Unleashing Climate and Energy Knowledge with Linked Open Data and Consistent Terminology

—Transcript of a webinar offered by the Clean Energy Solutions Center on 12 March 2013—
For more information, see the clean energy policy trainings offered by the Solutions Center.

Florian Bauer	Operations and IT Director, Renewable and Energy Efficiency Partnership (REEEP)
Jon Weers	Senior Web Applications Engineer, U.S. National Renewable Energy Laboratory

Vickie Healey:	Hello, everyone. I'm Vickie Healey, with the National Renewable Energy Laboratory, and I'd like to welcome you to today's webinar hosted by the Clean Energy Solutions Center. Our discussions today are focused on how opening up and linking data, also known as Linked Open Data, and categorizing the data automatically through use of consistent terms can help dissolve barriers to gaining access in information.

Next slide please.

Before we get started, one important that I mention, that I need to state is a disclaimer which basically states that the Clean Energy Solutions Center does not endorse or recommend specific products or services. However, information provided in this webinar is featured in the Solutions Center's resource library as one of many informative and best practice resources that are researched and provided by our technical experts.

Next slide please.

Before we begin, I'll quickly go over some of the webinar features. First for audio, you have two options. You may either listen through your computer or over your telephone. And, if you choose to listen through your computer, please select the "mic and speakers" option and audio pane so we will eliminate the possibility of feedback and echo. And, if you select the telephone option, a box on the right side will display the telephone number and audio panel you should use to dial in.

We ask that you please mute your audio device before the presentations began. One other thing, if you are having technical difficulties with the webinar, you may contact the "Go to Webinars Helpdesk" at 888 259 3826 and the Helpdesk will be happy to provide assistance to you. We also encourage our attendees to participate in the webinar by asking questions and also providing relevant comments if you have them and you would like to contribute. And, if you would like to ask a question or add a

comment, we ask that you use the questions pane where you may type in your question.

And, if you are having difficulty of viewing any of the materials through the webinar portal, you will find PDF copies of the presentations being presented today at cleanenergysolutions.org/training and you can pull off the PDF slides and follow along as our speakers present. Also, I would like to let you know that an audio recording and the presentations will be posted to the Solutions Center training page within a few weeks so feel free to check that so you have the opportunity to again listen to the presentation or others that aren't able to attend may go back and listen and view the presentation.

We have a great agenda prepared for you today, which will focus on how utilizing open data format and consistent terminology can provide a way to easily access the various multitudes of energy data that currently exists within the public domain. And, the first speaker is Florian Bauer and Jon Weers will begin their presentations. I'm going to provide a very short informative overview of the Clean Energy Solutions Center initiative and following the presentations; we'll have a question and answer session. We'll have a very short feedback poll where we will ask your feedback on the webinar and then we'll wrap up with any final discussion and closing remarks.

Next slide please.

This slide provides you just a bit of background in term of how the Solutions Center came to be. The Solutions Center is an initiative of the Clean Energy Ministerial and is supported through a partnership with UN Energy. It was launched back in April 2011 and it's primarily led by Australia, United States and other CEM partners. The outcomes of the unique partnership includes supporting developing countries to enhancement of resources on policies relating to energy access, we provide low-cost expert policy assistance and also, it serves as a forum for peer-to-peer learning and training tools including the webinar that you are viewing today.

Next slide please.

The Solutions Center has four primary goals. First, it serves as a clearinghouse of clean energy policy resources. It also serves to share policy best practices and data and analysis tools that are specific to clean energy policies and programs. The Solutions Center delivers dynamic services that enable expert assistance, learning and peer-to-peer sharing of experiences. And lastly, the Center fosters dialogue on emerging policy issues and innovation occurring around the globe.

Our primary audience is energy policy makers and analysts from governments and technical organizations in our countries but we, also, very much try to engage with the private sector, NGOs and also civil society.

Next slide.

Now, I'd like to go over what we consider to be a mark key feature that the Solutions Center provides which is our expert policy assistance. We call this service "Ask an Expert" and it's a very valuable service that's offered through the Solutions Center. We've established a broad team of over 30 experts from around the globe who are available to provide remote policy advice and analysis to all countries. And, I'd like to add that this service, this advice, this analysis, is provided at no cost to the requester so if you have a need for policy assistance on anything from micro-grids, smart grids, regulations, renewables, energy efficiency, clean transportation or any other clean energy sector you can think of, we welcome and we encourage you to use this very useful service.

And again, to repeat, this assistance is provided free of charge and to request assistance, there's a very simple process, you may submit in your request by registering through our "Ask an Expert" page feature in energysolutions.org/expert and we also invite you to spread the word about the service to those in your networks and organizations that might have a need or interested in receiving this assistance.

We also encourage you to explore and take advantage of the Solutions Center resources and services including the expert policy assistance, which I just mentioned. We invite you to subscribe to our newsletter, participate in webinars such as you are doing today. We welcome you to recommend to us relevant resources that you feel would be added value to our resource library and we invite you to test and provide feedback on our global renewable energy opportunity tool, which is now going through review and is currently available on our website. If you're interested in testing the tool and giving us your feedback, we welcome that as well.

