www.bsc.es

Barcelona Supercomputing Center
Centro Nacional de Supercomputación

The EUBrazilOpenBio-BioVeL Use Case in EGI

Daniele Lezzi – BSC

EGI Technical Forum 2013 - Madrid

EU-Brazil OpenBio
EU-Brazil Open Data and Cloud Computing e-Infrastructure for Biodiversity

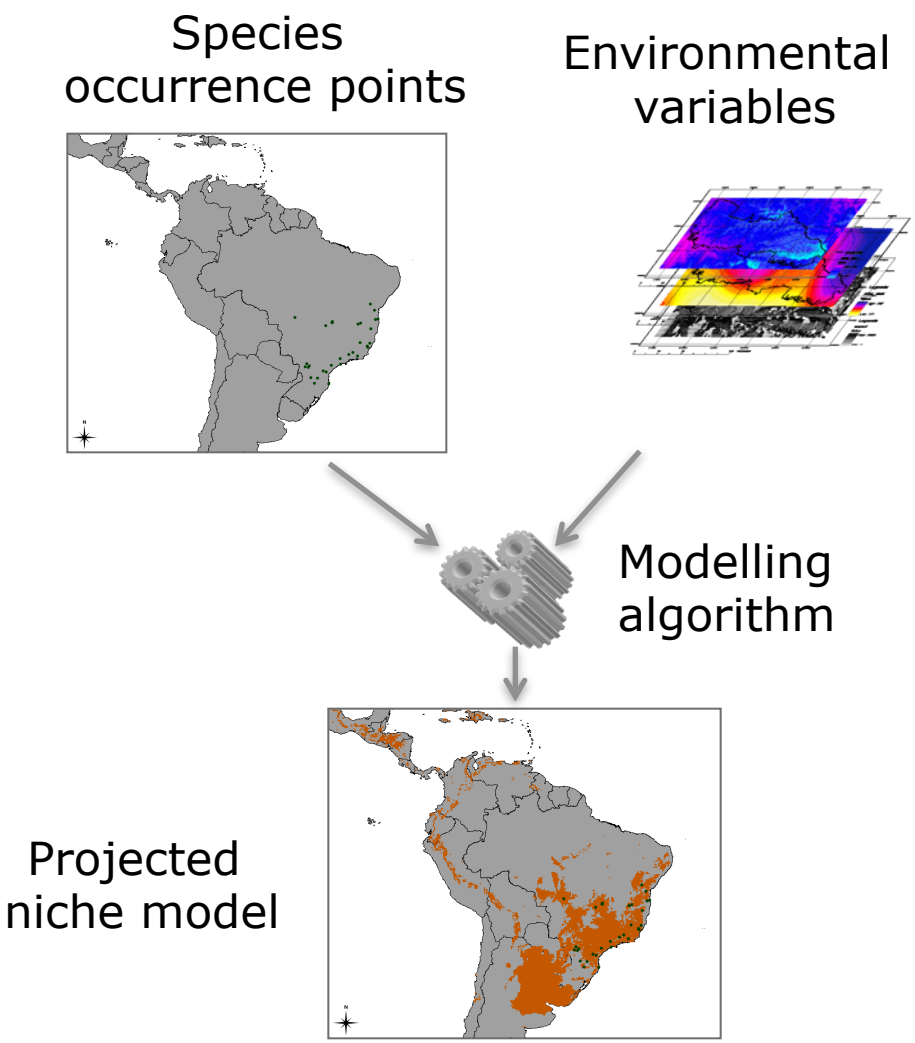
BioVeL

The Use Case: Ecological Niche Modelling

Ecological niche:

