Linked Data in Renewable Energy Domain

Henry Abanda and Joseph Tah
Oxford Institute for Sustainable Development
Department of Real Estate and Construction
Faculty of Technology, Design and Environment
Oxford Brookes University, Oxford, OX3 0BP, UK
fabanda@brookes.ac.uk; jtah@brookes.ac.uk

Abstract: The vision of the Semantic Web is to build a global Web of data that links different resources. To this end, the first step is to put data on the Web in the form that machines can naturally understand or convert to useful forms. This can be achieved through Linked Data technology, an emerging publishing paradigm in which not only documents, but also data are linked over the Web. The ultimate goal underpinning Linked Data is to enable different individuals and organisations to efficiently contribute and share data about different domains. Climate change impacts and/or energy scarcity are among the leading challenges that the international community currently faces. The domain of renewable energy (RE) is widely believed to be an innovative domain that can contribute significantly in overcoming these challenges. Consequently, the institutionalisation of policies/strategies that can potentially lead to the uptake of RE technologies by various organisations is now too common. Employing Linked Data paradigm in managing knowledge provides opportunities for disparate organisations and people to efficiently learn and share knowledge about RE technologies, hence potentially leading to their applications in different projects by end-users. However, given the nascent nature of Linked Data technology, knowledge on how it has been employed in managing RE knowledge is still very limited. In this paper, an exploration of the use of Linked Data in the RE and its relationship to other Linked Data of different domains will be conducted. Given the scope of the study, an exploratory desk-study methodology has been pursued. Furthermore, two Linked Data applications have been discussed.

Key words: Linked Data; Renewable energy; Semantic Web

1 BACKGROUND

The current Web has been successful in providing information about different domains to different people in different regions of the world. Despite this success, it has been plagued by many short comings [Allemang and Hendler 2011]. One can easily get lost in search of information over the Web or does not even obtain the required and precise results. Key to this cause is the overwhelming disparate data being loaded over the Web by different end-users from different terminals. Thus, the idea of linking data over the Web for efficient domain knowledge management has emerged. This concept is generally called “Linked Data”.

Depending on different societal needs different research communities have developed linked dataset. Perhaps, partly because of the promises and opportunities in Linked Data set and hence the Semantic Web, interest in both has widely grown. Different agencies including government departments and renowned institutions have recently engaged in developing Linked Data sets in different fields.
Recently, climate change and energy scarcity have occupied the centre stage of most governments and international organisations. The uptake of RE technologies has been among the top strategies in combating climate change and providing clean energy [Connolly et al. 2011; Deichmann 2011]. While advanced information technology including Linked Data paradigm has been used in diffusing knowledge about different domains such as life sciences, media and publications; little effort has been made towards linking data in the RE domain. Partly, this constituted the rationale of the studies conducted by Abanda [2011] and Tah and Abanda [2011] where they focused on investigating better ways of modelling knowledge about sustainable building technology domain. Furthermore, from an industry perspective, attempts are currently being made in developing some real RE datasets by reputable organisations. However, as earlier alluded to, RE datasets that exist are quite few compared to other fields despite its importance. Thus, more is needed in investigating better ways of developing RE datasets and linking them to other datasets. The objectives of this study are to: a) establish the state-of-the-art of Linked Data about RE; b) determine to what extent RE data is linked to other domain data; c) demonstrate the discovery of knowledge using Linked Data; d) establish a research gap in relation to RE linked data research.

Given the nascent nature of the Linked Data, a desk-study was used to achieve these objectives. The focus has been on the literature sources published from 2006 which corresponds to the time when the Semantic Web Education and Outreach Interest Group of World Wide Web Consortium was formed and entrusted with the task to develop strategies and materials to increase awareness among the Web community of the need and benefit for the Semantic Web, and educate the Web community regarding related solutions and technologies [SWEO 2006], which in fact laid the groundwork for Linked Data paradigm.

2 OVERVIEW OF LINKED DATA

2.1 Linked Data philosophy

In traditional Web, documents are accessed using Hypertext Markup Language (HTML) browsers. The end-user generally pursues the links between HTML pages where a particular resource has been defined. Similarly, data in the Semantic Web can be accessed using Linked Data browsers. Linked Data refers to a style of publishing and linking structured data on the Web [Berners-Lee 2006]. The fundamental idea that underpins Linked Data is that value and usefulness of data are maximised when interlinked with other data. Thus, unlike HTML pages for the traditional Web, data sources are required for the Semantic Web.