Next slide.

We also have what we call our "Policy Forum Discussions" and we encourage you to go to our, it's basically a blog page where you can read and comment on blogs which are located in the policy forum page. And on that page, you'll find many interesting and informative articles discussing progress of energy policy development and implementation on foreign countries all around the globe. And, we also follow similar articles posted by our partners at Renewable Energy and Energy Efficiency Partnership, also known as REEEP, that's where Florian Works, and we follow

Leonardo Energy articles and also, we follow podcasts that are developed by Bloomberg New Energy Finance.

Next slide please.

And so now, I'd like to take this opportunity to provide brief introductions of our very distinguished panelists. We're joined today by Jon Weers, a Senior Web Applications Engineer at the National Renewable Energy Laboratory and is one of the lead developers on OpenEI.

Next slide.

And also, I'm pleased to introduce Florian Bauer who's Operations and IT Director at Renewable Energy and Energy Efficiency Partnership, also known as REEEP. Florian is also responsible for overseeing REEEP's clean energy information portal, known as REEGLE, and Florian, you are our first presenter for today. So this time, I'd like to turn the presentation over to you. And, Florian, welcome!

Florian Bauer: Hello! Thank you very much Vickie for this introduction. I'm very pleased to have this opportunity today to talk to you about a topic that is very close to my heart, which is making data and information available to everyone. I'm working since few years in the Renewable Energy and Energy Efficiency Partnership on knowledge management on information dissemination and as part of that, we have developed a tool which is called REEGLE.info, which is a clean energy information gateway.

For the past years, we have moved much to using a technology that is called Linked Open Data and we are also very much focusing on consistent terminologies. Jon Weers and I, we're both trying to bring you a bit closer to these concepts. We'll explain you the details so please don't day time what this is and what it needs, what it can do for you and I will focus on a little bit more later on the take. But let me start with showing you some slides on why we think that data is so relevant. Actually, all of us know that we have the victim on the data. We know that it is very important to have access to data, to get data in a format that it allows us to use it. I have put together some examples on where you usually have to see the use of lots of data.

First example here is something that was developed out from a brilliant project, which is a tool that's online on the web where you can visualize different energy statistics and so on. Another example is that showing how you can use data and make it better accessible and readable for people. Of course, showing data on maps and this is an example from our own REEGLE.info information portal where we show energy statistics on country maps and show the values as part of the map.

4

The other one known I used as an example, where you use a lot of data is, or course, when you want to show complex relationships of different things like this graphic from the US Energy Information Administration that has quite a lot of data in the back. And of course, we all know these very nice "infographics" that you see every day at the moment which are very good to understand and very good tool to show complex things to keep off but, of course, we need to develop those because we need to have access to reliable and machine-readable data to be able to trade those infographics.

Another need, another example that we see is the use of big amounts of data, presently, is when you work with geographic geo-location data and this is an example of showing potentials of solar energy in different parts of Eastern Europe who are in the map and this is another example where, of course, a lot of data is needed in structured form. And with that, I will hand over to Jon who will show a few more examples and then explain you what Linked Open Data is and how it works. Jon, I'll turn over to you.

Jon Weers: Thank you Florian. All right. A couple more examples of Open Data. This one from OpenEI, we use data for collecting underlying analysis and research. This map actually shows coverage of utility rates collected for utilities within United States that then gets filed into a ton of technologic applications and research platforms and analyses. We also use data for profiling, generating big pictures to help fuel analysis in an international level. So a lot of stuff going on, a lot of great uses for data. This is all good but how do we get the data and why do we need Linked Open Data?

That's really the question of Florian and I are going to spend the next 20 or so minutes answering and I really love this quote at the bottom here from IBM last year, "Every day we create 2.5 quintillion bytes of data, that's basically 90% of the data in the world today has been created in the last two years alone." This includes all of data in all the libraries, all the works of Shakespeare, all the ancient texts; we are replicating that volume of information just about every two years. And, I wouldn't be surprised if two years from today, that statement is still true. Our data is expanding exponentially. So knowing that there's this dramatic increase of data headed our way, why do we need Linked Open Data? And, that's what I'm going to answer now.

So this right here is a model of the typical way to store and share data. This is the internet as it is now today, 90% of the websites out there use this approach and we call this as "Cyload Approach". Each website has its own database, feeding information to the website is displayed on the site and then that site is pushed out to the world. And, everything is very vertical. Cyload. This results in a lot of duplication of effort and information.

For example, if the site on the left wanted to use some data from the site on the right, even if that data was open and shareable and downloadable, they have to download a copy of it. Typically, in Excel or, hopefully, CSV format and then they put that data into their database so that they can put back content unto their website and push it out to the world. And, this is great, sharing data is fantastic but this isn't good enough and I'll tell you why.