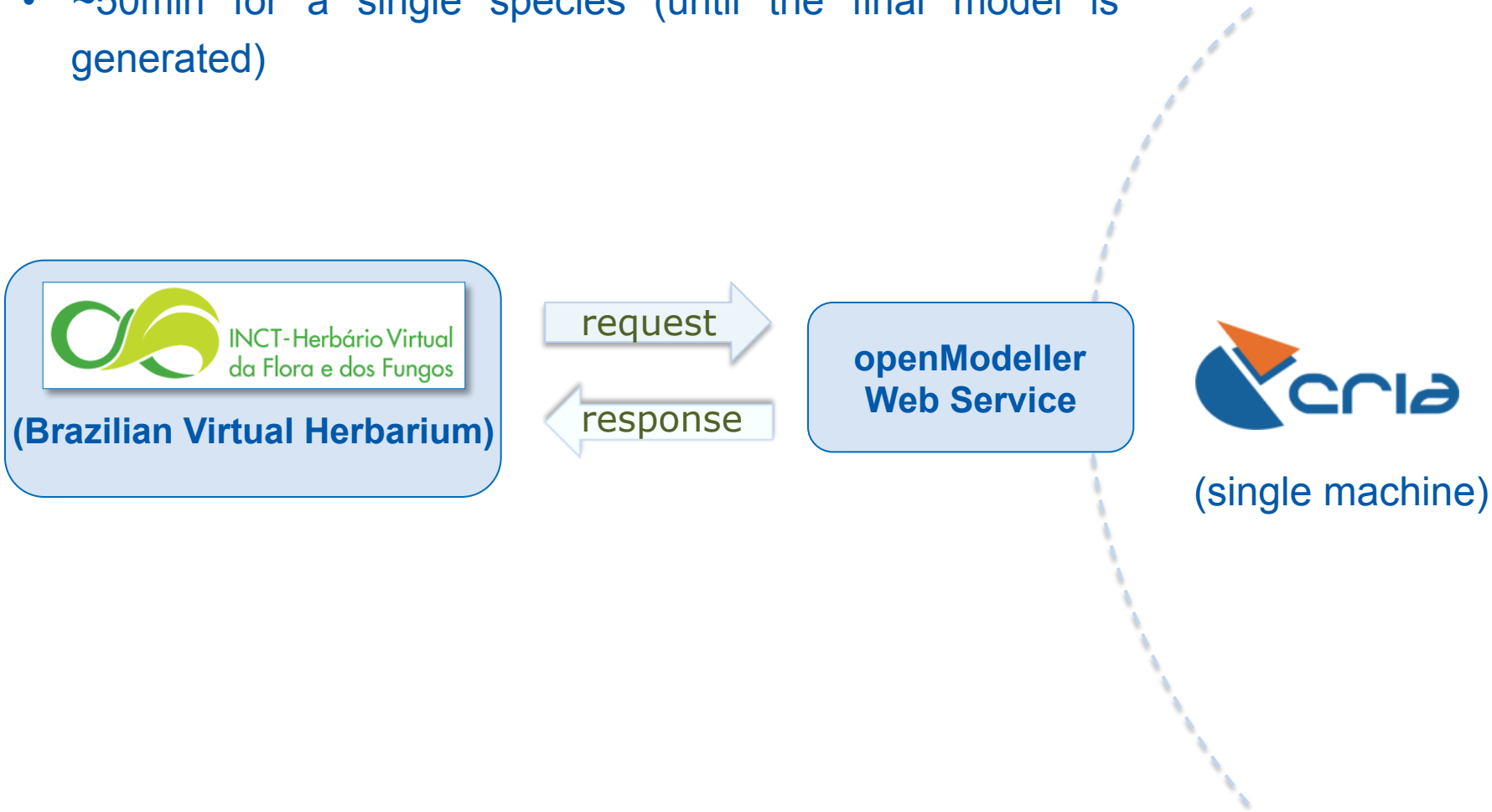
"Set of ecological requirements for a species to survive and maintain viable populations over the time."

(Grinnel, 1917)

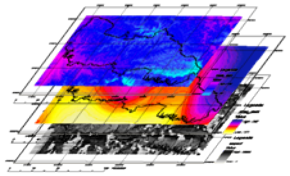


The Use Case: Ecological Niche Modelling

- One WS instance and one oM server
- ~50min for a single species (until the final model is generated)



