

Review for: “Open Chemistry, JupyterLab, REST, and Quantum Chemistry”

Roberto Di Remigio¹

¹UiT – The Arctic University of Norway

September 2, 2020

Abstract

The manuscript describes the features of a platform developed by the OpenChemistry consortium that bridges many prominent technologies in data sharing, analysis, and visualization. The platform unifies a web-based GUI with few computational backends and can be used to visualize and analyze pre-existing data or newly produced computational results.

The development of this platform is timely and the description in the manuscript is clear and compelling. I recommend the paper be **published with minor revisions**, suggested in the following.

I hereby give permission to publicly associate my name to this referee report.

Editor’s Note: *Accepted version of this manuscript* <https://doi.org/10.22541/au.158687268.81852407/v2>

1 Referee Report

I would first like to thank the authors for describing their work in a language approachable also to those that are not steeped into novel web-based technologies. The submission is timely and is extremely well-suited for publication in an interactive format. Finally, I couldn’t agree more with the last statement in your conclusions: *As a community it is important to embrace open source, open data, open standards, and open access to reproducible research.* I think your work does represent an important step forward in this direction.

This said, I have some questions and suggestions which the authors should consider in revising the manuscript:

- The use of the Chemical JSON format seems essential for the inner workings of the platform described in the manuscript. It is my impression, however, that the format is not as widely known as it should be. Are there standard examples in your repositories of how to write a Chemical JSON file from commonly used compiled languages, such as C++, C, and Fortran? Are there any quantum chemistry codes that can already emit their output in this format? Or is there an intermediate Python layer that translates from, *e.g.* a checkpoint file, to Chemical JSON?
- Is there a formal standardization process for the Chemical JSON format in place? Who is participating? Could you describe the workflow used in the definition of the open standard?
- How widespread is the adoption of Chemical JSON so far? Could it be merged with the QCSchema efforts of the MolSSI?
- Are there any limitations to the format? I could think that storing basis set and MO coefficients information for very large molecules would make it rather impractical. Is this the case? If yes, how do you plan to solve this problem?
- What about QM/MM simulations? How would the format need to be extended, if at all?

- Are there any plans/thoughts to support output from non-GTO-based codes? For example, numerical basis sets, plane waves, multiwavelets.
- Are there guides available on how to deploy the platform on local HPC infrastructure?
- The MolSSI QCArchive initiative has an overlap of functionalities with the platform described in this paper, or so it seems to me. Would you clarify the relationship between your platform and QCArchive? What are the use cases for which your platform is specifically designed? Could the platform benefit from integration with QCArchive? Would QCArchive benefit from integration with your platform?
- I think the manuscript would benefit from a short description of the code development workflow used by the developers.
- The authors should put more emphasis on the fact that quantum chemical program packages in the backend are accessed through Docker/Singularity/Shifter images. Containers make it possible to share the code used to generate computational results without violating licenses. Even without the (undue, in my opinion) barrier imposed by licenses, some authors find unpleasant any obligation to share their research code. The use of containers removes a significant barrier not only for reproducibility, but also for collaboration. This is, to me, an extremely compelling feature of the platform and I think it deserves to be highlighted more in the manuscript.

2

2.0.1