

Atmos. Meas. Tech. Discuss., referee comment RC1
<https://doi.org/10.5194/amt-2022-67-RC1>, 2022
© Author(s) 2022. This work is distributed under
the Creative Commons Attribution 4.0 License.

Comment on amt-2022-67

Anonymous Referee #1

Referee comment on "SMEARcore – modular data infrastructure for atmospheric measurement stations" by Anton Rusanen et al., Atmos. Meas. Tech. Discuss., <https://doi.org/10.5194/amt-2022-67-RC1>, 2022

The manuscript introduces the SMEARCore data infrastructure framework, a collection of modular programs and workflows aiming to collect raw data and monitor ongoing processes, do some analysis and give access to raw data in a common format.

It is presented as a help for the operator to monitor the station and do some data analysis.

The concept is interesting and worth bringing it to the attention of the atmospheric science community, specially the part allowing the collection and monitoring of a measurement station.

However the manuscript needs some substantial rewriting. The general plan is not presented clearly enough.

The abstract states that the processing is under the control of the data owners with a focus on the station level but later it is mentioned that data postprocessing can be done outside the station, and in the conclusion the authors mention multiple stations supervised by the same people. Data harmonization is emphasized which is a very good idea, but in that case the data owner can not really do whatever she/he wants anymore. Data harmonization is a key concept in networks like ICOS where the data processing is centralized leading to the harmonization. The SMEAR concept endorses European infrastructures like ICOS but the authors claim their concept to be different ; however in the conclusion they suggest networking stations together, cross-referencing their data and sharing storage between the stations.

Another point which needs clarification is the use of SMEARCore in the frame of campaigns as it needs some hardware resources.

Specific comments:

L18: why a faster installation of new measurement will allow a station to benefit from the experience of SMEARCore ?

L 33-39: the paragraph focuses on important amount of data and big data but this is not related to the accuracy of mass spectrometry. The raise of number of stations implies indeed more data but not at a single station and SMEARCore focuses on the station level.

L 55: big infrastructures are by default not thought to be interconnected with each other, it is a plus when they do. Presenting the lack of coordination between them as a default is not correct. It is true that processing data can be labor intensive and lack documentation, mentioning this to highlight SMEARCore features is fair but saying that the large infrastructures do not automatize, trace and document their process is incorrect.

L 101: point two is more a plus than a default requirement.

L 121: indeed but there is no conclusion related to SMEARCore features.

L 123: « in a conceptual framework » would not be better said as « conceptually »?

L 138: the required procedure corresponds to the workflow.

L 139: one workflow or branching workflows?

L 150: figure 1 legend, are each box a workflow ? If yes it would be to mention it.

L 194: column form time series data is not clear.

L 200: maybe « raw data » will be more clear than « data itself »

L 204: « This way it is ... » is not clear, something like « Multiple views allow to get » may be easier to understand.

L 219: « restrict instruments so they can access only their own data.» is not clear.

L 222: Apache Airflow is only use for the analysis workflows, what about the others?

L 252-253: « SMEARest » SMEAR Estonia? Grafana processes the metadata? The whole sentence is not really clear.

L 262-274: too long for the purpose of the manuscript.

Figure 3 legend: the information of the custom plugin is interesting and it would be more appropriate to move it to the core of the manuscript.

Figure 5 and 6 are redundant, one would be enough with the legend specifying that inorganic data can be presented the same way.

L 306: remove « but », an email is also sent of the operator.

L 318: remark on SMEARCore, it should trace the removal of the instruments from the station in order not to display false status.

Figure 8: may be it can indicate how far in the pas the data are available.

L 338: « parallel implementation » is not very clear.

L 346: « running the workflows as graphs in Airflow » sentence is a bit to technical. Maybe something like « the workflows in Airflow are defined by graphs ».