There's a potential here for the online community to present it with conflicting information and that happens whenever that original site, represented by the number 2 in this case, decides to update their data, released version 2 of that data, and the site that previously downloaded it and installed it, they no longer have the same data. Now, unless they've been very diligent in time stamping and attributing their data and telling where exactly they got it from and when they got it from, which most people don't, the users of that sites are going to be confused.

The users of both sites don't know when and where the data came from so when they go to these two websites, they're going to see conflicting values for a particular data point and they're not going to know which one is newer or which one is more accurate. Linked Open Data, not only solves this problems, it essentially renders it irrelevant. And, I'm about to tell you why.

Before I do that, I'm going to go through the five steps of Linked Open Data. We're going to define what will Linked Open Data is and this comes from Tim Burners-Lee's Five Stars to Linked Open Data and Florian's prepared some wonderful slides here I'm going to read through, illustrating the differences.

As you can see, we slowly move up to staircase from one star, being the bare minimum, to five stars, being the best way you could possibly embrace the idea of Linked Open Data. The first step in the staircase is to make your stuff available on the web in whatever format. You got to get it out there. This is essentially open data. It's an open license. It's free. It's shareable and these are typically PDFs or whatever formats handy.

The next step is just to spend a little more time thinking about it, making it structured, making it reusable for somebody else. If you can imagine an Excel file or a spreadsheet, instead of just a scanned version of the table or a picture of the spreadsheet. If I can download the actual spreadsheet, then I can get in there and start crunching the numbers and doing things with the data. It becomes what we liked to call more 'reuseful'. There's a way to improve upon that though. Use of non-proprietary formats like CSV, instead of Excel, means I don't actually have to own a copy of Microsoft Office before I can use that data. That might seem like a small step but it's pretty huge especially in developing countries where funds are limited.

The fourth step is to use URIs to identify things. Now, URIs are very similar to a URL, except of instead of a link, it's an identifier. And, really the only difference here is a URI is a URL that leads to directly to a thing. If you've got your CSV dataset out there on the web and it lives at a specific place on the internet at an unchanging location, www.mysite.com/mydataset.csv, that's a URI. That's something that I can point to and rely on and essentially have more or less a guarantee that it's always going to be there.

Now, I don't even have to download that dataset, I can actually take my programs or my analyses and point it to that URI and I can count on that being there. Google is famous for doing this with all of their Google Maps APIs, allowing people to do geo-location, mapping services without actually having to build their own map client. That we point our stuff to Google Maps and it displays on the map in the web browser, instead of us having to download all the maps in the world and post them on our own.

And, the last step is to link your data to other people's data to provide context. This is the really the most important part of the Linked Open Data and this is what really enables sharing. When you get to this stuff, you start talking about the semantic web and its all semantics. The reason that they call the semantic web is because the literal definition of the word 'semantic', we want to make sure we're talking about the same things. And, the way we do that is we link our data to other data to provide context.

There's a great slide coming up that's Florian's going to be talking about. It shows a good example of how context can get confused but the other classic example is what we refer to in the States as the Georgia Problem. If we have two websites talking about Georgia and sharing the information about the population of Georgia, we want to make sure that we're talking about the same things. Semantically, we want to ensure that we're talking about Georgia, the country, versus Georgia, the US state. And, so that will link, connecting to other people's content is how the machines verify that's true.

Now we have an idea of how we climb the staircase to Linked Open Data. Let's go back and revisit our model with Linked Open Data and see how this works. With Linked Open Data, datasets are shared behind the scenes. Each site can focus on key data and import supplemental data. You'll see the arrow at the bottom of the graph; the database is shared behind the scenes. It's not like an API or some other technical device where you're sharing data from website to website; you are actually sharing data from database to database.

And, this is really important because as the data on the site on the right updates to version two, because it's a shared database, it's instantly updated to all site that are link to it. The content changes instantly on both

sites. Even more important, those arrows pointing towards the middle, to the sign that says "Other websites", this is where it gets really fun.

Linked Open Data actually enables third-party websites to tap into those databases and do what we call a "Mash Up", combining the databases in new and exciting ways. Innovating content, new research, new ideas, doing things with the data that the data was never even intended for. And, that's in a nutshell while other D and clean energy is still important. There's a need to focus efforts, want to display all the relevant information about the topic, want to provide only the information with subject matter experts for.

We want to reuse existing dataset to avoid replication of work already done. It saves cost. We're not duplicating efforts. We want our dataset to update in real-time so we're delivering that consistent message to all the users of the internet or sharing data and making sure that they are not visiting out-dated sites and giving out-dated data. And, we're moving towards semantic linkages and interoperability, data mash-ups and utilizations never before imagined, fostering innovation.

The SPARQL language that you see at the bottom is this very technical query language for querying Linked Open Data. We don't want to get too technical but I do want to mention that it could span multiple data sources. So in a single query, you can actually query OpenEI or REEGLE, Wikipedia's database which is DBpedia. You could query all three sites at the same time and that probability is what enables these mash-ups. To be able to ask a question programatically, it can only be answered by three or more or five different datasets combined.

