



DELIVERABLE 4.4: SUMMARY OF MODIFICATIONS DELIVERED TO THE SCHEMA.ORG STANDARD FOR ENERGY

Report

Authors: Pietro Zambelli, Valentina D'Alonzo, Samuele Zilio (EURAC)

Revised by: Diane von Gunten, Jakob Rager (CREM)



Table of Contents

List of tables	1
List of abbreviations.....	1
1. INTRODUCTION	2
2. SCHEMA.ORG: WHAT IS IT?	3
2.1. The existing “Energy” schema.....	3
2.1.1. MAIN LIMITATIONS.....	4
3. THE SUGGESTED MODIFICATIONS	5
3.1. Introduction.....	5
3.2. Proposal	5
3.3. Next steps	9
4. REFERENCES	10



List of tables

Table 1. Properties that might be used to characterize an Energy value based on the current schema definition.	4
Table 2. Properties that might be used to define the proposed EnergyDataset entity.	8

List of abbreviations

CSA	Coordination and Support Action
D	Deliverable
FAIR	Findable, Accessible, Interoperable, Re-usable
H2020	Horizon 2020
JSON-LD	JavaScript Object Notation for Linked Data
M	Month
OEPO	Open Energy Platform Ontology
RDFa	Resource Description Framework in Attributes
URL	Uniform Resource Locator
W3C	World Wide Web Consortium

1. INTRODUCTION

EnerMaps is a Horizon 2020 (H2020) Coordination and Support Action (CSA) project that aims at improving data management practices in energy research and management. Currently, energy data is often difficult to find, mixed in different repositories, and fragmented, which can slow project progress, increase project costs, and create an overall lack of efficiency in the field of energy research. EnerMaps will act as a quality-checked database of crucial energy data that will communicate and disseminate data effectively and efficiently using practices to make the data findable, accessible, interoperable, and re-usable (FAIR).

This document is the Deliverable 4.4 (D4.4) of the EnerMaps project, which describes the activity performed to improve the Energy ontology in the Schema.org project (1). The purpose of this report is to synthesize the main issue of the current schema and to initialize a discussion with the main subjects involved. The result of this deliverable is a list of possible modifications to integrate and extend the current Energy ontology definition to improve energy classification.

To follow the main FAIR principles, it is important to provide data with a rich meta-data, to facilitate the understanding of the meaning of the different variables/data sets and to facilitate the orange-to-orange comparison. Schema.org provide a set of ontologies for different entities like: Dataset (2) and Energy (3). However, as will be discussed in this report the current energy definition is link only to a number and not to a Dataset, and even for a single number it misses relevant information like: Is it from a renewable source? What is the primary source used? Which technology has been used to convert the primary source? In what physical form is the energy?

The report briefly introduces the main Schema.org concepts, then presents the main limitations of the status quo and provides some possible modifications to the current ontology schema.

2. SCHEMA.ORG: WHAT IS IT?

Schema.org is a collaborative and community activity that aims to create, maintain, and promote schemas for structured data on the Internet, web pages, email messages, etc.

Schema.org vocabulary can be used with many different encodings (e.g. RDFa, Microdata, JSON-LD). These vocabularies cover entities, relationships between entities and actions, and can easily be extended through a well-documented extension model. A shared vocabulary makes it easier for webmasters and developers to decide on a schema and get the maximum benefit for their efforts. The main objective of the founders, together with the larger community have come together, is then to provide a shared collection of schemas.

Many applications like Google, Microsoft, Pinterest, Yandex, and others, are already using these vocabularies. In addition to people from the founding companies (Google, Microsoft, Yahoo, and Yandex), there is an important participation by the larger Web community, through public mailing lists such as public-vocabs@w3.org and through GitHub (4).

The daily operations of Schema.org are handled by a steering group, which includes representatives of the founding companies, a representative of the World Wide Web Consortium (W3C) (5) and a small number of individuals who have contributed substantially to Schema.org. The discussions of the steering group are public. The steering group is responsible for high level oversight of the project (including approval of new releases), and the larger community group handles the activities of schema evolution, discussion, and integration.

