

Using LOD¹ to Share Clean Energy Data and Knowledge

Denise Recheis, Florian Bauer

REEEP – The Renewable Energy and Energy Efficiency Partnership, Wagramerstrasse 5,
1400 Vienna, Austria,
(denise.recheis, florian.bauer)@reep.org

Abstract. This paper explains why and how some of the most essential features of reegle.info, a one-stop information gateway in the renewable energy and energy efficiency sector, are based on Linked Open Data technologies. reegle focuses heavily on clean energy and related issues and draws most of its content from a wide range of open data sources – in such a way it maintains itself and ensures that reegle users have access to the latest high quality information, presented in a visually appealing way. Services such as reegle’s comprehensive dossiers on individual countries take advantage of the future web of (LOD) data. All content is also available for external use via reegle’s data portal.

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

1 reegle – a Clean Energy Open Data Portal

Reegle² has already established itself as a popular information portal in the field of renewable energy and energy efficiency. With more than 220,000 users per month,³ its services have been well perceived by the community and proved their usefulness. Reegle was first launched by REEEP⁴ and REN21⁵ in 2006 as a free specialist search engine and information gateway. In response to recent changes in the way that data is retrieved, offered and displayed on the web, a complete re-design in its style and content, as well as in its technology and services was deemed necessary. This makeover has been ongoing since 2010 and has opened up many new avenues for reegle and its users.

As a consumer as well as a provider of open data, reegle displays data retrieved from various authoritative sources in new and interesting ways while at the same time offering all such data for free integration into other websites. In fact, Reegle now offers all of its data under the W3C standards, as open and linked data in a non-proprietary (RDF⁶) format.

¹ Linked Open Data

² www.reegle.info

³ as of April 2012

⁴ Renewable Energy and Energy Efficiency Partnership – www.reep.org

⁵ Renewable Energy Policy Network for the 21st Century – www.ren21.net

⁶ Resource Description Framework, Linked Open Data format

1.1 reegle's Decision to Go LOD

Reegle was founded to ensure access to timely information in the clean energy sector. Originally a search engine was viewed as the most useful tool to connect users and relevant content, but the onset of Open Data and Linked Open Data today allows such information to be made directly available to users on the reegle website.

Reegle operates on a now-profit basis and is funded by UK's Department of Energy and Climate Change, the Federal Ministry for the Environment, Nature Conservation and Nuclear Safety of Germany and most recently also the Climate and Development Knowledge Network (CDKN). As reegle is financed by public money, using the most efficient and cost-effective systems to operate is mandatory.

Going back to the portal's inception in 2006, modern technologies have always been seen as key to the upkeep of the portal; much more so than large manpower. The idea of sharing data und avoiding duplication thus has a very strong appeal to its operators. Being a relatively new field, Linked Open Data technology is also an opportunity for reegle to take a spearheading role. In this position, reegle now acts as much more than an information gateway. It is also a source of expertise on the implementation of LOD technologies for REEEP partners who are keen to take advantage of the possibilities offered by the semantic web.

2 Web 3.0 and the Future of Data Providers

Over the last few years, a new paradigm has emerged in the IT scene: the shift from the traditional "web of documents" towards a new "web of data." This is driven by the vision of a semantic web, a world where machines can "understand" how things are connected and thus greatly increase efficiency, effectiveness and enjoyment for providers and users of the WWW⁷. The shift from hypertext as html to hyperdata as RDF has added a new dimension of semantics to information that is processed by machines.

This new technology has been embraced by several major providers of information such as the British government portal data.gov.uk and its US equivalent data.gov. Publishing Linked Open Data is increasingly the method of choice for those organizations supplying data.

⁷ World Wide Web

2.1 Data on the Web or Data in the Web?

The real difference between the traditional web and what has become known as the semantic web is the way in which information or data is actually published.

To understand the difference, Tim Berners-Lee's⁸ 5 Star Model, first presented in 2012, is useful. Michael Hausenblas⁹ explains the cost and benefits for publishers based on this model.

