



Virtual workshop on requirements for a fast and scalable evaluation workflow

BoG2: Data processing machines

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



- Objective: list user requirements for data processing machines
- Users' experience:
 - Machine access
 - DLR is using the DKRZ Infrastructure. Train users to use these services. Users are not aware of all the services.
 - Online documentation and regular talks
 - Every user has their own specific needs
 - NLESC is also using DKRZ.
 - Is difficult to know how/who to get access in those infrastructures (national/local level, is a local partner needed).
 - What to do with users who don't have a project behind?
 - CERFACS
 - Access to many systems (JASMIN, METEOFRACTANCE, own data processing server - downloading the most usual datasets, NCAR systems, Climate4impact to download specific data).
 - How about moving from different systems? Documentation in internal websites, monthly team meetings to get people on board.



- Deploying software
 - Relying in conda environments.
 - What to do when is not available? Using containers is an alternative.
 - Some problems when using these solutions and filesystems.
 - Default modules available (user friendly way to start working). Provide an initial environment (with documentation and example) for the user to start.
- Training
 - Jasmin offers training quite often.
 - BSC offers training when a tool is released.
 - Training should always be asynchronous (users should be able to do the training alone).
- Cloud
 - KNMI is doing a big change (in terms of offering computational services within the institute)
 - Dedicated cloud accounts (AWS). Team: cloud platform support. You need a support team to deal with this change. Pushing researchers in getting familiar with cloud services.
 - Young researchers are asking sandboxes with data to work on.
 - Applying agile methodologies. Always have a peer to ask questions. Professional software development → encouraging good practices



- Sharing developments
 - Using solutions like Confluence to gather and share knowledge.
 - How to share this information between research centers?
 - Share pre-computed data between teams
- Incremental backup
 - This is needed when you are the source of the data for other users



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