Interactive comment on "OWFgraph: A graph database for the offshore wind farm domain" by Erik Quaeghebeur et al.

Erik Quaeghebeur et al.

e.r.g.quaeghebeur@tudelft.nl

Received and published: 11 July 2019

In this author response to reviews, we will first give a general response that contains the most important or impactful reaction. The aim is to address the main issue pointed out by the reviewers. After the general response, we will address remaining issues, which we also judge to be important, per review on a point-by-point basis.

Before starting, we would like to thank the reviewers for their efforts and also for their openness. Their willingness to discuss our paper and their reviews at WESC 2019 has helped in forming our thoughts for the revision. The setup of the review process was an enabler for this, so we would like to encourage the journal editors to continue with the current approach used.

1 General response

Both reviewers are positive about the work we have done and find it valuable.

We thank the reviewers for so explicitly sharing their appreciation for the work. This provides encouragement for the revision we need to do and for continuation of this research line.

Both reviewers indicate that they think the presentation of the paper is currently not ideal. They miss focus on the most important aspects and feel that readers may not fully appreciate the relevance to the wind energy community, perhaps also discouraged by the paper's length. Reviewer 1 suggests splitting the paper in two parts, one part presenting the general idea for a broad audience and a second part presenting practical applications. Reviewer 2 indicates that much of the material in Sections 2 and 3 is 'documentation', which we interpret as being distractingly detailed.

We take seriously the evaluation of the paper's presentation by the reviewers. Therefore, for the revision, we plan to indeed split the paper into two parts. The first part will give a general overview and the second part will go into detail. We will precede both parts with a general introduction where we make clear to the readers what the aim is of each part and where we will make sure the paper's structure is also clear. We think that this will result in a revised paper that addresses this main issue pointed out by both reviewers.

2 Point-by-point replies

2.1 Review 1

- The authors should clearly state that use of the graphic database is a general tool applicable to any wind farm project. The environment can be defined by a further property (metadata) listing possible terrain types, e.g. as done by Sempreviva et al 2017, property "external conditions" containg the attribute "offshore", Complex terrain, coastal etc.

  We agree that the tool is indeed more generally applicable than offshore wind farms (our specific interest). Therefore, we will make this clear in the paper and plan on renaming the database to not be offshore-specific anymore. We will also make sure readers understand the current focus on offshore wind farms.

- [C]oncerning the Conclusions. Expert elicitation is still a practice for professional taxonomists. To avoid biases, the group of experts must include the right choice of experts. However, I am not sure what the authors mean stating that OWP-Graph is not biased. Is it not designed by experts?

  Thanks to the file with detailed comments, we understand that the reviewer's concern is about our statement that the approach we suggest for defining disciplines using the database could reduce bias. We agree with the reviewer's sentiment that experts provide the database's content, but still back our statement, as the database's contents can provide concrete backing for the disciplines defined. To address the concern, we will clarify this part of the text.

- Figures: In my opinion, the colors of the circles in all figures are too dark, I suggest considering pastel colors with a black font.

  We understand that the figures can be difficult to read for some readers. However, we have already done quite some effort to ensure legibility by making the font as

large as possible and paying attention to contrast. We do not feel much can be gained by further tweaking, unless the figures would be made larger still, which would blow up the page count. We have ensured that the figures are vector-based and so the can be zoomed without loss of quality in the electronic (pdf) version of the paper. We hope this is sufficient to address this comment.

- Other comments on the language, poor sentences and request of clarification are in the attached PDF.

  We thank the reviewer for their effort to signal issues concretely in the text. We will consider each of the language, style, and text-structure suggestions and follow those that we agree with. Each request for clarification will be acted on. We will consider the highlighted pieces of text without explicit comments as potentially unclear parts and deal with them as such.

2.2 Review 2

2.2.1 Issues

- Size of the graph should be provided in the abstract or introduction.

  This will be done.

- Publishing the graph in the form of an attached .zip file is not the best way to start building a community around it. My suggestion would be to use GitHub and publish the repository via Zenodo. GitHub would work well for tracking changes proposed by the community in the .json file, while Zenodo would make it citable and would keep track of future releases (supports versioning).

  We understand that the reviewer did not access the on-line database itself because it was not clear to them how to access it. The on-line version is and will

remain the main location for consulting and working with the database. However, we will follow the suggestion of the reviewer to make database snapshots available as citable JSON files, for example using Zenodo or an institutional data repository. We will consider the option of allowing community members to provide input through GitHub.

- While none "phenomenon" nodes have been defined yet, it would still be useful to provide some examples of what concepts you would like to see there in the future.

The reviewer is right that this would help clarify its role. We will seriously consider adding such an example, but may refrain from it in case we are not yet satisfied with the examples we come up with also in how they connect to other nodes.

2.2.2 Comments

- The variable names don't seem to follow any existing naming convention. By defining over 800 of them it seems like an attempt to create a yet another standard, which I believe to be more harmful than helpful when it comes to developing standards and work on tool interoperability. This could be addressed by defining a new relation for variables "same_as". The mapping between naming conventions could be stored in a separate .json file (see next point), for optional use.

The node naming is indeed our own, where brevity was usually a big factor. The reviewer is right that using and promoting existing naming conventions would be good. We will at least add a remark about this to the paper. Actual implementation in the database (as an extra data field with convention/name pairs or by adding extra nodes for standard names) will be put high on our to do-list. The resources and tools created by the reviewer will serve as inspiration.

- Dividing the .json file into smaller ones would make it easier to work with and

to trace changes. The division could be made by discipline, or it could separate attachments to the core graph which might not be of interest for the whole community (institution specific nodes like "ECN install").

Given that the database is accessible on-line using an interactive web interface and can also be accessed through APIs, we do not plan to do this. Others can export any subset of the data as they see fit.

- APPEARS_IN relation feels like an unnecessary replacement of input and output relations, probably created for compatibility with the visualization software. If possible, I would consider using separate input and output relations or renaming, for clarity APPEARS_IN to INPUT_OUTPUT_TO.

The APPEARS_IN relation was consciously created without consideration to visualization. We will consider a renaming, but currently think it unlikely that we will follow the reviewer's suggestion. The rationale behind the APPEARS_IN relation was that some models deal with inputs and outputs implicitly, without clear distinction. Instead all variables APPEAR_IN the model, and the user may choose which are inputs and which are outputs.