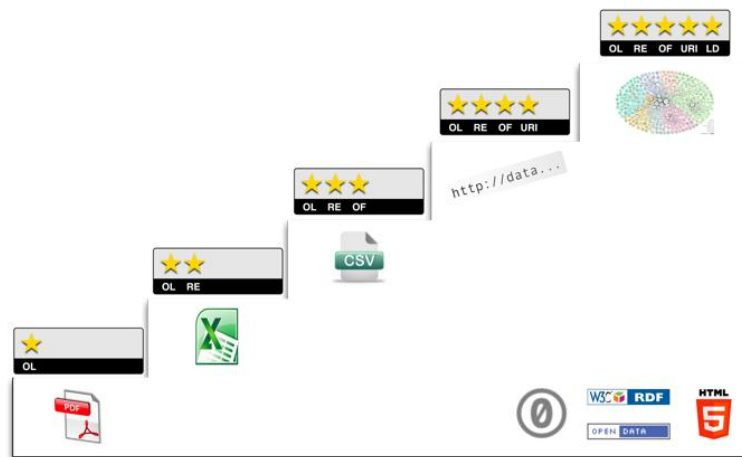


Fig 1: Visualization of 5 Star Model, including the corresponding formats. (Source: <http://5stardata.info>)

★

One star data is the first level of Open Data¹⁰, data that is available to be used freely in any possible context. Yet if information is contained in documents, for example in a PDF file, it is difficult for others to use it in their own web portals.

Reegle will include PDF documents when they are available under an Open License, but their contents cannot be directly integrated into reegle services.

★★

Even though still locked within a document, two star data is structured in a machine-readable format. Proprietary software is needed to be able to use the information.

Some of the data that is relied on for reegle services, including some statistics, is extracted and converted to RDF from structured Excel sheets.

⁸ http://de.wikipedia.org/wiki/Tim_Berners-Lee

⁹ <http://sw-app.org/mic.xhtml#i>

¹⁰ Data available in the web under an open license

★★★

Three star data is similar to two star data, only the data is now structured in a non-proprietary format and is available to use for everyone on the web. Yet the actual data is still “on the web” rather than “in the web” as an integral element.

Similar to two star data, reegle uses this kind of structured and freely available three star data to provide relevant information in the clean energy context.

★★★★

Four star data is data “in” the web, meaning that important data now has its own URI¹¹ and is an integral part of the web itself. The native way to represent such data in the web is RDF, but there are also other options.

Reegle has taken the decision to publish its data in the web by modeling the data and giving things their URI (e.g: www.reegle.info/glossary/492/hydro-power.rdf). All of the information on reegle is available via its SPARQL¹² endpoint at data.reegle.info as RDF.

★★★★★

Five star data is what the semantic web actually consists of: data in the web that is interlinked and forms a tight net.

To actually weave the semantic web it is crucial to link such data in the web so that people as well as machines can explore the related content. Reegle links its data to external data in the web to support the increasing growth of the semantic web. In order to ensure that its data is well linked to semantically related data, Reegle has forged collaborations with other key organizations in the clean energy and climate arena. One example is the “concentrated solar power” entry in the thesaurus, which is linked to the related Wikipedia and ClimateTechWiki definitions (<http://www.reegle.info/glossary/1367/concentrated-solar-power.htm>)

A quote from Rick Jelliffe¹³ explains the difference between “on the web” vs. “in the web” well: “...make a distinction between a resource being “on” the web or “in” the web. If it is merely “on” the web, it does not have any links pointing to it. If a resource is “in” the web, it has links from other resources to it. [...] A service that has no means of discovery (i.e. a link) or advertising is “on” the web but not “in” the web, under those terms. It just happens to use a set of protocols but it is not part of a web. So it should not be called a web service, just an unlinked-to resource.”

¹¹ Unique Resource Identifier

¹² SPARQL as the query language of the semantic web offers a powerful API to retrieve data and manage complex queries over several data sources

¹³ <http://www.oreillynet.com/pub/au/1712>

2.2 The Sematic Web is a Smarter Web

Keeping one's data in "silos" is not a very cost-effective way of running a database. It must constantly be updated and adapted to reflect new use cases or greater amounts of information. In contrast, LOD technologies are very flexible in regards to these issues, and after an initial set-up can significantly reduce running cost of maintenance.

Once a template for a certain webpage has been designed, the actual data can be retrieved from the source regularly and is only cached for a specified amount of time. This means that once a dossier has been laid out as a template, the information will stay up-to-date and relevant without further manual effort. This is a great benefit that is exploited by reegle to ensure that its users have access to the latest data from their field of interest.

Before LOD, the data providers that produced most data to publish or sell were regarded as the most valuable. Now, under this smarter approach, information provider value is judged on whether it can channel data to wherever it is most useful. For smaller organizations that combine (Linked) Open Data to suit a certain use case and cater for it in a much more tailored fashion while also being able to rely on external data providers, this can be a great advantage.

Publishing one's data in the W3C standards as Linked Open Data also means specializing in what an organization does best – the web of data doesn't know "data silos" but views the entire net as one huge database. Therefore it is also a system of de facto shared responsibilities, where different sources provide different facets of the full picture. The full picture itself can also differ, depending on the mandate of an organization. Each web publisher can enrich their own data with whatever data is necessary to round off the experience, make their points, and increase understanding for users. Since each provider is only responsible for their particular share of the data, this increases quality and efficiency while decreasing cost and effort for all data providers.