The Use Case: Ecological Niche Modelling

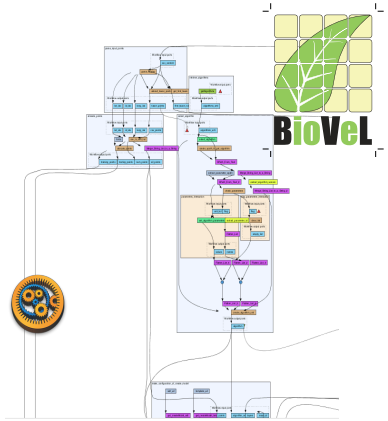


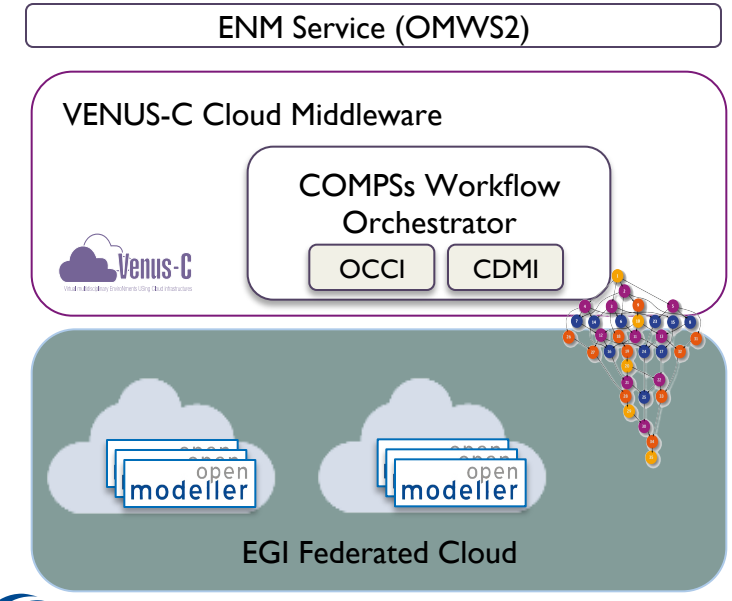
2012 lista de espécies flora do brasil

43203 species (18 Sept. 2012)

- 31 718 angiosperms (flowering plants)
- Assuming that 30% will have enough points to generate models (~9 000 species):
  - 495k models, 540k tests, 90k projections
- 10 months to generate all models!
- But what if we want to generate models for
  - All ~43 thousand plant species from Brazil?
  - Using more than one spatial resolution?
  - Projecting into different environmental climatic scenarios?
  - With global coverage?
- Note: models may be regenerated every time new data is available for each species...

