A Geo-knowledge Graph Based on RDF, OWL and GeoSPARQL

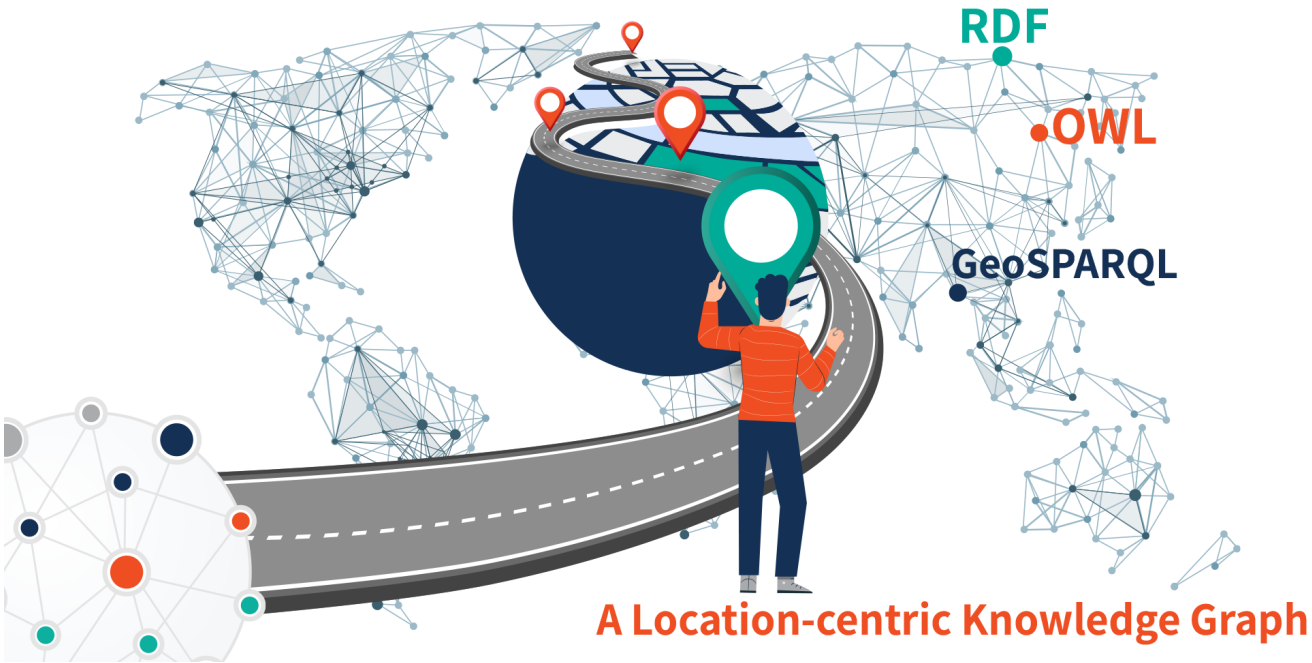
The first paper in focus is Know, Know Where, KnowWhereGraph: A densely connected, cross-domain knowledge graph and geo-enrichment service stack for applications in environmental

Copyright: Ontotext New York | Sofia | Basel
Reach out to us: info@ontotext.com

intelligence by Krzysztof Janowicz, Pascal Hitzler, Wenwen Li et al. published as a special topic article in AI magazine, Volume 43, Issue 1, Spring 2022.

The paper introduces KnowWhereGraph (KWG) as a solution to the ever-growing challenge of integrating heterogeneous data and building services on top of already existing open data. KnowWhereGraph is a location-centric knowledge graph of data at the interface between humans and the environment, based on existing standards like RDF, OWL and GeoSPARQL. It incorporates custom ontologies and uses a hierarchical discrete global grid for spatial representations.

Motivating the use of knowledge graph technology for environmental data, the authors explain why spatial data requires special treatment. They suggest application areas for the graph in disaster mitigation, (food) supply chain management and the broader Environmental, Social, and corporate Governance (ESG) market. They also introduce services on top of the graph such as GeoEnrichment – the process of augmenting local data on-demand with custom tailored contextual information.



KnowWhereGraph's value proposition is that it can deliver area briefings for any place on

Copyright: Ontotext New York | Sofia | Basel
Reach out to us: info@ontotext.com

earth within seconds to power environmental intelligence applications and data-driven decision-making more broadly.

More concretely, KnowWhereGraph answers questions such as:

- “What is here?”
- “What happened here before?”
- “Who knows more?”
- “How does this region/event compare to other regions/events?”

To do so, KWG draws from over 30 fully integrated and semantically homogenized data layers. The current graph release (called *Vienna*) contains 12.5B triples. This number is expected to grow beyond 20B by next year, as reported by the authors of the project. These 30 layers can be split into two kinds: a location-reference layer and a topic layer. KWG contains ten location layers, including millions of named places, administrative borders, weather zones, climate zones, zip codes, fips codes, market areas, etc.

As neither natural disasters nor the curiosity of data scientists stops at fiat borders, KWG also contains the S2 discrete global hierarchical grid down to a level thaty allows users to freely combine cells (covering about 1 km2) to define their own region of interest. For each region, known or user-defined, the graph contains topologically-registered data from about 20 thematic layers, including population characteristics, transportation infrastructure, crop time series data for agricultural applications, soil characteristics, past events , and climate predictions, to name but a few. The triple store is powered by GraphDB, while interfaces, GeoEnrichment services and so forth are custom-developed, e.g., as add-ins to Esri’s ArcGIS Pro Geographic Information System.