The whole idea can be compared to the onset of Adam Smith's notion of increased productivity through specialization. Division of labor has traditionally been seen as increasing efficiency in production, and with the creation of content and data, this claim is more valid than ever.

2.3 Semantically Linking People and Data

An important lesson learned during reegle's shift to an LOD portal is the recognition that just throwing out so-called Open Data is not enough. Since the inception of LOD, more and more organizations have begun to publish their own data according to W3C standards. But if this data is never taken up and used by external websites or applications, most of the added benefits of the semantic web go to waste. For this reason, reegle has been actively approaching potential partner organizations to ensure a lively exchange of experiences as well as integration of each other's data.

One such fruitful collaboration has been established between OpenEI¹⁴ and reegle. Like reegle, OpenEI produces and consumes Open Data and spearheads this devel-

¹⁴ OpenEnergyInformation, a service from the US National Energy Laboratory

opment in the US clean energy sector. Reegle uses OpenEI data within its clean energy dossiers, the country energy profiles. Reegle's thesaurus-based glossary is linked with corresponding terms from openei.org, and both sites display each other's definition on top of other LOD definitions.

OpenEI is also one of the first beneficiaries of the new reegle API¹⁵, a service based on the SKOS¹⁶ -thesaurus for automated tagging and categorization of energy and climate documents.

This project has also strengthened ties between reegle and other organizations such as weADAPT and Eldis; these organizations are also part of the reegle API project. Further collaborations have been established between reegle and ci:graps and the Clean Energy Solution Center.

The technology is used to forge connections not just between people, but the data itself is available in RDF graphs¹⁷; the non-proprietary RDF format that connects concepts across different datasets through RDF triples¹⁸.

3 Application Areas of Linked Open Data

Decision-making on complex topics should be based on reliable and timely data. A future web of Linked Open Data will enable several applications to support this process.

3.1 Mashups Combining Related Datasets

So-called mashups are a valuable use case where data from several different sources is combined in a defined way to allow the user an overview of the most relevant, reliable information on a certain subject.

For reegle this is a particularly interesting method to provide users with tools necessary to accelerate the uptake of clean energy technologies. The first example of this kind of mashup on reegle are the country energy profiles.

In its country energy profiles, Reegle now offers comprehensive and well-arranged energy-related information on each of the world's 243 countries and regions. A template was originally designed and data sources explored. Reegle enriches relevant external open data with information drawn from REEEP's own database. On top of a Wikipedia definition, all country profiles display a flag and relevant energy statistics from established sources, and all reegle stakeholders ("actors") active in the relevant

¹⁵ Application Programming Interface,
http://en.wikipedia.org/wiki/Application_programming_interface

¹⁶ Simple Knowledge Organization System

¹⁷ RDF graphs consist of RDF triples

¹⁸ An RDF triple consist of: the subject (URI reference or a blank node), the predicate (URI reference), and the object (URI reference, a literal or a blank node)

country. The statistics come from trusted sources such as UN, World Bank and Euro-Stat and graph relevant trends such as clean energy generation and consumption.

Tools are another section of reegle's country energy projects and act as a useful gateway for users looking for a wide range of tools.

Projects, Programmes and Projects outputs give a unique snapshot of activities on the ground with information pulled from a variety of internal and external sources.

New sources are constantly being reviewed to see if they fit reegle's high quality requirements. If yes, and they also enrich the quality of the country energy profiles, they will be added. Right now the country profile mash-ups are based on (Linked) Open Data from UN Data, World Bank Data, DBpedia¹⁹, Eurostat, OpenEI²⁰, DFID, RESLegal, and REEEP with SERN. Since these sources are known keep their data up-to-date, the country energy profiles always display the latest available information.

3.2 Complex Queries

Sematic search engines such as Wolfram Alpha²¹ already give a hint of what is possible when machines can actually "make sense" of data. In the clean energy sector, questions such "Which regions have the highest potential for solar power?" or "What country has the highest feed-in tariff for renewable energy?" are of great relevance, yet the answers can be difficult and tiresome to research. A sematic search engine will be able to give an answer to such questions rather than display many documents that may, or may not, include this information. This is another exciting method that could help people take the right decisions based on reliable and accessible information.

3.3 Increased Social Context

Services such as Quora²² already exploit the connection between personal interest, available resources and the fact that often several people are interested in the same thing. Quora aggregates questions and answers to topics and allows users to collaborate on them by editing questions and suggesting edits to other users' answers.