OpenBio in EGI: An interoperable and scalable solution

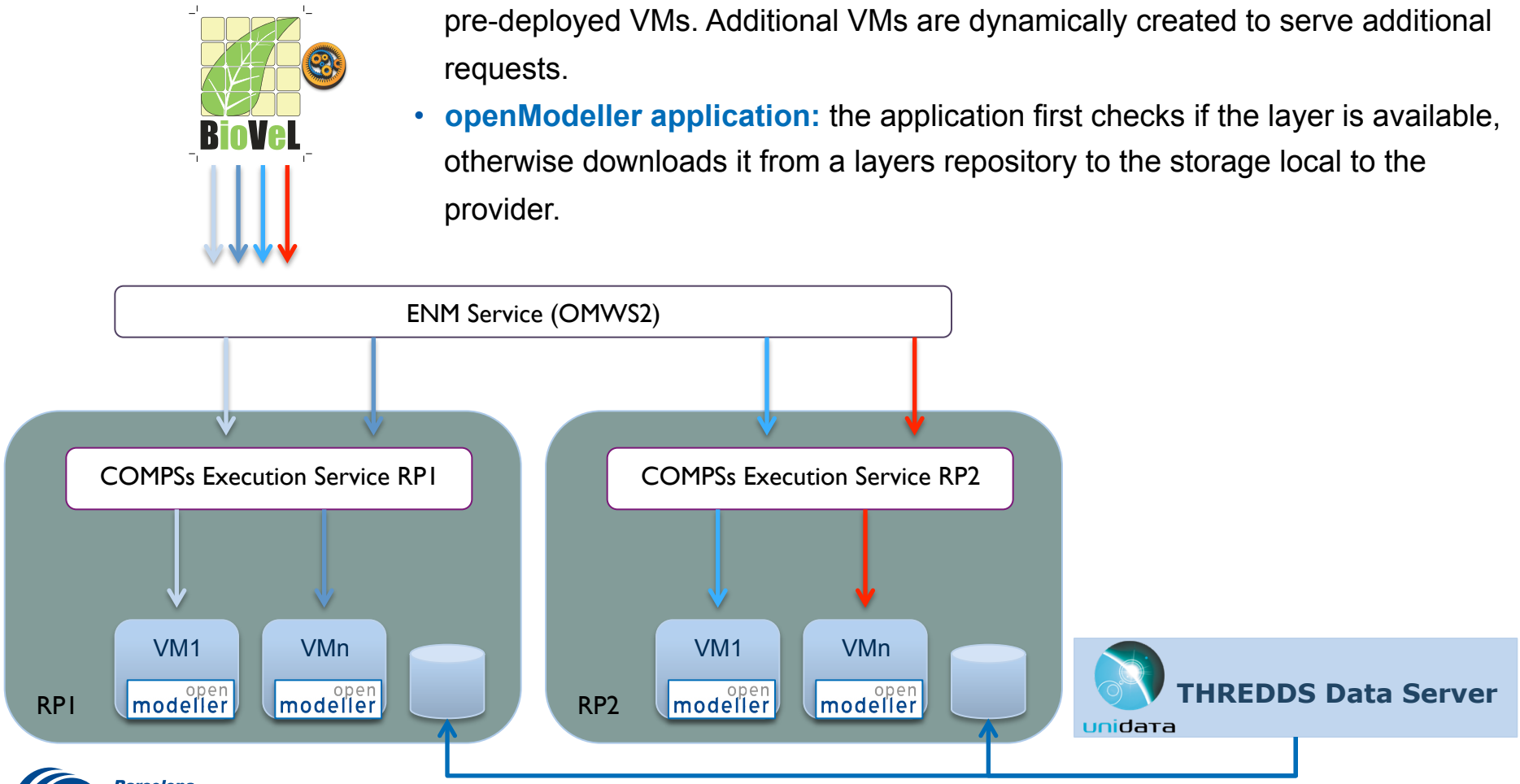




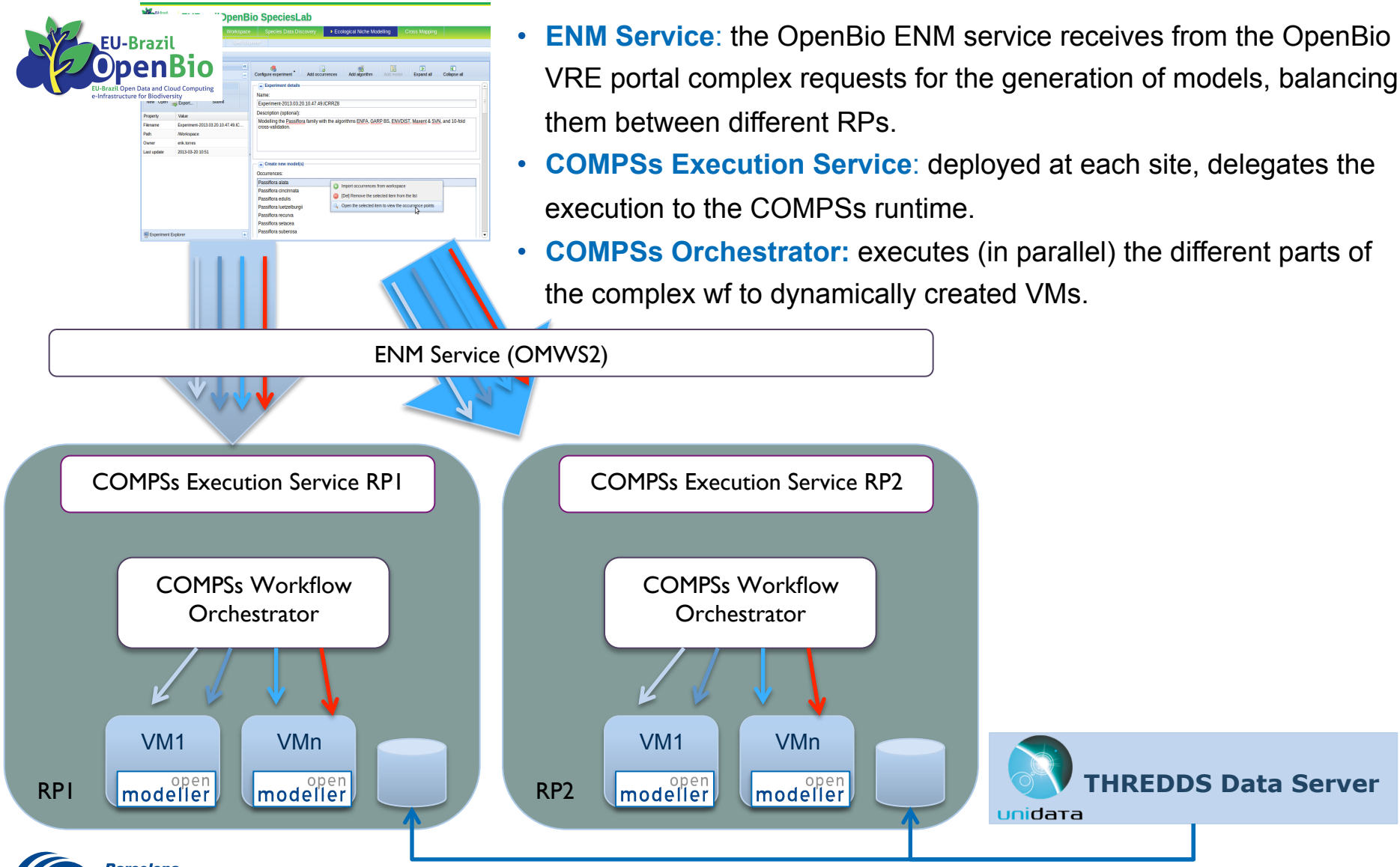
- **Shared requirements** between EUBrazilOpenBio and BioVeL
- The **EUBrazilOpenBio ENM service** is exposed through an extended openModeller Web Service interface (**OMWS2** in the picture).
- Such interface in EUBrazilOpenBio supports multi-staging and multiparametric experiments implemented through **COMPSs** and the openModeller software and managed through a Virtual Research Environment (VRE) portal.
- The OMWS extensions are backwards compatible with the original specification, allowing existing clients, as the Taverna Workflow Management System in BioVeL, to be fully supported in the new implementation.
- In the case of the EGI Federated Cloud, the **VENUS-C** middleware is used to instantiate openModeller workflows on **cloud resources** from different providers **dynamically deployed by COMPSs**.
- An **OCCI connector** is used for the VMs management while data management supports **CDMI** endpoints.