And, really those linkages are what put us together and connect us in what Richard Cyganiak calls the "Cloud of Linked Open Data Cooperation". I like to show these slide because the more integrated you are, the closer to the middle you appear on this, partially eclipsed by the call-out pock is DBpedia, which I mentioned is the Linked Open Data backbone of Wikipedia, but you'll also see REEGLE and OpenEI represented there and moving towards the middle.

Let me give you a little bit of background on OpenEI. OpenEI is a collaborative knowledge-sharing platform which is free and open access to energy related data. More the 840 datasets, more than 56,000 content pages, a ton of apps including green button apps which is an initiative we're pushing, a lot going on there. I'd often to refer to OpenEI as a platform of platforms with many different aspects to it. One of them is our semantic media wiki and that's basically the same platform that Wikipedia is working on with Linked Open Data enabled at wiki level. And, one of the things that we do in our wiki is we have these country profiles.

For India country profile, for example, against slightly eclipsed if you look, it's sort on that shadow of the call-out box is creating, you might be able to catch a glimpse of what we're capturing population and GDP. Berneath of the map which, of course comes from Google Maps, is a paragraph describing India in a very high-level context. That paragraph comes from Wikipedia, the population and the gross domestic product or GDP comes from CIA World Factbook. We are importing this information, just like I said a few slides ago.

We don't want to have to be the stewards of this information. We at the National Renewable Energy Lab want to focus what we are good at and that's making maps, doing analysis and building these datasets that fuel decision-making. I don't want to have to update my website every time the population in India changes, for every time the gross domestic product fluctuates and I don't want to have to spend money, time and effort building my own policy and regulatory overview when I know that REEGLE has done such a fantastic job with theirs.

So we have a Linked Open Data partnership or we're pulling in the REEGLE policy and regulatory overview into our site, we're pulling in the population and GDP, we're pulling in the maps, the high-level description information so that our site can be new and fresh and produce a comprehensive overview of each country but all we really have to contribute is what we are good at. That allows us to focus our efforts on what we consider ourselves as subject matter experts of and that's really whether the true value of embracing Linked Open Data comes from.

This is a slide I would get too much into the technical details but this is a slide of behind-the-scenes and how it works. The REEGLE data is easily found and accessible at data.reegle.info and you can find the RDFs, the SPARQL points, once you find these when you have a developer versed in these languages, you can grab the information you want using SPARQL and plug it into your site at the database level.

Like I said earlier, it comes through as raw data. We basically have a direct line to REEGLE's database so that gives me the freedom to format this information in however I want to display in a way that works on the OpenEI site, unlike conventional UPIs or other things, I'm not bound by whatever format another website might choose to present the information in or to release it in. With Linked Open Data, I have access directly to the data as it exists in Florian's databases and that allows me to do all the customization I want of it. So that is Linked Open Data, an overview and in practice on OpenEI. And now, I'm going to turn it over back to Florian to see REEGLE and how they're using it.

Florian Bauer: Thank you very much Jon. So I think that we already showed the very important aspect of Linked Open Data and I think it's the most important

aspect there is that while there are several different data portals with information on energy all around like OpenEI and REEGLE, both have their specialties, both have their own target audience. It can avoid to replicate with others who already did it and Jon showed it a second ago by showing how OpenEI integrated the REEGLE policy and regulatory reviews which is really something that really focusing on in updating and so on and it makes sense to push this information to OpenEI directly via Linked Open Data connection.

So just very briefly about REEGLE, REEGLE is an energy efficiency and climate compatible development information portal. We have 220,00 users per month which is we're not quite remember I think and this was possible by using Linked Open Data and Open Data alone and I will show an example in a second on how we did that. The important thing for me is to mention that we provide all the datasets that we create or we have our expertise on in Linked Open Data format to be able to be reused on other portals like OpenEI.

So I will show you one example on where we, for example, use a lot of open data into patients. We have this so called "Country Energy Profiles" where we put together information on all energy related aspects for each country in the world and this is just an example on Germany but there are all other countries in the world available. There's chance that you don't want to update the population of the country so you get this information from other open data providers. In our case, we get the description of the country from Wikipedia. We get statistics that we put in our own charts from the UN, from EuroStats, from the World Bank. All data providers that provide data in a format that allows it to reuse it.

The beauty of that also is that we can put the datasets into a chart that looks like we want to display it. So we are not bound of any graphic code thing that one of these data providers provide but we can put it in a format that we like and that looks that fits to our website and our portal. Of course, there are other information from REEEP, from our own website, we have our own database stand because we have data from OpenEI on each of the countries. So let me finish with this Linked Open Data from, I think you have rather quite good introduction from Jon on about this.

