The generation, transmission, distribution and sale of electrical power generates a lot of data needed across a variety of roles to address reporting requirements, changing regulations, advancing technology, rapid responses to extreme weather events and more.

One major driver is the European Green Transition, Europe's new growth strategy, designed to transform the Union into a modern, resource-efficient, and competitive economy.

This initiative increases roles for prosumers, energy communities, and distributed energy resources, precipitating increased and more complex communications for organisations such as TSOs.

The European Union has had standards in place for several years to impose structure on the variety of data involved with energy generation and delivery. One technology getting attention in this area is knowledge graphs (KGs), as the related standards and tools can contribute semantics, stronger reliability and greater interoperability to this data so users can optimise and leverage its full value.

Current electricity data standards

EU legislation, such as REMIT, charges the European Network of Transmission System Operators for Electricity (ENTSO-E) with collecting electricity market data from TSOs.

TSOs must provide this data in a format from the IEC's family of standards known as the Common Information Model (CIM). The CIM is defined as "an abstract model that represents all the major objects in an electric utility enterprise" and it covers activities that can take place in an enterprise ranging from optimisation and congestion management to enterprise resource planning and asset management.

Another important standard is the Energy Identification Code (EIC), which is used for global identifiers of energy resources and parties. A resource might be something like a power plant or a specific generator within a power plant.

The CIM requires that objects defined in its data have resource identifiers but, currently, there is no requirement that they be globally unique identifiers. This creates a challenge, as two sets of data can identify the same resource differently, so aggregated data will not show

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

accumulated figures relevant to that resource that might reveal useful patterns.

The use of EICs in these identifiers is a step toward a knowledge graph approach and the benefits that it can bring.

What are knowledge graphs and how can they help?

Unlike relational databases, graph databases store information as networks of nodes and edges.

Like the network of an electrical grid, this adds flexibility in finding connections between nodes. It also adds flexibility in accommodating new kinds of data, including metadata about existing data points that lets users infer new relationships and other facts about the data in the graph.

When you connect different sets of related contextual data and add schemas of metadata about that data's potential structures and relationships, a more powerful version of a graph known as a knowledge graph is developed.

Knowledge graphs that are created with the W3C data model standard RDF (Resource Description Format) – along with related W3C standards such as RDFS for schemas, SPARQL for querying, and SHACL for enforcing data quality constraints – help to connect the data to even more datasets, offering opportunities to find connections and patterns in the data that contribute to the accomplishment of business goals.

Another specification that is typically used with these W3C standards is Uniform Resource Identifiers (URIs), which enable the creation of unambiguous identifiers that make it possible to link common entities described in different datasets.

Support for RDF and related semantic technology standards makes it easier to mix and match data from different sources. This use of additional data that may have otherwise remained in silos makes it easier to identify new trends and patterns that can drive cost savings, market opportunities and other efficiencies.

For example, for an electricity grid data manager who is facing challenges in the assessment

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

of data quality and low-quality data that is unreliable for analytics, a knowledge graph that leverages existing electricity data standards can take advantage of:

- Data validation with an extensible set of rules
- Automated re-validation as part of daily updates
- Automatic data correction to increase the relevance of subsequent rules.

An energy trader struggling to find correlations across disparate public data can also use the additional context of a knowledge graph's explicit relationships among data values for more effective querying and analytics.

An early version of CIM used an RDF-based syntax that is still used for some CIM messages. But, because it was done a few years ago, it could not build on later developments that added to RDF's capabilities such as RDF Schema and SHACL.

Stakeholders can get more out of CIM with several applications of knowledge graph principles. Several of these ideas also play an important role in the Linked Data model of sharing data across organisational boundaries, including:

- **Using persistent URIs for entities referenced in the data**. The EIC standard can provide a great foundation for these, and the urn:eic namespace was registered for this (URNs being a type of URI), so this is a positive step forward.
- **Publishing accumulated CIM in a standardised form** which applications can reach on the public Internet instead of only using this data for transient messages.
- **Using a more modern RDF syntax** such as JSON-LD or Turtle.
- **Taking additional advantage of the W3C RDF Schema** (and optionally OWL) standards to publish data models describing the structure of published data. Schemas are an example of how the right metadata can add value to the data it describes. These also enable inferencing, which lets users derive new knowledge from an existing collection.
- **Linking the data to related data in other collections** and adding other data to this collection. Contextual data is one of the most valuable contributors when building a regular data graph into a knowledge graph.
- **Using the W3C SHACL standard to define machine-readable constraints** results in higher quality data as arriving data can be automatically checked for the presence of

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

required values, the use of the correct data types, and other constraints on numeric, date, and string data.

Knowledge graphs in action: Enter the transparency energy knowledge graph

As a prototype of an electricity data knowledge graph that demonstrates this potential, Ontotext worked with ENTSO-E to take a subset of the data in their Transparency portal, publish it as the Transparency Energy Knowledge Graph (TEKG), and add interactive features that demonstrate the value of the new knowledge graph.

This data was combined with sources such as VIES for VAT validation and OpenStreetMap for power plant and transmission line maps and coordinates.

SHACL rules as part of the knowledge graph helped to identify several categories of data quality issues.

In fact, some EIC resources are described multiple times, and entries referring to the same resource have different descriptions; some resources have a null or invalid value in their 'function' field, which should describe the nature and market role of the resource; some resource functions have variations in spelling (for example. 'Power Unit' vs 'Power Plant'), and some market participant VAT numbers are syntactically invalid or expired.

The TEKG's web-based dashboards allow analysts with no background in knowledge graphs to perform faceted searches of constraint violations and CIM data, and those with more knowledge of the standards can explore the data with their customised SPARQL queries.

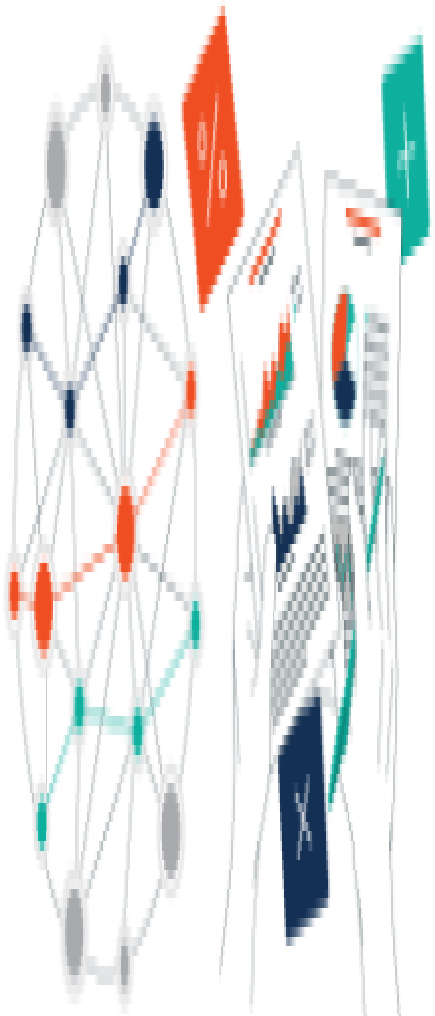
Easier identification of addressable data problems is just one example of how a knowledge graph like the TEKG can let members of the electrical power industry achieve their goals more efficiently.

Combining multiple datasets to create a whole that is greater than the sum of its parts – and doing so with tools whose standards support lets you mix and match them – is truly enlightening and offers a growing choice of possibilities.

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

ontotext

Interested to see a knowledge graph in action?



Try our knowledge-graph-powered search

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