OpenBio in EGI: Execution Scenario 1

- **ENM Service**: the OpenBio ENM service receives from the BioVeL portal simple requests for the generation of models, balancing them between different RPs.
- **COMPSs Execution Service**: deployed at each site, executes the requests to pre-deployed VMs. Additional VMs are dynamically created to serve additional requests.
- **openModeller application:** the application first checks if the layer is available, otherwise downloads it from a layers repository to the storage local to the provider.



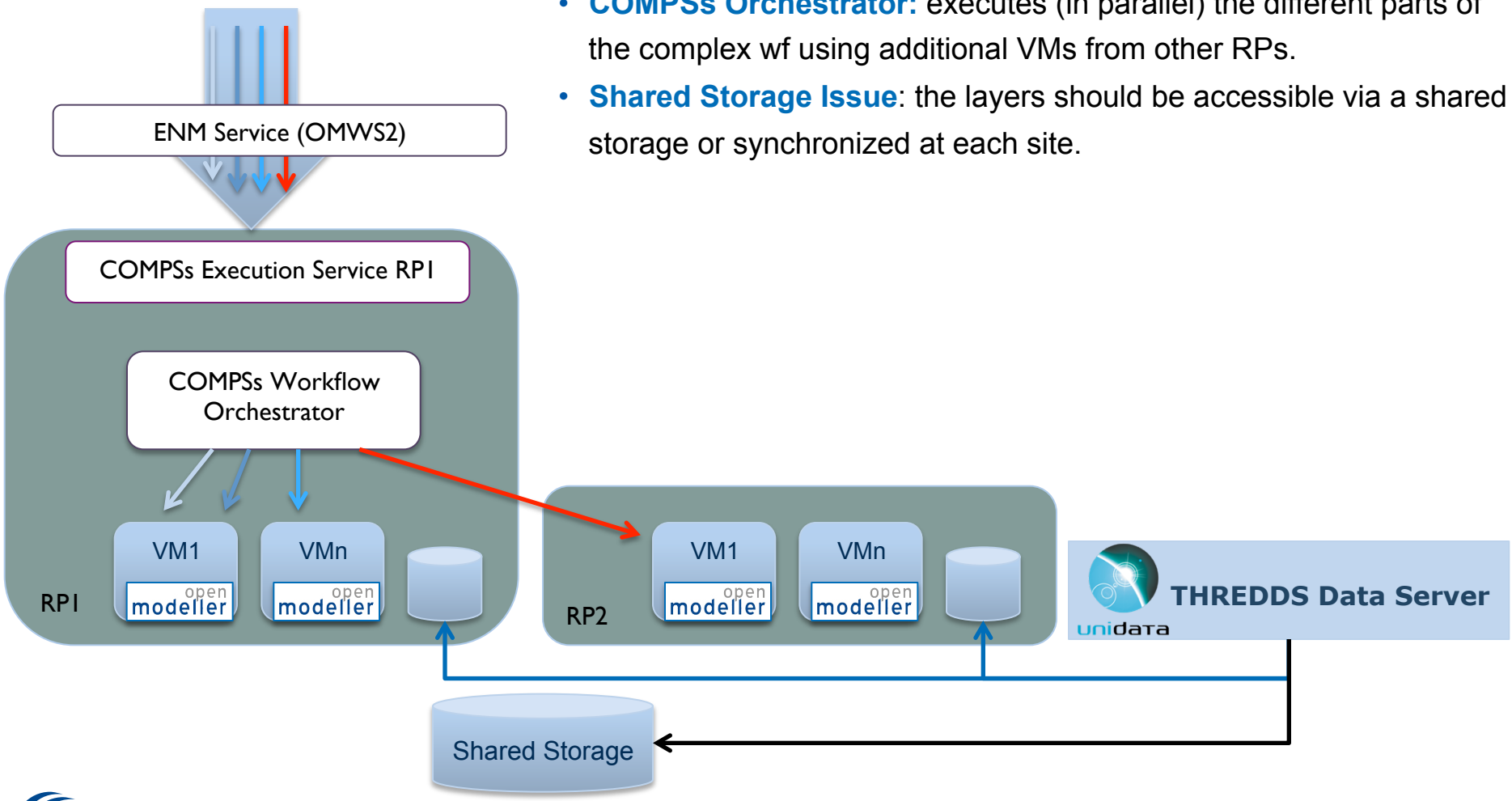
OpenBio in EGI: Execution Scenario 2



- **ENM Service**: the OpenBio ENM service receives from the OpenBio VRE portal complex requests for the generation of models, balancing them between different RPs.
- **COMPSs Execution Service**: deployed at each site, delegates the execution to the COMPSs runtime.
- **COMPSs Orchestrator:** executes (in parallel) the different parts of the complex wf to dynamically created VMs.