The steering group has a mailing list, whose notes are posted to public GitHub and linked from issue #1 in the project's issue tracker. The community group has mainly the role to propose, discuss, prepare, and review changes to Schema.org, for final review and publication by the steering group. The community group also represents the hub for discussion with other related communities, at W3C and elsewhere. The Schema.org GitHub repository is considered a tool of the community group and is currently the primary focus for community discussion.

2.1. The existing “Energy” schema

The Uniform Resource Locator (URL) to reach the standard for “Energy” is the following: <https://schema.org/Energy> (3) (under [Thing](#) > [Intangible](#) > [Quantity](#)) and its properties are represented in the following table. Currently, the properties that take Energy as values are of the form '<Number> <Energy unit of measure>'. As mentioned, Table 1 shows the properties that are currently available to characterize the Energy value.



Table 1. Properties that might be used to characterize an Energy value based on the current schema definition.

Property	Expected Type	Description
Properties from Thing		
additionalType	URL	An additional type for the item, typically used for adding more specific types from external vocabularies in microdata syntax. This is a relationship between something and a class that the thing is in. In RDFa syntax, it is better to use the native RDFa syntax - the 'typeof' attribute - for multiple types. Schema.org tools may have only weaker understanding of extra types, in particular those defined externally.
alternateName	Text	An alias for the item.
description	Text	A description of the item.
disambiguatingDescription	Text	A sub property of description. A short description of the item used to disambiguate from other, similar items. Information from other properties (in particular, name) may be necessary for the description to be useful for disambiguation.
identifier	PropertyValue or Text or URL	The identifier property represents any kind of identifier for any kind of Thing , such as ISBNs, GTIN codes, UUIDs etc. Schema.org provides dedicated properties for representing many of these, either as textual strings or as URL (URI) links. See background notes for more details.
image	ImageObject or URL	An image of the item. This can be a URL or a fully described ImageObject .
mainEntityOfPage	CreativeWork or URL	Indicates a page (or other CreativeWork) for which this thing is the main entity being described. See background notes for details. Inverse property: mainEntity
name	Text	The name of the item.
potentialAction	Action	Indicates a potential Action, which describes an idealized action in which this thing would play an 'object' role.
sameAs	URL	URL of a reference Web page that unambiguously indicates the item's identity. E.g. the URL of the item's Wikipedia page, Wikidata entry, or official website.
subjectOf	CreativeWork or Event	A CreativeWork or Event about this Thing. Inverse property: about
url	URL	URL of the item.

2.1.1. MAIN LIMITATIONS

The current schema applies only to a numeric value and there is no possibility to describe an energy dataset. Furthermore, it is not possible to define the energy carrier in the schema (e.g. electric, mechanical, chemical, heating, gas, etc.), the primary source of energy (e.g. Oil, Solar, Wind, Nuclear, etc.), the technology used to convert/use the energy (e.g. boiler, heat pump, photovoltaic power plant, etc.), the energy "type" (e.g. consumption, production, primary energy, final energy, energy demand), the source of the data (e.g. measured, simulated, statistically assessed, physically modelled), and finally it is not possible to specify the spatial and temporal extent of the energy numeric value.

3. THE SUGGESTED MODIFICATIONS

3.1. Introduction

The modifications that are structured and described in this document have been shared with the Schema.org community in a dedicated discussion #2957 (6). Before being integrated on the next release of the Schema.org they need to be deeply discussed with the community. Each of these suggestions needs to be discussed with the community to understand the best way to integrate them. Therefore, this document reports just the start of an ongoing process. Furthermore, we are aware that other vocabularies exist on this topic, particularly the Open Energy Platform Ontology (OEPO) (7). However, the OEPO vocabulary is well known and less used respect to the Schema.org vocabulary, therefore integrate at least some basic classification and specification properties to the Schema.org Energy definition might have an important impact on the larger community of energy data users.