Similar to the navigation from one HTML document to the other which exploits the fact that some key words have been tagged and URI (Uniform Resource Identifier) referenced, the navigation between the different data sources over the Semantic Web is based on ontologies expressed in Resource Description Framework (RDF) [Heath and Bizer 2011]. RDF is by far the most mature data model that has been integrated into many enabling Semantic Web technologies. RDF helps in making statements about resources in the form of subject-predicate-object expressions. For example, in RDF, the notion “The PV-module has the colour of blue” is represented in the RDF “subject-predicate-object” paradigm as:

"the PV-module" (subject), “has the colour” (predicate) and “blue” (object)

In RDF, each resource is defined by an URI as well as the different components of the “subject-predicate-object” triples. Thus, Linked Data browsers enable users to navigate between different data sources by pursuing RDF links. This allows the user to start with one data source and then move through a potentially endless Web of data sources connected by RDF links [SWEO 2006]. For example,
suppose an end-user is searching for “aluminium frame system” over a hypothetical Semantic Web. The end-user might be interested in the supplier’s location. By pursuing an “aluminium frame system” RDF link, the user can navigate to information about supplier’s location contained in another dataset. In other words, let’s suppose “aluminium frame system” is contained in the ASC Complete Resource for the United Kingdom (UK) building products [ASC 2011]. The ASC Building Products Directory is one of the largest sources of information on building products and building materials in the UK construction market. It contains information about building products and their respective suppliers. Let’s also assume that the supplier’s information is found in the online Yellow Pages Directory. The Yellow Pages Directory is a resource that contains information about businesses in the UK. By pursuing the “aluminium frame system” in the ASC directory, information about the aluminium frame system and its supplier can be obtained. If by definition “supplier’s location” had a URI and also formed part of the RDF triples, it will be possible to pursue the RDF link to obtained more information about the supplier’s information in the Yellow Pages Directory. It is important to note that this is on the proviso that data stored in the Yellow Pages and ASC directories are in RDF format.

2.2 Some linking open data efforts

Over the last five years, the research community has engaged in the development of Linked Data in different domains. The Linked Data that so far exist relates to the different geographical locations, people, companies, books, scientific publications, television, music, etc. The different Linked Data of various domains and their different relations is represented as a data cloud presented in Figure 1.

![Figure 1. Linked Data cloud [Source Bizer et al. [2011]]](image)

From Figure 1, the different Linked Data sources can be categorised into seven main groups including life sciences, geographic data, media, user-generated content, publications, media and government. From Figure 1, statistics for the main Linked Data groups is presented in Table 1.
<table>
<thead>
<tr>
<th>Domain</th>
<th>Number of datasets</th>
<th>Triples</th>
<th>%</th>
<th>(Out-) Links</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Media</td>
<td>25</td>
<td>1,841,852,061</td>
<td>5.82</td>
<td>50,440,705</td>
<td>10.01</td>
</tr>
<tr>
<td>Geographic</td>
<td>31</td>
<td>6,145,532,484</td>
<td>19.43</td>
<td>35,812,328</td>
<td>7.11</td>
</tr>
<tr>
<td>Government</td>
<td>49</td>
<td>13,315,009,400</td>
<td>42.09</td>
<td>19,343,519</td>
<td>3.84</td>
</tr>
<tr>
<td>Publications</td>
<td>87</td>
<td>2,950,720,693</td>
<td>9.33</td>
<td>139,925,218</td>
<td>27.76</td>
</tr>
<tr>
<td>Cross-domain</td>
<td>41</td>
<td>4,184,635,715</td>
<td>13.23</td>
<td>63,183,065</td>
<td>12.54</td>
</tr>
<tr>
<td>Life sciences</td>
<td>41</td>
<td>3,036,336,004</td>
<td>9.60</td>
<td>191,844,090</td>
<td>38.06</td>
</tr>
<tr>
<td>User-generated</td>
<td>20</td>
<td>134,127,413</td>
<td>0.42</td>
<td>3,449,143</td>
<td>0.68</td>
</tr>
</tbody>
</table>

Source Bizer et al. [2011]

3 WHY THE RENEWABLE ENERGY DOMAIN?

The contribution of RE towards minimising climate change impact and the amelioration of energy poverty cannot be understated. As of 2004 it was estimated that RE contributed about 14% of world energy [UNDP 2004] and reduces enormous amount of greenhouse gases that impacts on the environment. Consequently, many governments are now recommending the use of RE. To this effect, various voluntary and mandatory policies aimed at encouraging the uptake of RE are now common. In the UK, feed-in-tariff and RE obligations are among the leading policies to promote their use [DECC 2011]. Thus, RE market is maturing at an alarming rate on daily basis. In relation to the existence and application of RE, it is interwoven with many other domains. Buildings upon which RE is applied, policies that regulate its use, the involvement of manufacturers and installers in RE, location of manufacturers are domains related to RE in one way or the other. This makes RE a suitable area to investigate its relationship with other domains. The key question is how is RE dataset linked to other domains?

Despite the importance of RE vis-à-vis climate change and energy scarcity, it does not appear as a key concept in the Linking Open Data (LOD) cloud as depicted in Table 1 and Figure 1. On further analysis it emerged that conventional energy and RE data have been captured under the “government” dataset. As in Table 1, the government dataset links to other domains is only 3.84%, indicating how weak the government dataset is linked to other domain despite having the highest “triples” or RDF links. This implies RE has far fewer links than the 3.84% as it is a subset of the government concept. Thus RE is an important and suitable area that warrants investigation with regards to the Linked Data paradigm.