Second reading for the topic that's related to the topic that is very important for me is talking about standards and consistent taking and context and it's kind of fits. What you see here is a tag cloud from the word chart of the city of Edmonton in Canada. They talked about how the government could move towards opening up things and the tag cloud we present, we used for discussion we are that they have. And, you see at the upper left side, this is the whole data topic that we have described already and now, at the last minutes. On the right side, you see those in very big

topic around standardization, around context, around format and so on. Now we get to these topic a little bit.

Based on the experience that we have with establishing multitude of portals, we know that it is really, really important that of those of really strong need to increase the consistency when tagging and describing planet and energy resources. What does this mean? You have different portals. Every portal provides their own information, their own documents and they tag this information so they describe this information with metadata, and they describe this information with text like a note from blogs and so on. But very often this is not consistent.

So one site has a similar document than another site that is probably in a different way and that's confusing. We think that it's very, very important to ensure the consistency of message that is delivered to the public because if you don't do that, if you provide confusing messages, provide inconsistent messages, this is very confusing for users and this is providing uncertainty. And to do this, to overcome this, this really needs standardization of the used categories and tags when describing information.

So the question for me was "Can we support that with an automated system?" and the easy answer is "Yes, I think we have to." because just doing it manually, it'll never get us to the point that we have consistency because every person thinks different and we describe things in a different way. So we need an automated system that helps us to provide this consistency.

The question now is "Can an automated system do this and what are the other portals and what are the issues with automated systems?" Well, one important thing is an automated system has to understand synonyms and relations and has to understand the context of what this document is about and what it is talking about.

And, that's very important because here, you see a picture of the Internet, a visualization of the Internet, which is called the Internet Map, and just imagine on the left hand side, you have a website with a document talking about 'photovoltaics' and they tag it with photovoltaic. There is another website with the same document but they tag it with PV so obviously this is the same but it's tagged in a different way. And, this is difficult to process. We need to have a system that understands this.

And, the second very important thing is we need a system that is able to deal with this "ambiguition." And, this picture here, shown in the left hand side, a plant on the right hand side, a plant. But obviously, for us, the left hand side picture is a flower; the right hand side is the famous Springfield atomic power plant. The system has to understand what are you talking

about in a document and choose the right concept, choose the right meaning.

So there is a new tool that can help with that and we just released them a few months ago. It's called the REEGLE tagging API and I will show that to you very quickly. The main issue is that very often you have left sides or you have your own information storage that's unstructured data. You have lot of data, you have lot of documents, established papers and so on. And, you need to structure this information. As at with manually, it's very difficult so we want to support it with an automated system.

The idea is that this REEGLE tagging API, API is an application programming interface, turns this data into knowledge, turns it from unstructured data into knowledge. And, how do we do this? The system allows to send a file to this tagging API. This can be done by integrating the tagging API into a website or a web application. So you send the file and this file is analyzed by the tagging API. This is an important step. This is a fully automatic process based on knowledge that is built into the system, knowledge that is built into the system by using a thesaurus that understands relations between terms and understand synonyms, understands the context of terms.

So what does the tagging API do? The tagging API extracts the most relevant terms from a document. It screens the document and extracts the most relevant terms. These terms will be fed back to the user of the tagging API and can be used to structure information. Another thing that the tagging API extracts is geographic location so if you have a document that is talking about a specific region, a specific city, a specific country, the tagging API recognizes this, knows this, and feeds it back to the user.

And, that's very exciting up the most relevant extractions you'll have contextual information so you don't only get the tags that were extracted but you also get the information that we have in our thesaurus along this tags. So for example, the definitions, the explanations, the synonyms links to other available thesauri that have this term 'onslaught'. And finally, there is a way of how you can use this tagging API to better share your documents and also get other related documents from other information providers because the tool allows that your document that was sent to the tagging API is safe in other.. a location of this document is safe in our systems.

So our systems know where it can find it on your website and you extract it turns a site also in the system. and if someone submits a related or similar document, we can suggest to have a look in your website, to have a look in your document, and it goes both ways, of course, so you're on the one hand side came that recommendation for similar documents from

other sites and you, of course, can push your documents to these content to give another way for other sites to get this as a recommendation.

The tagging API is used to ensure the consistency in tagging of information and by basing it on a thesaurus making consistent tagging. It makes the documents better searchable and accessible. If you tag them with consistent terms, it would be better searchable. It will increase your positioning in search engines like Google and so on. It can share the documents to others by using this content pool and you can, of course, receive suggestions for similar other existing documents as I've shown before.

This tool is for free so it's a no-cost tool. You can have a look on api.reegle.info and let me know if you have questions and Jon will now show you on how OpenEI is using the REEGLE tagging API. How and well they implement this tool into their website to take back use from it. So I'll give the presenter remote back to Jon.

Vickie Healey:	Jon, are you still with us?
Jon Weers:	Yes, I messed up on you. Thank you.
Vickie Healey:	No problem.
Jon Weers:	Thanks Florian. I'm going to show a couple of examples of how we use the REEGLE API on OpenEI. The first one I'm going to show is how we took an exhaust spreadsheet of definitions provided by the International Smart Gird Action Network of smart grid and renewable energy related terms and pass them through the API to generate relationships between those terms before we include them in the OpenEI glossary.