3.2. Proposal

The main objective of the suggested modifications is to improve the discoverability of energy datasets by large search engines, such as Google. A notable improvement is also the addition of different important properties such as energy type (primary, final, etc.) and type of calculation methodology (measurement, modelled, etc.).

The proposal would like to add the following properties to the existing Energy schema:

1. **Primary energy source** to identify the main source that has been used to produce the energy. This property could be an Enumeration entity with, for example, values like:
 - Non renewable
 - Fossil fuels
 - Oil
 - Coal
 - Natural gas
 - Mineral fuels
 - Natural uranium
 - Natural thorium

- Renewable
 - Solar energy
 - Wind energy
 - Water
 - Failing
 - Flowing
 - Tidal
 - Biomass
 - Geothermal
- Mixed

2. **Energy system component/technology** that might be a new dedicated entity like: *EnergyComponent* that takes some flow energy as input and converts into another form of energy or produce something. A not exhaustive example of an Energy Component might be:

- Photovoltaic power plant
- Fossil fuel power plant
- Nuclear power plant
- Gas boiler
- Mixed
- Etc.

These Energy Components might be characterized for instance by an installed capacity (kW), an energy efficiency to perform the conversion (%), the date of the installation, the space position, etc.

3. **Energy carriers**, the energy that is exchanged using one or more of the following carriers, like for example:

- Liquid fuel:
 - Gasoline
 - LPG
 - Alcohol
 - Bio-fuel
 - Etc.
- Electricity
- Mechanical work

- Thermal:
 - Heating
 - Cooling
- Gas:
 - Bio gas
 - Hydrogen
 - Syngas
 - Etc.
- Mixed

The Energy Carrier might be defined as an Enumeration entity or using a dedicated *EnergyCarrier* entity where it is possible to further detail the main energy carrier characteristics.

4. **Energy type** that defines what kind of energy as an Enumeration entity, possible values might be:

- Production (the energy produced by a plant / energy component)
- Demand (the energy required by a system e.g. a residential house to satisfy its needs)
- Consumption (the energy that has been used by an energy component)
- Primary energy (before any transformation to a secondary form of energy)
- Final energy (that includes the losses for the energy transmission)

5. **Type of source of data** that describes / characterizes the source of the number, like for instance:

- Measured
- Simulated
- Assessed
- Modeled
- Unknown

Furthermore, a common use case is to apply the above properties to characterize and describe not a numeric value but a whole dataset. To cover this use case, the EnerMaps project suggests adding a new Schema.org entity, called *EnergyDataset*, that would extend the Dataset schema with the new energy properties used already to extend the Energy schema (Table 2).

Table 2. Properties that might be used to define the proposed EnergyDataset entity.

Property	Expected Type	Description
Properties from EnergyDataset that extends the Dataset schema (2)		
PrimaryEnergySource	Enumeration	It defines the primary energy used to generate the energy. Valid values are: Oil, Coal, Gas, Natural uranium, Natural thorium, Solar energy, Wind energy, Failing and flowing water, Tidal energy, Biomass sources, Geothermal energy, Mixed.
EnergyCarrier	Enumeration or EnergyCarrier	It defines the energy carrier of the dataset Valid values are: Electricity, Heat, Fuel (solid, liquid and gas), Hydrogen, Mechanical work.
EnergyComponent	Enumeration or EnergyComponent	It defines the energy component of the dataset. Valid values are: Oil refinery, Fossil fuel power station, Nuclear power plant, Thorium breeder reactor, Photovoltaic panel, Solar power tower, Solar thermal panel, Wind farm, Hydropower plant, Wave farm, Tidal power plant, Biomass power plant, Geothermal power plant.
DataSourceType	Enumeration	It defines the data source of the dataset. Valid values are: Assessed, Measured, Modelled.
EnergyType	Enumeration	It defines a qualitatively classification of the energy in the dataset. Valid values are: Production, Demand, Consumption, Primary energy, Final energy.

