

data.reegle.info – A New Key Portal for Open Energy Data

Mag. Florian Bauer (Operations Director REEEP),
Denise Recheis (Junior Researcher REEEP),
Martin Kaltenböck, CMC (CEO Semantic Web Company)

REEEP – The Renewable Energy and Energy Efficiency Partnership,
Wagramerstrasse 5, 1400 Vienna, Austria,
{Florian.Bauer, Denise.Recheis}@reep.org, m.kaltenboeck@semantic-web.at

Abstract. Over the last few centuries knowledge and information has become much more widely available. It is no longer limited to a select few; more often it is available to anyone, anywhere with an internet connection and access to a computer. A major step forward was the invention of the printing press, resulting in greater access to information, higher literacy rates and improved levels of education. This sharing of knowledge has been radically transformed again recently with the onset of the internet combined with rapid technology advances. Linked Open Data is the next step along those lines, offering us opportunities to use the latest technologies and techniques available to share pertinent, up-to-date information, from a wider variety of sources free for use and re-use. One of the main challenges now is to sort through this ever increasing volume of information, of varying quality; to select the most relevant, up-to-date and reliable information from the different sources and formats in use; and to make this information available in a useful, reliable and standardised way. The information gateway *data.reegle.info*, which is at the forefront of this development, is consuming and publishing energy related data in a machine-readable format so as to allow the flexible sharing and re-use of information in new and innovative ways. It filters and enriches content by sorting and adding pertinent information from various reliable sources. It enables data providers to easily link and publish datasets in a standardised, machine-readable format. In this way it can be easily accessible, and shared with 3rd parties. This is a real benefit for the users of reegle which targets policy- and decision makers, developers and the general public with an interest in renewable energy and energy efficiency, with a focus on developing countries.

Keywords: semantic web, open government data, energy data, linked data

1 Introduction

Several governments and official agencies have pledged to make all their collected data publicly available, a move that makes sense considering it is public money that is being spent on gaining this data. These days, such data is published on the internet, meaning there is more and more of it available online. Sharing data is expected to deliver more transparency, efficient public service and better use and re-use of data in the public and commercial domains.

Especially in the field of (clean) energy, it is crucial to have access to latest and accurate datasets and to be able to link, compare and mash-up this information to provide decision makers, financiers and project developers with the necessary set of information to accelerate the marketplace of clean energy. The Renewable Energy and Energy Efficiency Partnership (REEEP) has taken up the challenge of developing and running a web portal (<http://www.reegle.info>), which acts as a single point of access to these datasets, offering its datasets according to the Linked Open Data W3C Standards and enriching its own information with relevant data from other open data providers to offer the full benefit of information on clean energy to the user.

Linked Data best practise will allow the extension of the worldwide web to a global data space that can connect data from various domains like scientific data, businesses but also film and radio programmes. New applications using Linked Open Data can incorporate new data appearing on the web automatically by recognizing its relevance, and thus the user will receive more holistic results. reegle is joining this movement by providing and using Open Data from renewable energy related themes, and by offering its own applications as well as making its data available for applications of external sites. By using only high-quality sources, concerns about the trustworthiness of data are appeased for users as well as developers working with data retrieved through reegle.

1.1 Linked Open Data W3C Standards

When great amounts of data are published, it is important to structure, catalogue and document it in a way that makes it retrievable – not only by humans but also by machines. In order to make data actually re-useable, the distribution of raw data is crucial. This raw data has to be in a well defined machine-readable, structured format.

The Linked Data standards, techniques and technologies allow information to be published in an open and standard format with common web based access methods. This standard web-based data format allows data of various kinds, vocabularies, sources, semantics and authors to be federated, queried and analyzed. It is based on

the “Resource Description Framework” (RDF) – a very open and general way to publish and consume data¹.

1.2 SKOS and RDF Format

SKOS (Standard Knowledge Organisation System)² is a formal and controlled vocabulary used in the context of the semantic web for creating taxonomies, thesauri and ontologies. Knowledge organized in such a way uses, for example, the Resource Description Framework (RDF) which encodes the information in a way that it can be used in decentralized metadata applications. RDF triples consist of a subject, a predicate and an object. The subject and object are labelled as a `skos:Concept`. The predicate describes the “properties” between subject and object, which could be `skos:broader` if the subject is a child of the object, like for example if “jeans” were the subject then the predicate to its object “trousers” would be `skos:broader`. So RDF is built from classes/concepts and the properties between them, and each class and property has its own unique identifier URI. This “third dimension” added when converting data to RDF triples allows mash-ups to be created with data from various sources to give it new meaning. This adds lots of value and results in highly interesting and informative applications, especially in the context of resource and energy data.