Information on "who knows about what" is well-suited to be integrated into the knowledge of the sematic web, and the added dimension of personal interests will bring a new depth to Linked Open Data.

¹⁹ Wikipedia datasets

²⁰ OpenEnergyInformation (US National Energy Laboratory)

²¹ <http://www.wolframalpha.com/>

²² <http://en.wikipedia.org/wiki/Quora>

4 LOD and Sustainable Development in Developing Countries

Ensuring relevant information about technologies, policy, best practice, statistics and events is readily available for implementers and public servants is an important pillar for sustainable development. Since reegle has a special focus on clean energy in developing countries, it is important that such local players can benefit from the increasing amount of information that becomes available.

Recent studies show that by early 2012, some 65%²³ of the African population and 76%²⁴ of India's population were users of mobile phones. In the future, an increasing share of internet services will be accessed by smartphones. This is an exciting opportunity to increase outreach to those on the ground. By providing the right data in a useful format, reegle appeals to app developers who come up with simple but effective web-based applications to assist in making the right decisions. An example of relevant information in this context could be the details about feed-in tariffs in a given region, which could actually determine the size and type of a solar project.

Another reason why connected datasets are so valuable for sustainable development is that they can increase the grasp of abstract issues. Particularly in the energy sector, information can be hard to understand. Keeping reegle's target group in mind, English will not be every users' mother tongue. This is where context provides deeper understanding, and where semantically linking content is a contribution to sustainable development that Linked Open Data can make.

5 Publishing and Consuming Clean Energy Data

Putting open energy data out there and providing it with added semantics via reegle's data portal increases the impact of knowledge brokers in the clean energy sector. Like reegle, many such portals are funded by public bodies and aim to provide a service for their users free of charge. Such organizations can certainly profit from the integration of clean energy data and other appropriate Open Government Data, but at the same time provide extended services for their users:

- As mash-ups: provide knowledge dossiers on certain sectors by drawing together from several sources and making this information available in one place
- Publish their own datasets as (Linked) Open Data for other organizations to continue making new connections

²³ <http://www.bbc.co.uk/news/world-africa-15659983> (November 2011)

²⁴ Census India 2012, retrieved 2009-11-10 from www.censusindia.gov.in and <http://www.trai.gov.in/WriteReadData/WhatsNew/Documents/PR-TSD-Mar03052012.pdf> for mobile phone penetration.

- Combine previously unconnected datasets to highlight new conclusions, possibly in a visual way

There is certainly a need for more Open Data on energy-related matters, and with more datasets being published, more innovative models of integration will develop. Reegle sees its role as a pioneering actor in the development and usage of clean energy data and will support other data providers who decide to head into the same direction.

5.1 reegle's Available Clean Energy Data

reegle offers all its data in the LOD-standard RDF for external developers; they can easily extract and use for free all reegle/REEEP data through the SPARQL endpoint at reegle's developers portal²⁵.

Reegle datasets are published with resolvable URIs²⁶, in the approved RDF format. There are at least 50 RDF links to other LOD datasets and the entire dataset can be accessed via reegle's SPARQL endpoint.

The reegle dataset is also part of CKAN's²⁷ data hub.

Reegle's data is already used for mash-ups by numerous other players in the field of energy, such as OpenEI, and this can be viewed as a good sign that the information is relevant

6 Outlook

The main drivers behind Linked Open Data are NGOs, governments and the media. At the moment, the media are started to catch up with the idea of providing their readers with semantically related resources, while governments and NGOs see public money best spent by allowing citizens to make free use of the data they produce – and theoretically already own.

Reegle will continue to use and produce energy-related data while at the same time providing the most cutting-edge information to the casual browser in a cost-efficient way. Reegle also wants to ensure a swift uptake of Linked Open Data technologies in developing countries by sharing its own experience with interested parties and supporting newcomers to the LOD scene.

²⁵ <http://data.reegle.info/>

²⁶ e.g. www.reegle.info/glossary/1367/concentrated-solar-power.rdf

²⁷ <http://thedatahub.org/dataset/clean-energy-data-reegle>

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. 5 Star Open data, <http://5stardata.info/>
5. Bauer, F., Kaltenboeck, M., Linked Open Data, The Essentials, 2011 (<http://www.reeep.org/LOD-the-Essentials.pdf>)
6. Alonso, J.M., Boyera, S. 2011 Open Government Data, Feasibility Study in Ghana https://public.webfoundation.org/2011/05/OGD_Ghana.pdf
7. Linked Data – Connect Distributed Data across the Web, <http://linkeddata.org/>