We essentially just wrote a small script pushing them through the API using Python, which is a relatively simple programming language. The result returns extracted keywords which we were then able to cross-reference with existing definitions in our glossary and assemble multiple relationships between these terms. You can view the relationships under the related links that were created by the API on the OpenEI glossary.

There's an even better example that I'm going to show you, here. We then took the REEGLE API and use it to provide users with real-time definitions of those key terms by taking select articles on the OpenEI wiki and passing them through the API. We were able to get a list of the extracted terms again cross-referenced with our internal list of definitions and then use that final list to highlight existing terms and here, you'll see 'energy' highlighted with the dotted green line and when a user mouses over that, the definitions pops up automatically from OpenEI glossary. Clicking that link to feel the full definition on OpenEI will take you to the

definition page where you can actually see the link to the REEGLE definition and we complete the beautiful circle of Linked Open Data.

We also use the REEGLE API to recommend related articles of searching similar terms. There's actually several ways we got about doing this while Florian and his team were still developing their content pool, we used the API to suggest key terms from articles within the wiki and we pass those terms to our internal search API and that produced a list of pages on OpenEI. We're now currently working on integrating the content pool functionality so that's sometimes seen the list of related articles includes sites other than OpenEI and allow user to branch out and view content on other members of the pool including REEGLE and some of the other partners in the API projects.

Both of these examples are currently live, you can see them online at openei.org/wiki/Gateway:Wind or you can just go to OpenEI.org and click the big 'WIND'. And, that's all we have for you. I'd like to turn it back over to Vickie or encourage the key to close us out and thank you very much for your time and we'll look forward to answering your questions.

Vickie Healey: Right. Florian and Jon, thank you so much. For that outstanding presentation, I learned a lot. I know several other people on the attendee list probably did as well. I have a quick question coming in and I'm not sure if you could answer this one and the question is "How can you deal with disambiguation?" I'm sorry. "For example, this problem is very frequent in names of cities and geographical locations?"

Florian Bauer: Well, basically, the way how the system deals with that is trying to understand the context of the words in a document. By using a thesaurus, which models the context of these words, the system tries to detect which kind of words exist in this document and how they fit in to this context that is modeled in the thesaurus. Therefore, it is possible for the system to decide which this one word that the system just extracted.

It's this part of a contextual feed of, for example, cities or is it part of something different. So it's really about finding out which related words are also in the document and understanding the context between those words and that is by using a very comprehensive thesaurus which was developed by asking the last years and currently covers renewable energy topics in a geo-efficiency topics, climate change, deforestation, green growth and we're just working on extending it to cover adaptation related topics so that the other fields of knowledge that are already included in the system.

And, an important thing to add to that is it understands five different languages so it is also able to deal with Spanish, French, Portuguese and German documents and because we just have to translate the thesaurus, it

	also understands all the different relationships between the words no matter what of the five languages the documents are in.
Jon Weers:	I'd add to that that the use of Linked Open Data really makes that disambiguation easier because when you connecting your information to other sites, if you can imagine having to choose cities with the same name, it's confusing, if the document is linked to another document that you know is talking about the same city that you want to talk about. There are special links in the Linked Open Data that we hope that you can use that tell machines, "Hey, I'm talking about the same exact thing as this." and when we have those links present, it makes the disambiguation incredibly easy to determine programmatically.
Vickie Healey:	Okay. Great, thank you. Excellent answer. And actually, this is a question from me. We've talked about this at the Clean Energy Solutions Center about how we can incorporate open data and the unique tagging in such as that. How complex is it for someone like in the Clean Energy Solutions Center who has just tons and tons of document and information available through the website? How complex is it to incorporate the open data concept and the API because there's so much information already contained in that particular site?
Florian Bauer:	Jon, do you want?
Jon Weers:	Sure, I'll start on this. This is a great question and this is one of the reasons why we really like to highlight the five stars and the five steps to that staircase. Getting all the way to the fifth step is very complex. You'll need developers that are trained in Linked Open Data technologies and all of the time and money but the first step is incredibly easy and nobody is going to jump to the top of the staircase right after that. So definitely take a look at your content and take a look at the data and information in your site and see what you can do to get to that first step and reap the benefits from that and then move up to the second step and so on.
Florian Bauer:	I would like to add, in terms of how easy it is to use the tagging API in your system. Well, it definitely means that you have a good developer to have a look at it because implementing a tagging API means that a web developer has to do something in your website, has to write a little bit of code. But the way on how APIs are built is that they are very well documented so it's very easy for a web developer to understand what they can do with it and implement it. So the best case is just a few lines of code that has to be implemented into the system.

And, coming back to your question about you already have lots of documents, well, that doesn't make a difference because once you have implemented it into your system, you can send all the documents to the tagging API and you can get the results back for all of the documents |

immediately. So it is not about how many documents you already have, it's really about getting it implemented at one stage and then you can submit all the documents you already have and feed the ones to the API.

