

Optimisation in IS-ENES2

Context of optimisation in IS-ENES2

Summary of Is-ENES2 requirements for optimisation

Detailed requirements

1. Related to your answer to the generic question 7 (What part of your RI needs to be improved):

i. What does it mean for this to be optimal in your opinion?

- Easy, standardized interfaces for command line usage as well as portal integration ,
- Faster, more robust and fully automated replication procedures. Fast replication across continents is key to accelerate data access at an early stage of a major project.
- Policies etc. for assignment of compute resources to user (groups)
- Funding for community computing resources.

ii. How do you measure optimality in this case?

- Are there any existing metrics being applied? RIs have a set of KPIs that will progress if those areas are improved.
- Are there any standard metrics applied by domain scientists in this discipline? End user satisfaction. Number of publications should progress faster than before if we progress following those directions.

iii. Do you already know what needs to be done to make this optimal?

- Is it simply a matter of more resources, better machines, or does it require a rethink about how the infrastructure should be designed?

A rethink is necessary on one hand to get the end users ("data analysts").

iv. What would you *not* want from an 'optimal' solution? For example, maximizing one attribute of a component or process (e.g. execution time) might come at the cost of another attribute (e.g. ease-of-use), which ultimately may prove undesirable.

Due to the amounts of data, we would not like to lower network performance. Also fundamental is the "ease of use" of the RI by scientists and by engineer.

2. Follow-up questions to answers from other sections which suggest the need for the optimization of certain RI components.

Data citation is currently not an easy task because our data collections are extremely complex. We can progress along that line.

3. Do you have any use case/scenarios to show potential bottlenecks in 1) the functionality of your RI, for example the storage, access and delivery of data, doing processing, handling the workflow complexity etc. 2) ensuring the non-functional requirements of your RI, for example ensuring load balance in resource usage etc.

Ensuring load balance when computing services will be made widely available will be a challenge. Network resources are also a potential bottleneck because of the data volume we are dealing with.

4. To understand those bottlenecks:

- what might be the peak volume in accessing, storing, and delivering data? Previous project (CMIP5) had up to about 10 TB over all (mainly 3) European nodes daily. We expect CMIP6 to show significantly higher values.
- what complexity might the data processing workflow have? We presently need to handle rather complex workflows.
- Are there any specific quality requirements for accessing, delivering or storing data, in order to handle the data in nearly real time?No.

Formalities (who & when)

Go-between	Yin Chen
RI representative	Sylvie Joussaume < sylvie.joussaume@lsce.ipsl.fr > Francesca Guglielmo < francesca.guglielmo@lsce.ipsl.fr >
Period of requirements collection	Oct -Nov 2015
Status	Completed