The Need for Crowdsourcing Energy Data

Posted by **Rembrandt** on December 18, 2012 - 12:00am

This is a guest post by Andreas Ligtvoet a PhD. Researcher at TU/Delft department of Energy and Industry in the Netherlands. Andreas is a contributor to [EniPedia (energy wiki),](#) a site that his colleague [Chris Davis created and maintains.](#)

The effort to get a better grip on peak oil runs into the problem of data availability, data accessibility and data quality. As most TOD readers will recognise, there seems to be data asymmetry between the oil producers (NOCs and IOCs), international energy agencies, and the general public. Some transparency has been achieved by streamlining and organising data collection, e.g. through the [JODI initiative](#). However, this encompasses top-down data collection that runs the risk of being polluted by non-data-driven incentives (the political need to over- or under-report, for example).

Click below the fold to read this discussion on the need for savvy open-source data gathering and to learn more about [EniPedia.](#)

There have been a lot of bottom-up attempts to collect and combine data, many of which have been reported in TOD. This high-quality information is often represented in such a way, that does not allow others to build upon the work. Excellent collective efforts like the [megaprojects taskforce](#) on Wikipedia seem to have died. One of the reasons for this demise could be the lack of analytical power the current setup of Wikipedia allows: timeline data, conflicting sources, and large sets of relatively unworthy facts (e.g. location of wellheads) are not handled well. But also the Wikipedia community may not be aware of the notability of energy data.

This article is an attempt to re-ignite the spark of combined cognitive efforts in the TOD community, to spend some of your mental surplus on checking and updating facts and figures on oil projects and production, a topic you are passionate about to begin with. Why do it? Because credible data underpins a well-informed debate. Even commercial databases that cost loads of money and are inaccessible to the general public face the problem that they are full of mistakes, because the time and effort required to update them is enormous. Many of the contributors to TOD discussions have access to excellent data and spend time analysing it. A combined effort is possible (as projects in other domains have demonstrated) and arguably leads to better curated information.

Quality control of crowdsourced data

Quality control is naturally an important issue. Who do you trust: Encyclopaedia Britannica or Wikipedia? One could argue that only paid experts provide the necessary comprehensiveness, accuracy and oversight. However, paid experts' time is limited and possibly biased. TOD community has ample well-informed members who can check facts & figures. That' s what TOD is all about to begin with! There is a lot of collective cognitive energy being spent on TOD, and it would be beneficial if we could channel this more effectively so that we can leverage each other's efforts instead of repeating them. We also notice that the cognitive energy of experts is often wasted, as they find it easy to point out problems, but there are not always systems in place that

permit them to contribute their knowledge to improving the data.

There are several ways in which data quality can be managed, based on the type of systems that are used. For example, wiki systems employ a revision control system that records who did what when, meaning that every edit is logged, and mistakes can easily be reverted. This means that it is harder to vandalize a page than to fix it. Without this functionality, Wikipedia would have failed a long time ago. There are also systems in place like ScraperWiki , which allow people to write scripts that gather and clean up data from across the web (example from oil megaprojects) and then create different views that enable people to visualize the contents of the data. Having different means of visualizing and interacting with the data is key in order to expose errors in the data. Some issue may be spotted using a map or table, while for others, a more in-depth statistical analysis may be needed. Overall, this is about collectively building the modules that allow different people to contribute different steps in a process of gathering, analyzing and improving data.

Enipedia, an example of a crowdsourcing data platform

Of course this is no easy task. We argue, however, that a relatively large group of motivated individuals can curate data more effectively than one or two paid professionals who have to wade through tens of thousands of data points. To show some interesting examples of what can be done with open data, in particular OpenStreetMap, ScraperWiki, and Wikipedia, we have set up a portal on our own Delft University of Technology site Enipedia.

The project uses open source semantic software that is readily available. However, to our knowledge it has not been used in an online attempt to gather, curate, and display scientific data on energy infrastructures. It opens up possibilities to a community of interest that was until now unavailable in a dynamic fashion: to contribute to each others' work and to critique and improve the available information.

Why did we not focus more specifically on e.g. nuclear power plants in Germany? First of all because such a limited set can still be handled by single (research) organisations. Because more is different: no one entity could seriously claim to know all the details or have the manpower to find out all the required details. Because we don't know where the interest of the community lies: our largest non-colleague contributor seems to be French; (s)he may not have participated if the project were about Germany. Because we think people need this broad information to make useful policy decisions.

We most actively work on information on power plants, combining e.g. the Carma.org, eGrid, E-PRTR databases that each have (limited) information on these infrastructures. For a list of all the data sets we used, see Energy and Industry Data Sets

Figure 1 - Power Plant map of Europe from Enipedia

To provide an example, overviews and insightful maps such as the one above can be generated, which depicts all power plants for France. This is a picture taken from our interactive map where you can obtain an overview of country by country power plants in existence. For some more examples, check out http://enipedia.tudelft.nl

The tools for more productive cooperation are available. There is a large and perpetual need for people to check and update the information regularly. The more computer savvy enthusiasts can design new analyses.