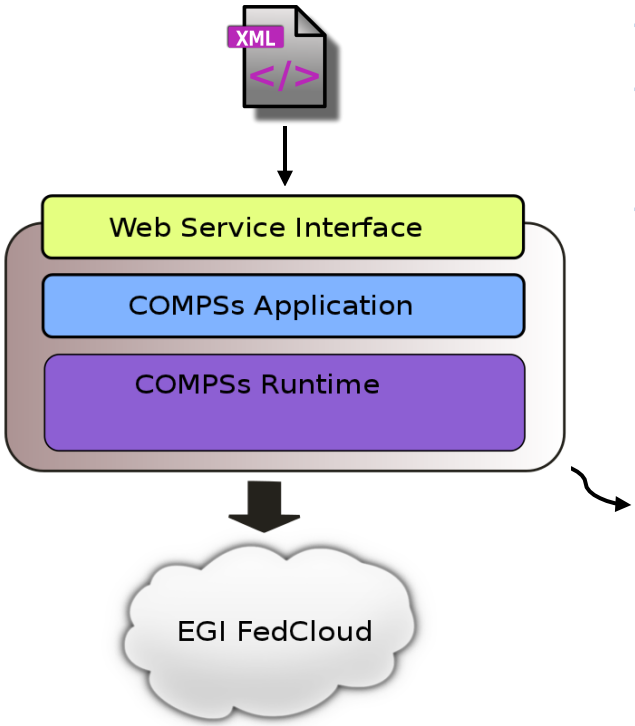
OpenBio in EGI: Hybrid Scenario

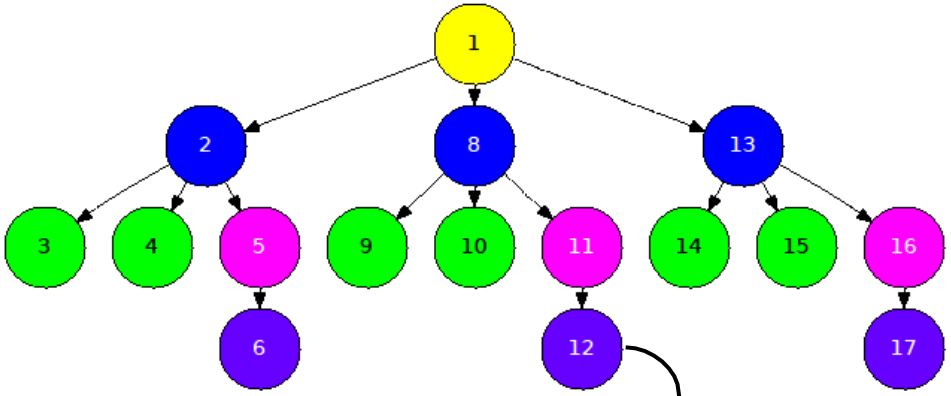
- **COMPSs Execution Service**: deployed at each site, starts the execution of the COMPSs workflow using resources of a RP.
- **COMPSs Orchestrator:** executes (in parallel) the different parts of the complex wf using additional VMs from other RPs.
- **Shared Storage Issue**: the layers should be accessible via a shared storage or synchronized at each site.



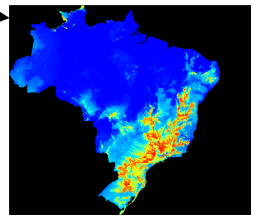
The Modeller Service (operations)

- **Convert:** Converts multi-job request in a single requests.
- **Model:** models a specie distribution for a given request.
- **Test:** Checks the accuracy of the distribution.
- **Project:** Project each distribution over geographical layers in raster format.
- **Translate:** Translates the raster projection into an image.