The existing schema “Dataset” already has a good set of properties that can be used to provide useful information of an *EnergyDataset* as well. Of particular interest are the attributes on spatial (*spatialCoverage*) and temporal (*temporalCoverage*) coverage to provide information on the spatial and temporal dimensions. Furthermore, the current schema already defines properties to describe the authorship (person or organization), how to cite the dataset, an abstract to summarize the content of the dataset, how to describe the methodology (*measurementTechnique*), how to specify the variable that is measured, the copyright and the licensing.

In the following, some examples of energy datasets that could be described with the new properties are listed.

- Dataset of estimation of solar photovoltaic potential on the building roofs in Europe.
 - *PrimaryEnergySource*: Solar Energy

- *EnergyCarrier*: Electricity
- *EnergyTechnology*: Photovoltaic panel
- *DataCalculationMethod*: Assessed
- *EnergyType*: Production
- Dataset of estimation of thermal energy demand of buildings in Europe:
 - *PrimaryEnergySource*: Multiple
 - *EnergyCarrier*: Heat
 - *EnergyTechnology*: Multiple
 - *DataCalculationMethod*: Modelled
 - *EnergyType*: Demand
- Dataset of monitoring data of thermal consumption in the residential buildings in Bolzano:
 - *PrimaryEnergySource*: Multiple
 - *EnergyCarrier*: Heat
 - *EnergyTechnology*: Multiple
 - *DataCalculationMethod*: Measured
 - *EnergyType*: Consumption

3.3. Next steps

The report presents a set of changes that have been suggested to the Schema.org community, we expect that in the next months these points will be further discussed with the Schema.org community to assess the technical feasibility of these changes and to further refine/revise the proposal.

In the next months we will spread this proposal through our network channels to receive feedbacks and collect ideas to further improve it. We will contact by email people that, at different level, might be interested in improving and extending the proposal. We will open a post in the forum of the Open Energy Modelling Initiative (OpenMod) and, through their mailing list, we will open a post on LinkedIn to engage and give visibility on the discussion to other professionals. Finally, through a person-by-person contact with the running projects that might be interested in using/adopting this schema, we will discuss and refine this proposal.



4. REFERENCES

1. **Schema.org**. *Schema.org*. [Online] 2021. <https://schema.org/>.
2. **Schema.org**. Dataset. *Schema.org*. [Online] 2021. <https://schema.org/Dataset>.
3. **Schema.org**. Energy. *Schema.org*. [Online] 2021. <https://schema.org/Energy>.
4. **GitHub**. GitHub. [Online] 2021. <https://github.com/>.
5. **W3C**. *World Wide Web Consortium*. [Online] 2021. <https://www.w3.org/>.
6. **Zambelli, Pietro**. Discussion #2957: Additional properties to better characterize: Energy . *GitHub*. [Online] 2021. <https://github.com/schemaorg/schemaorg/discussions/2957>.
7. **Open Energy Platform**. Open Energy Family - Open Energy Ontology (OEO). *GitHub*. [Online] 2021. <https://github.com/OpenEnergyPlatform/ontology>.



The Open Data Tool empowering
your energy transition.

WHAT IS ENERMAPS?

EnerMaps Open Data Management Tool aims to improve data management and accessibility in the field of energy research for the renewable industry.

EnerMaps tool accelerates and facilitates the energy transition offering a qualitative and user-friendly digital platform to the energy professionals.

The project is based on the FAIR principle defining that data have to be Findable, Accessible, Interoperable and Reusable.

EnerMaps project coordinates and enriches existing energy databases to promote a trans-disciplinary research and to develop partnerships between researchers and the energy professionals.

Project Coordinator

Jakob Rager, CREM
jakob.rager@crem.ch

Communication Coordinator

Clémence Contant, REVOLVE
clemence@revolve.media



The EnerMaps project has received funding from the European Union's Horizon 2020 research and innovation programme under [grant agreement N°884161](#)