



Virtual workshop on requirements for a fast and scalable evaluation workflow

**Parallel breakout groups -
Reproducibility and provenance**

***18-19 May 2021
Barcelona***

- Reproducibility and provenance in climate data evaluation:
 - Come back to questions from previous presentations
 - Standards for the input data
 - CF conventions is one of the requirements but more is needed, especially for observations/non ESGF data→ Force providers to adapt to a standard? Potentially talk to Copernicus CDS or ECMWF. To take into account that the modelling community is not the only community using observational products, and that their needs may be conflicting with the needs of other communities. Efforts are needed in order to have a common access to observational products formatted properly that could be used for the whole community. Possible opportunity to work with ESA, as they have become the CMIP tender office.
 - Recording of provenance information: structure and formats
 - Used in esmvaltool to generate the citation information.
 - Some work done on provenance done in Clim4impact and in DARE H2020(with PROV-O standard, also used in DARE H2020). In DARE, the provenance info was generated by the workflow itself. For now only the processing steps are listed.

- Reproducibility and provenance in climate data evaluation:
 - Recording of provenance information: structure and formats
 - Provenance is key but people (scientists) are not fully aware of where to find the information. To increase reuse, the necessary information would be to take the provenance info and re-run the diagnostic. Having a nice viewer for the information would be a good feature to have scientist be more interest in provenance information. IPCC people really use provenance information and it would be interesting to share standards to work with this type of information. Try to find a more comprehensive approach to go from the information to the visualization.
 - Cache with the history of the results of previous analysis to avoid having to re-compute an already existing diagnostic if it was computed by someone else recently. The mapping between the “recipe” and the results is done through the provenance information.

Have a nice viewer to show provenance. An standard is needed.

automation of storing of the generation of the provenance info is needed (done in esmval, not in all tools)



This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N°824084



Our website
<https://is.enes.org/>



Follow us on Twitter !
@ISENES_RI



Contact us at
is-enes@ipsl.fr