Vickie Healey: That's impressive. It sounds a lot easier than I anticipated it would be so it's very good to know on that. We have another question and it just came in. And, the question is, "Are there any sites or efforts to make Linked Open Data for climate change actions of countries such as country profiles on climate change mitigation, adaptation, finance, UNF, CCC, national communications, etc.?"

Florian Bauer: Yes, well, I think its starting point of the other two portals that you have seen, REEGLE and OpenEI, both have Linked Open Data content on the mentioned topics. There are other portals that follow that road of going towards Linked Open Data. We have a lot of positive feedback from different organizations including, for example, World Bank, on this topic. I would like to go back to what Jon said before. Linked Open Data is really deprives them of profit, already starting with 1-2-3-4 star models. This is a very good step forward and we found already a lot of data portals that have the 2-3 star approach like World Bank, the UN, EuroStats, all these big FEO, all these big international organizations. They have seen that this is an important concept and adopted it already so there is lots of data already available in these 1-2-3 stars format and we have to get towards the 5-star in the next years. I think this is a good progress going on.

Vickie Healey: Okay, great, thank you so much. Next question is "Can you provide an idea about the size of data interchange?"

Florian Bauer: Well, it's probably one for you Jon.

Jon Weers: I'm not entirely sure what that question means but I'll do my best to answer it. I'm going to guess that they're asking about the size of data being transmitted back and forth between REEGLE and OpenEI because we're dealing at a database level and on a per request basis. It's very small. You can, using SPARQL language, you have full control over the query you're requesting. You can request the entire dataset all at once if you like. That's very impractical.

It's much more practical to, on each web page load, request only the specific information you need so we're transmitting less than a kilobyte of information back and forth. There are no size limitations in Linked Open Data. We have datasets in house that we've linked for research and analysis purposes that are in the neighborhood of terabytes. It's, again, the reason why we expose them via Linked Open Data was so we wouldn't have to download the entire terabyte dataset all at once.

Taking a more broad approach to interpreting that question, I'd really like to talk for a moment again about that IBM quote. Another article I've read once referred to as the 'Impending Tsunami of Data Coming Our Way". Not only is the amount of data we generate increasing dramatically over two years, but also the size of datasets we're creating are getting bigger and bigger now that we're collecting wind data, for example. It's a multiple elevations at virtually every tower across the US at intervals as short as 15 minutes. We're amassing unwieldy datasets every couple of hours and if you wanted to do analysis and research or even just visualize the wind data coming out of our wind network, you would not be able to download that dataset to your computer and then load it in a program or put it into Excel. It's just too big.

We are actually creating the data faster than you could possibly download it so when you start looking at datasets of that caliber, Linked Open Data becomes the only answer and instead of me trying to put a copy of that dataset on my computer and look at it, I'm just linked to it. So I can essentially, the question that I want to ask, the visualization that I want to run, it's almost is as if I send that question out to the dataset and I just get the answers back. And so with Linked Open Data, we're able to do analysis on these immense datasets but really the only data we're transferring is the answer to the question. So we're transferring tiny amounts of data. I hope that answered the question.

Vickie Healey: It did. I verified with the requester and you answered the question perfectly so thank you Jon. Going through, it looks like I do have one more question. It's just hiding down at the bottom. The question is "How can LOD cut my organization's share if qualitative research outputs such as research papers and policy briefs with a wider audience?" And, second part to that question is "We are already posting our outputs on our webpage in PDF form." Actually, that wasn't it. That was just a little more of a clarification to the question.

Florian Bauer: Well, I think Linked Open Data can effectively help..

Jon Weers: Okay, I'll answer this one. Okay, Florian, go ahead. After you, Florian.

Florian Bauer: I think Linked Open Data can effectively help in getting your information out. Just think back about the example that Jon has shown about our policy and regulatory reviews that we publish. In the first instance, we only published it as PDF which meant every one could download it but it made it hard to reuse.

So now, we are publishing it in Linked Open Data formats that means that Jon is able to push-pull the information that we have on policy reviews. It is in our information that the first instance we're also in PDF and now, it's able to be used in Linked Open Data formats. He is able to put that on his

website and make the style of integration with our service to always get the latest version of it.

And that, of course, increases the outreach of the information that we want to publish enormously because it's not only available on our site but it's also available on OpenEI always in its latest version. I think that's a massive increase of outreach that we can generate because it also spreads to various different websites, for example, that she has take on this policy and regulatory reviews and that's a massive increase of outreach that we achieve just by publishing it in Linked Open Data formats.

Jon Weers: Thanks Florian. Yeah, that's exactly what I wanted to say. Another thing that I would add is that we talked a lot about Linked in data. It's important not to skip over the open part. If you really want to get to a research out there, you want to make sure it's open and reusable and you want to do that by giving it a good shareable, share-alike, free open to reuse, free of attribution license.