KnowWhereGraph’s value proposition is that it can deliver area briefings for any place on earth within seconds to power environmental intelligence applications and data-driven decision-making more broadly. Share on X

Copyright: Ontotext New York | Sofia | Basel
Reach out to us: info@ontotext.com

Transforming the European Transportation Ecosystem One Triple at a Time

The second paper we want to talk about is SPRINT: Semantics for PerfoRmant and scalable INteroperability of multimodal Transport by Mersedeh Sadeghi, Petr Buchníček, Alessio Carenini, Oscar Corcho, Stefanos Gogos, Matteo Rossi, Riccardo Santoro, published in Proceedings of 8th Transport Research Arena TRA 2020, April 27-30, 2020, Helsinki, Finland.

The paper presents early results of the SPRINT project, which plays a central role in the Shift2Rail IP4 work programme. The authors address the challenge of interoperability in the digitalization of mobility systems and introduce a reference architecture for the Shift2Rail Interoperability Framework (IF). The IF's aim is to enable multimodal travel in a highly diverse environment and with many transport modes.

The IF conceptual architecture includes various components, one of which is the Asset Manager. It acts as a catalog of assets that are involved in various publication processes. The catalog stores the asset's metadata in RDF. It is also able to transform the asset descriptions

Copyright: Ontotext New York | Sofia | Basel
Reach out to us: info@ontotext.com

into RDF triples and to feed them to an RDF repository, which is GraphDB. This allows keeping a well-defined representation of the metadata of each asset and enables using a SPARQL endpoint to query it.

This paper concludes that the proposed reference architecture realizes the objectives of the IF in two ways. “Firstly, through the Asset Manager, which masks the complexity of interoperability to travel applications by publishing uniform abstractions of services, and which enables travel applications to communicate among them uniformly (e.g., web service/API interfaces and communication protocols). Secondly, by providing additional technical means to automate and facilitate seamless and smooth cooperation of heterogeneous and fragmented transportation actors and to enable them to operate on the web of transportation data.”

Using Data From Building Automation And Control Systems Semantic For More Efficient Analysis

The last paper in this selection is Data Integrity Checks for Building Automation and Control Systems by authors Markus Gwerder, Reto Marek, Andreas Melillo and Maria Husmann, published in Proceedings of CLIMA 2022 Conference, May 2022.

Copyright: Ontotext New York | Sofia | Basel
Reach out to us: info@ontotext.com

Authors: Teodora Petkova, Gergana Petkova

This paper addresses the challenge of using reliable and trustworthy data in Building Automation Systems. As data from building automation and control systems is growing in quantity, its quality needs to be improved in order to allow better management and data analytics over it. Towards that end authors introduce a system for integrity checks for building automation applications and using more reliable data for data analytics processes.

As the paper argues, the building automation industry faces a growing need for smart data integration in order to manage and utilize the data coming in from controllers. And while growing in quantity, the quality of this data is often poor “due to erroneous installation, commissioning, data recording or meta-information”. Focused on acquisition only, building automation engineers do not take into account the tagging quality and the need for analytics at a later stage. As a result of these data quality issues, the need for integrity checks arises.

The paper presents such data integrity checks for building automation applications, with examples using data recorded from a real building automation project – the very interesting case of Aspern Smart City Research – the largest and most innovative energy research project in all of Europe.

Copyright: Ontotext New York | Sofia | Basel
Reach out to us: info@ontotext.com

A major component of the developed solution for integrity checks is the semantic modeling of data. For prototyping and evaluation of data integrity check algorithms, a software test environment was designed and used for preprocessing and storing data, corresponding semantic model data access and automated execution of data integrity checks.

The software development environment consists of three main parts – a time series database, a graph database to store semantic information, and a data workflow management platform. Researchers used GraphDB to store semantic metadata. As authors highlight, semantic data is a key component to "achieve a high degree of automation in setting up the checks [and] it is only through them that the recorded time series data are given context and meaning." This use of semantic technology is already proven to be efficient in practice – two of the leading vendors of Building Management Systems already use GraphDB to support more efficient operations of tens of thousands of buildings around the globe.

Semantic data is a key component to achieve a high degree of automation in setting up the checks and it is only through them that the recorded time series data are given context and meaning. Share on X

Semantic Data For Meaningful Applications

The highlighted research showcased that it takes a semantic data approach to tackle the challenges of the increasingly complex environments we interact with. Building efficient applications for such environments relies on the right data. And the right data is the data put in context and made meaningful. Maturing over time, RDF graph databases and semantic technologies are proving to be a reliable solution for bringing context and meaning to data and factoring in interoperability, sustainability and repeatability in data projects.

Need a reliable RDF graph database for your research?

Copyright: Ontotext New York | Sofia | Basel
Reach out to us: info@ontotext.com

GraphDB in Action: Putting the Most Reliable RDF Database to Work for Better Human-machine Interaction

Authors: Teodora Petkova, Gergana Petkova

Ontotext's GraphDB
Give it a try today!
Download Now

Copyright: Ontotext New York | Sofia | Basel
Reach out to us: info@ontotext.com