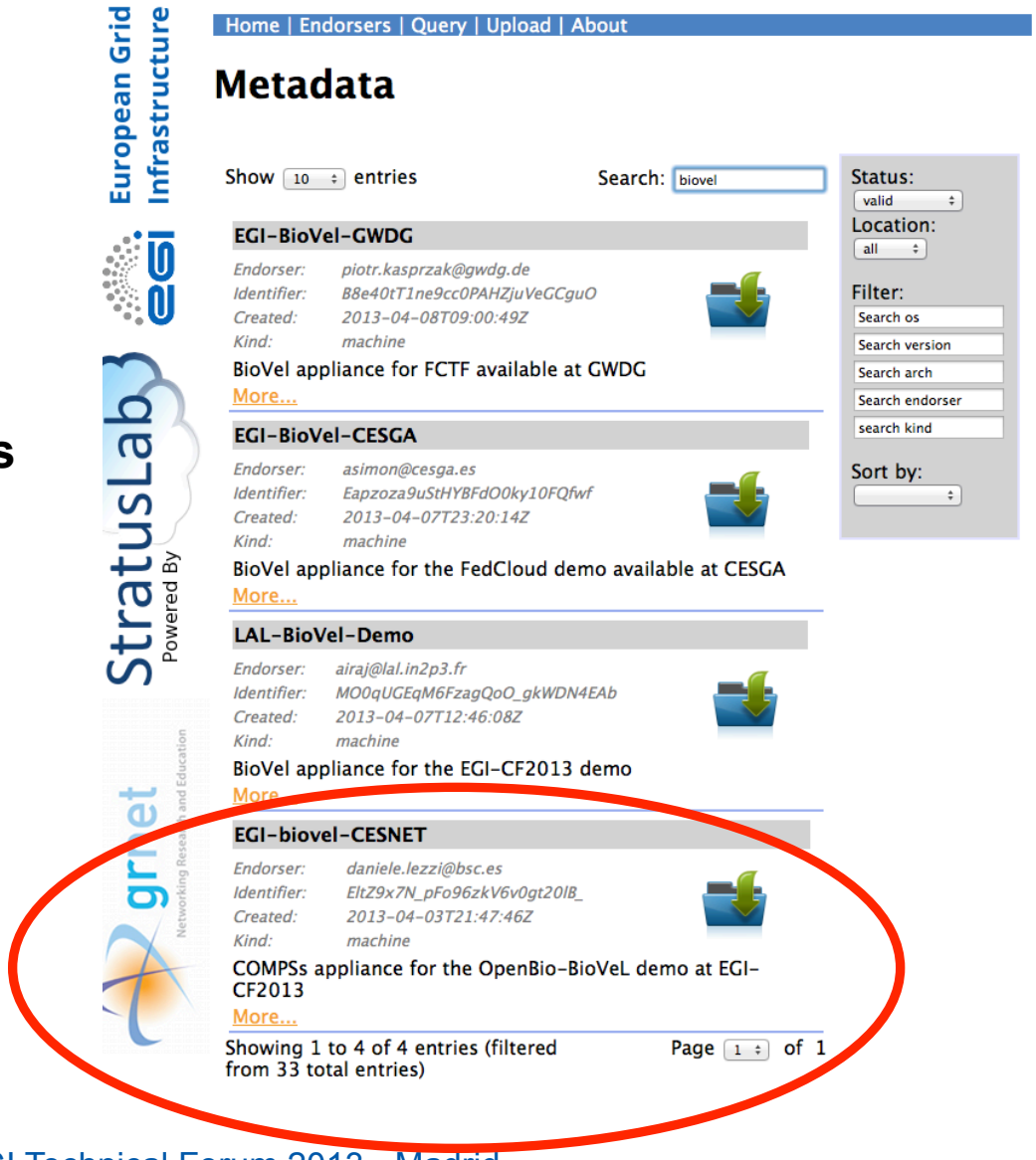


Type	Tasks
Convert	1
Model	2 8 13
Test	3 4 9 10 14 15
Project	5 11 16
Translate	6 12 17