1.3 Thesauri

Thesauri³ use controlled vocabulary in the sense of a closed list of named subjects as a means to index subjects/concepts. SKOS thesauri use a concept/subject-based classification. Controlled vocabulary has the advantage over uncontrolled vocabulary in that it avoids the use of different terms for identical subjects/concepts, like misspellings, old-fashioned names or simply singular forms. Information, data can be retrieved by its metadata which describes objects which are connected to the subjects/concepts they are about and which are used to classify the objects. An improvement over old-fashioned term-based thesauri is the fact that a concept/subject-based thesaurus allows poly-hierarchies, meaning that a subject/concept can have more than one broader relation. Each concept will always have the same narrow relations, whatever its broader concept may be.

Thesauri are a powerful way to classify objects and term subjects/concepts and to provide a way of searching for and browsing data. Furthermore thesauri enable very powerful search based applications as for instance similarity mechanisms et al. To

¹ W3C eGov Wiki (<http://www.w3.org/RDF/>)

² SKOS (<http://www.w3.org/2004/02/skos/>)

³Thesaurus (<http://en.wikipedia.org/wiki/Thesaurus>)

handle this amount of data becoming available now as open data, filter mechanisms and categorization are more crucial than ever.

The reegle Thesauri

In fact, the reegle thesaurus consists of several thesauri covering the subjects of renewable energy technologies as well as climate compatible development. Having been built with the software PoolParty⁴, it boasts full Linked Data capabilities and it is based on W3C principles which allow linking with other LOD⁵ sources. Having developed a controlled vocabulary for the fields it covers, the reegle thesaurus is capable of structuring and categorizing the many concepts of clean energy. As a hand-crafted piece of work, it can be adjusted when required but operates and powers some of reegle's most important features without the need for constant man power.

The fact that several domain experts from engineering as well as international development have contributed to the reegle thesaurus ensures a high quality and relevance for reegle users. They have the benefit of vast knowledge combined with the comfort and ease of a modern web-tool.

2 reegle.info as a Linked Open Data Publisher/Consumer

"reegle.info", the gateway for high quality information on renewable energy and energy efficiency acts as a main information source for more than 90,000 users per month in the clean energy sector. It is an offshoot from The Renewable Energy & Energy Efficiency Partnership (REEEP, <http://www.reeep.org>), a non-profit, specialist change agent aiming to catalyze the market for renewable energy and energy efficiency, with a primary focus on emerging markets and developing countries.

Originally launched in 2006, reegle was designed as a specialist search engine for renewable energy and efficiency issues. It didn't publish any content of its own, but eventually it developed further and an actors catalogue was produced for reegle. Two years later reegle underwent its first total make-over when a regular energy blog and a small glossary were introduced, and outreach could be expanded. Still, reegle was promoted foremost as a search engine and the website's design strongly focused on this feature.

Through experience and feedback, it became more apparent that user's tended to benefit more from reegle's secondary features, and more features were added. A reegle map visualizing energy statistics and potentials was introduced and again well-perceived by the public. As an ever evolving dissemination tool, it was decided in 2010 to add reegle to the LOD cloud as a user and producer of Open Data.

The new data portal (data.reegle.info), launched in 2011, has established reegle as a publisher and consumer of Linked Open Data in the energy sector in an innovative

⁴ <http://poolparty.punkt.at>

⁵ Linked Open Data

design. This means that data provided on reegle can be located and re-used by 3rd parties for free, and content is enriched by all relevant data from the LOD (Linked Open Data) cloud. It now provides key datasets free for re-use using Linked Open Data W3C standards. These datasets include extensive energy facts, a set of more than 1700 key stakeholders (actors catalogue) and a comprehensive thesaurus/glossary describing the clean energy sector. On www.reegle.info all available information is enriched by data from other data providers, resulting in comprehensive information dossiers, like the country energy profiles which summarize the most important energy related information (e.g. Regulations, Statistics, Organizations ...) for all countries worldwide. reegle data is already being used by external developers to enrich their own information systems as openei.org or the Solar-Med-Atlas (<http://dev.geomodel.eu/solar-med-atlas/>) and thus proves its relevance. This new development once again reinforces reegle's position as a one-stop-shop in the field of renewable energy, energy efficiency and climate compatible development. The energy sector is a vivid example of how Open (Government) Data can be utilized and benefits development and implementation of energy policies, targets and helps decision makers and project developers alike not only by offering the latest relevant data but also by offering new ways of using it.

2.1 The data.reegle.info SPARQL Endpoint

The SPARQL⁶ endpoint⁷ at reegle.info offers developers to retrieve data on energy statistics, key stakeholders and individual country energy profiles as well as the whole reegle.info thesaurus on clean energy for free and unlimited re-use.