4 OVERVIEW OF RE LINKED DATA

Currently, the LOD cloud in Figure 1 constitutes the best source where RE RDF can be found. On exploring the cloud, the following linked RE RDF knowledge models were uncovered: a) OpenEI - Open Energy Info.: This is a platform that connects world energy data; b) European Environment Agency Published Products: This consists of independent information about the environment; c) Linked Clean Energy Data (reegle.info): This is linked clean energy dataset which is comprised of policy and regulatory country profiles, key stakeholders (organisation profiles), project outcome documents and a thesaurus on renewables, energy efficiency and climate change for public re-use.
While the above RE datasets might be related or are linked, it will be more important to extend the link to include other related concepts. These concepts may be datasets that provide the different stakeholders involved in RE energy business, country in which the RE business is taking place, and even the datasets for cost and emissions from RE systems.

5 RE LINKED DATA APPLICATIONS

The examples of RE Linked Data discussed in section 4 require that information about the different domain be defined in a structured format, i.e. in RDF in particular. But there are already existing information over the current Web which end-users are interested in exploiting. Thus, the development of Linked Data browsers that can be used in acquiring information from the Web is gaining ground. These browsers can be used in the construction of powerful “datameshups” across heterogeneous data source collections without requiring any programming experience [ODE 2012]. Some main browsers are Marbles RDF browser, Zitgist RDF browser, OpenLink Data Explorer previously known as OpenLink RDF browser, Disco Hyperdata browser, Fenfire RDF browser and Tabulator. While most of the Semantic Web browsers have been developed to interface with data sources, the OpenLink Data Explorer and Tabulator are generic browsers that can be used to browse RDF data on the Web [Berners Lee et al. 2006]. In the ensuing section, applications for the browsing or finding RE energy information using Linked Data will be examined. The first application explores structured Linked Clean Energy Data (reegle.info) (i.e. a RE Linked Data example discussed in 4), while the second explores already existing RE information on the Web.

5.1 Exploring structured RDF RE Linked Open Data from Reegle website

RE is a key concept in the Linked Clean Energy Dataset. Its graphical presentation facilitates navigation to other related concepts. The symbol R denotes “related” terms. For example, “renewable energies” is “related” to “energy”. On the other hand N denotes “broader” or “narrower” terms. For example, “renewable energies” is a “broader” term to “renewable energy sources”. Reegle website provides a facility to generate RE RDF format through a single click. An excerpt of the output from clicking “renewable energies” is presented on the right of Figure 2.

Figure 2. Renewable energies Linked Data [Source Reegle [2011]]
Based on the left of Figure 2, it is possible to determine development institutions involved in “renewable energies”. This can be achieved by simple clicks on desired concepts as depicted in Figure 3. First, “renewable energies” concept is clicked. Second, “International Energy Agency” is clicked. Lastly, “development institutions” is clicked and the different institutions involved in RE then emerge as shown on the right of Figure 3. It is important to note that the RDF format of the different concepts in Figure 3 can easily be generated from Reegle’s website.

![Figure 3. Navigating RE information in REEGLE dataset](image)

5.2 EXPLORING RE DATA USING OpenLink DATA EXPLORER

Without any bias, OpenLink Data Explorer, one of the leading Linked Data browsers will be used in exploring RE information from a non-structured or non-RDF website. OpenLink Data Explorer is a Linked Data browser plug-in currently available for Firefox, Safari and Google Chrome [ODE 2012]. It facilitates the exploration of raw data and data relationships that underly a Web page. The plug-in allows end-users to navigate between hypertext and hyperdata on the Web. The installation of the OpenLink Data Explorer has been discussed in ODE [2012], and will not be repeated in this paper. However, it is important to note that for this example, it was incorporated into Firefox. The steps are:

A RE Web page is open from any Web link. In order to call the RE Web page into OpenLink Data Explorer, the RE Web page is right-clicked and “View Page Description” is selected as shown in Figure 4.

![Figure 4. Opening a RE Web page in OpenLink Data Explorer](image)
From Figure 4, another representation of RE object is via the “primary topic”. On clicking on the RE object captured as “primary topic” in Figure 4, Figure 5 is obtained.

**Figure 5.** Data object representation of RE

One main advantage of OpenLink Data Explorer is that it provides various views of any domain knowledge for easy understanding and visualisation. It also provides interfaces that can easily be queried. This is provided through the “Alternative Linked Data Views” as shown in Figure 5. On clicking on the facets link on the “Alternative Data Views” the different RE attributes are presented as in Figure 6.

**Figure 6.** Presentation of RE attributes using Alternative Linked Data Views

### 6 CONCLUSION AND FURTHER RESEARCH

In this paper the rationale for the need of Linked Data in the RE domain has been discussed. It emerged that current research in Linked Data development in the RE field is lagging behind other fields. However, some important RE linked datasets already exist and there is need to extend them so that it can be made more useful for communities interested in RE. While current RE datasets may not be much compared to datasets from other domains, in its current size it is a good starting point for how it can be best explored for Linked Data development and integration.
with other datasets. This entails linking most RE datasets to its related and needed domains so that information can easily be shared and re-used. This will be considered as part of future research.

REFERENCES