You can see this type of licenses in creative comments. They are often referred to as CC licenses or CC Zero, CC One, CC Share-alike, etc. Making your data open allows people, of course, to take your research results and use them in future research and reference them and link back to them. And, the more you get those linkages, the more you start building out that Linked Open Data relationship and the more users of your information you'll get.

Vickie Healey: Thank you so much. Well, one more question that came in and this is a little bit off of the energy topic but very relevant to the discussion we're having. And, the question is "Do you know of the similar development in area of gene frequency and bioinformatics with the data expression has become exceptional?"

Florian Bauer: If I'm right, this was one of the first few things that Linked Open Data was taken up. Jon?

Jon Weers: That's exactly what I'm just going to say. That's where Linked Open Data essentially came from was the Genome Project's medical databases that we're assembled. It was historically one of the first government funded Linked Open Data projects that was very successful and became a model for the rest of us to follow and when we're in the energy sector looking for inspiration, we often look back at the path they chose and the decisions that were made and trying to learn from those.

Vickie Healey: Okay, very good. Thank you so much. This next question, I'm not sure if this is something, I'll go ahead and toss it over to you but I'm not sure if it's something that you can answer based on the topic. But perhaps through the open data process and API, "Is there a way to fully engage people from

developing countries.." I'm going to pull these back and I'll send this to you privately because I'm not sure if this is a question that is relevant to the topic that we're discussing so I'm going to consider this and I'll get back to the person who address the question separately.

Florian Bauer: We'll just have to say a quick word on that.

Vickie Healey: Okay.

Florian Bauer: And, that's the Linked Open Data, the tagging API and all these stuff is really, really all boils down to sharing information and getting it out there and ultimately, increasing outreach so that's absolutely going to help people in developing countries whether directly or indirectly.

Vickie Healey: Okay. And, I'm just going through to make sure I did not miss anyone or any questions and it looks as though we have.. One more just comes in as soon as I say that but this is terrific. I'm glad to see the response from the audience. The question is "By chance, would it be relevant if on data regarding the funds or investment operating in clean energy or renewable energy specifically the fund holdings and performances? Is this something that you can address or perhaps a question that I could take offline?"

Jon Weers: I'm not sure if I understand the question. I'd be happy to answer the amount of investment we put in Linked Open Data but I'm not definitely qualified to answer the overall investment to clean energy.

Vickie Healey: Yeah, that's understandable. So I believe, at that, Florian and Jon, that wraps up our questions and I really want to thank you again so much for the outstanding presentation and we have had some really good questions from the audience and I appreciate the audience asking those questions. So I just based upon that, Jon and Florian, if you have any additional thoughts or closing remarks you'd like to make before we wrap up the webinar. You're welcome to provide those, at this time.

Florian Bauer: I think I'd would just like to say thank you very much for listening to us and if you have any questions regarding those topics, Jon's and my contact details are on the last slide of our presentation which is online already, I think, on the Clean Energy Solutions website. So just come back to us and we'll both be happy to answer any of your set of questions.

Jon Weers: Once again Florian says it best. Thank you all for listening.

Vickie Healey: Thank you and thank you guys so much for presenting today. And so with that, before we wrap up, I'd like to just, Heather if you could go to the first polling question, we're going to ask three quick questions to the audience to obtain your feedback on today's presentation. Just to let you know, these evaluation questions are very important to us and it allows our audience to inform us on what we are doing right and areas where we can

improve so we take your recommendations and your thoughts very seriously and we'll just take a few seconds to consider and answer each question thoughtfully.

Heather has displayed the first question on the platform. So the first question, "The webinar content provided me with useful information and insight." and if you could just take a few seconds to respond to that question, we'd very much appreciate it. Okay thank you. With that, we'll close the first question.

And, our second question to the audience is, "The webinar's presenters were effective." Okay thank you and with that, we'll close this question and move to the third and final.

Our final question is, "Overall, did you feel the webinar meet your expectations?" Right, thank you and we'll close that final question. And, thank you to the audience, to those of you who responded. We really do appreciate your feedback.

Next slide please.

So with that, on behalf of the Clean Energy Solutions Center, I'd really like to extend a hardy thank you to Florian and Jon and also to our attendees today for participating in the webinar. We've had a terrific audience and we very much appreciate your time in attending. And, I also invite our attendees to check the Solutions Center website every the next few weeks if you would like to view the slides and listen to our recording of today's presentations, as well as, previously held webinars.

And, just to let you know that the slides, the PDF copies of the slides of today's presentations will be available right after the webinar. I was informed by one of our attendees that it's not posted yet and I apologize for that but they will be posted prior to the end of today.

Additionally, you'll find information about upcoming webinars and other training events and also we invite you to inform your colleagues and those in your networks about the Solutions Center Resources and Services including our no-cost policy support that I spoke to earlier in the webinar. So with that, I wish you all a great rest of your day and we hope to see you again at future Clean Energy Solutions Center events. And, with that, this concludes our webinar. Thank you!