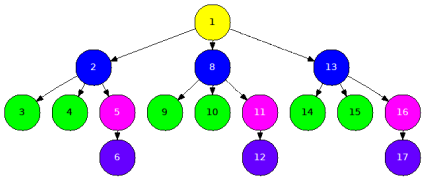
COMPSs integration with EGI FedCloud

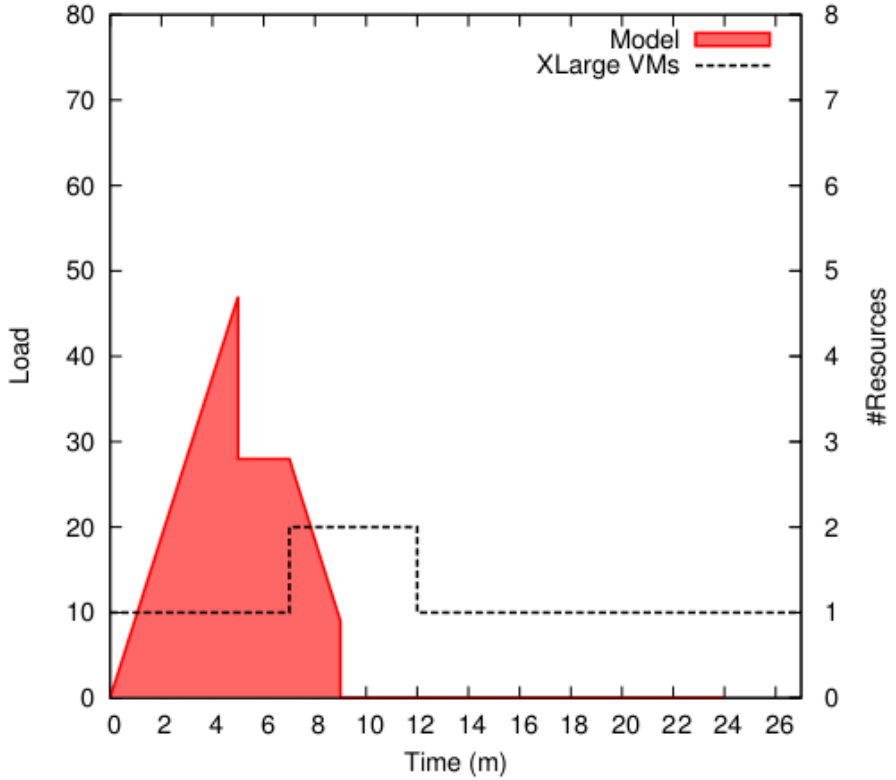
- **A COMPSs VM is available with the required software in the VM repository.**
- **The EGI Marketplace contains the list of providers offering this VM.**

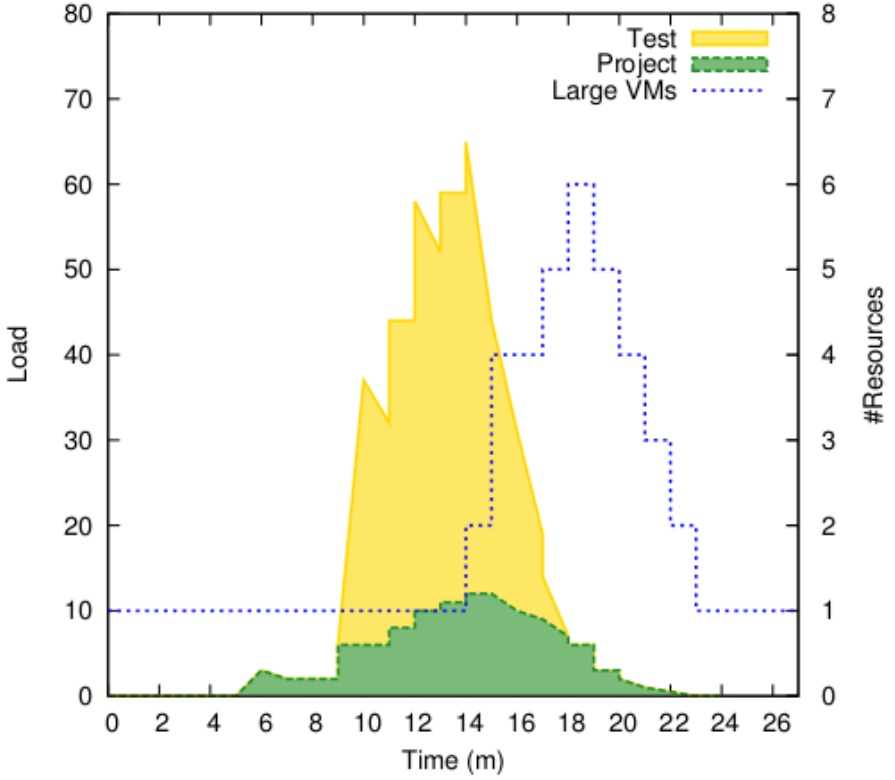


Results: Single request scenario

System load *vs.* available virtual resources







Conclusions

- **Integration of the EUBrazilOpenBio solution in the EGI Federated Cloud:**
  - Provision of the Ecological Niche Modelling Service Endpoint
  - EUBrazilOpenBio extensions to openModeller
  - Seamless execution from the OpenBio VRE and the BioVeL Portal

- **Use of COMPSs for the interoperability with the EGI FedCloud:**
  - Dynamic VM multi-provider management through the rOCCI connector
  - Parallelization of the execution of complex openModeller workflows
  - Optimized execution of BioVeL single requests
  - Support to CDMI endpoints

www.bsc.es

Barcelona Supercomputing Center

Centro Nacional de Supercomputación

Thank you!

For further information

http://www.eubrazilopenbio.eu

daniele.lezzi@bsc.es

EU-Brazil OpenBio

EU-Brazil Open Data and Cloud Computing e-Infrastructure for Biodiversity

BioVeL