As an example, a user could send a query regarding reegle stakeholders in energy and efficiency NGOs active in the Czech Republic. The query would look like this:

```
PREFIX foaf: <http://xmlns.com/foaf/0.1/>
PREFIX dcterms: <http://purl.org/dc/terms/>
PREFIX reegle: <http://reegle.info/schema#>

SELECT ?actor ?name
WHERE {
  ?actor a foaf:Organization.
  ?actor dcterms:subject <http://reegle.info/categories/11>.
  ?actor foaf:name ?name. FILTER (lang(?name) = "en" )
  ?actor reegle:activeIn <http://reegle.info/countries/CZ>.
}
```

(Example from data.reegle.info/developers/sparql-intro)

The user will receive a list of all relevant stakeholders, each with its own unique identifier, for example LaGuardia Foundation with its URI (Unified Resource

⁶ SPARQL as the query language of the semantic web offers a powerful API to retrieve data and manage complex queries over several data sources

⁷ http://semanticweb.org/wiki/SPARQL_endpoint

Identifier) <http://reegle.info/actors/2423>. This information can be used by developers for any application they can think of, free and without strings attached.

3 Benefits for reegle users

3.1 Glossary, Web search

Both glossary and intelligent web search are powered by the thesaurus. The glossary makes use of linked data principles and displays Wikipedia definitions beside reegle's own definition and informs the reader of synonyms, abbreviations or other terms that describe an identical "unit of thought"; they are included in the thesaurus as `skos:prefLabel` and `skos:altLabel` which allows for one preferred label and several alternative labels. Other sources like suitable glossaries or thesauri can easily be integrated to contribute more information when need arises. The web search can be refined through choosing related terms from the thesaurus in a box which can be ticked and which then will be added to the original terms. Related searches are also offered on basis of narrower terms at the bottom of the page. Auto-complete is another convenient measure that is offered, and common spelling errors are accounted for by including them as `skos:hiddenLabel` in the thesaurus. reegle users can thus find what they are looking for by refining their search and because the reegle web search operates on carefully pre-selected websites, reliability of the retrieved information is guaranteed.

3.2 Visualizing knowledge

The reegle thesauri don't only power the intelligent web search and clean energy glossary in the ways just described, but PoolParty also allows a visual display of all the hierarchical and associative relations of a concept. This can really benefit understanding complex and abstract concepts, and by clicking onto the relations the user can dig deeper into a subject and get the full graphic display for each unit of thought.

3.3 Country Energy Profiles - a reegle Mash-Up

Following reegle's re-launch in 2011, it now boasts comprehensive and well-arranged energy-related information on all of the world's countries and regions. These profiles are not researched and written by a person, but they are an example of making use of open data that is freely available for re-use. Valuable documents from REEEP's own database such as the regulatory overview and the reegle stakeholders (actors) are presented in the brand-new reegle design and this information is now for the first time enriched with easy-to-understand reegle graphs showing energy demand, production and emissions amongst other key facts retrieved from trusted sources, such as UN data and the World Bank. A short general description and the flag of each country is

contributed by DBpedia⁸ and traditional “links” to relevant web pages documents or other relevant information round up these comprehensive dossiers.

This new asset is attracting a lot of project developers, decision makers and other interested parties because they can now gain a complete overview in a user-friendly format, getting the relevant facts from the major sources in one place as well as re-use them for free.

Another advantage is that new sources that suit the high quality requirements are regularly reviewed and can be added very easy to improve reegle’s country energy profiles by such information.

4. Outlook

The ever increasing information flow can bring about many chances and opportunities, but now is the time to make some considerations and choose the right formats and create the right applications to show the public what can be done with Linked Open Data. One of the recent projects are the reegle.info country energy profiles which link and mash up data from different sources into one comprehensive package and offer machine-readable links to country information in other relevant data sources worldwide. reegle is determined to partake in this exiting development and be at the forefront of open energy data and to underline this progression it has just been re-launched in a new, modern and user-friendly design highlighting its most important features.

⁸ Wikipedia datasets (<http://dbpedia.org/About>)

References

1. The World Wide Web Consortium (W3C), <http://www.w3.org/>
2. W3C eGov Wiki (http://www.w3.org/egov/wiki/Linked_Data)
3. The Renewable Energy & Energy Efficiency Partnership (REEEP), <http://www.reeep.org/>
4. Garsho, L., Metadata? Thesauri? Taxonomies? Topic Maps! Making sense of it all (2004), <http://www.ontopia.net/topicmaps/materials/tm-vs-thesauri.html#N412>
5. Campell, L., MacNeill, S., The Semantic Web, Linked and Open Data; A Briefing Paper (2009)
6. Linked Data – Connect Distributed Data across the Web, <http://linkeddata.org/>
Bizer Ch., Heath T., Berners-Lee T., Linked Data, The Story So Far; In: International journal on semantic web and information systems (2009)