

Ecological Niche Modelling in the EGI Cloud Federation

Daniele Lezzi¹, Roger Rafanell¹, Francesc Lordan¹, Renato De Giovanni², Erik Torres³, Ignacio Blanquer³

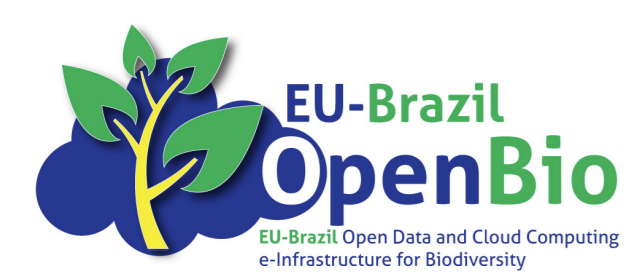
¹ Barcelona Supercomputing Center, Spain

² Centro de Referência em Informação Ambiental, Brasil

³ Universitat Politècnica de València, Spain

Objectives

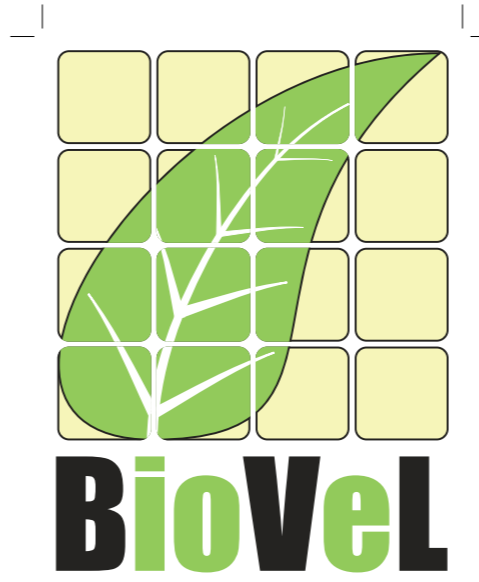
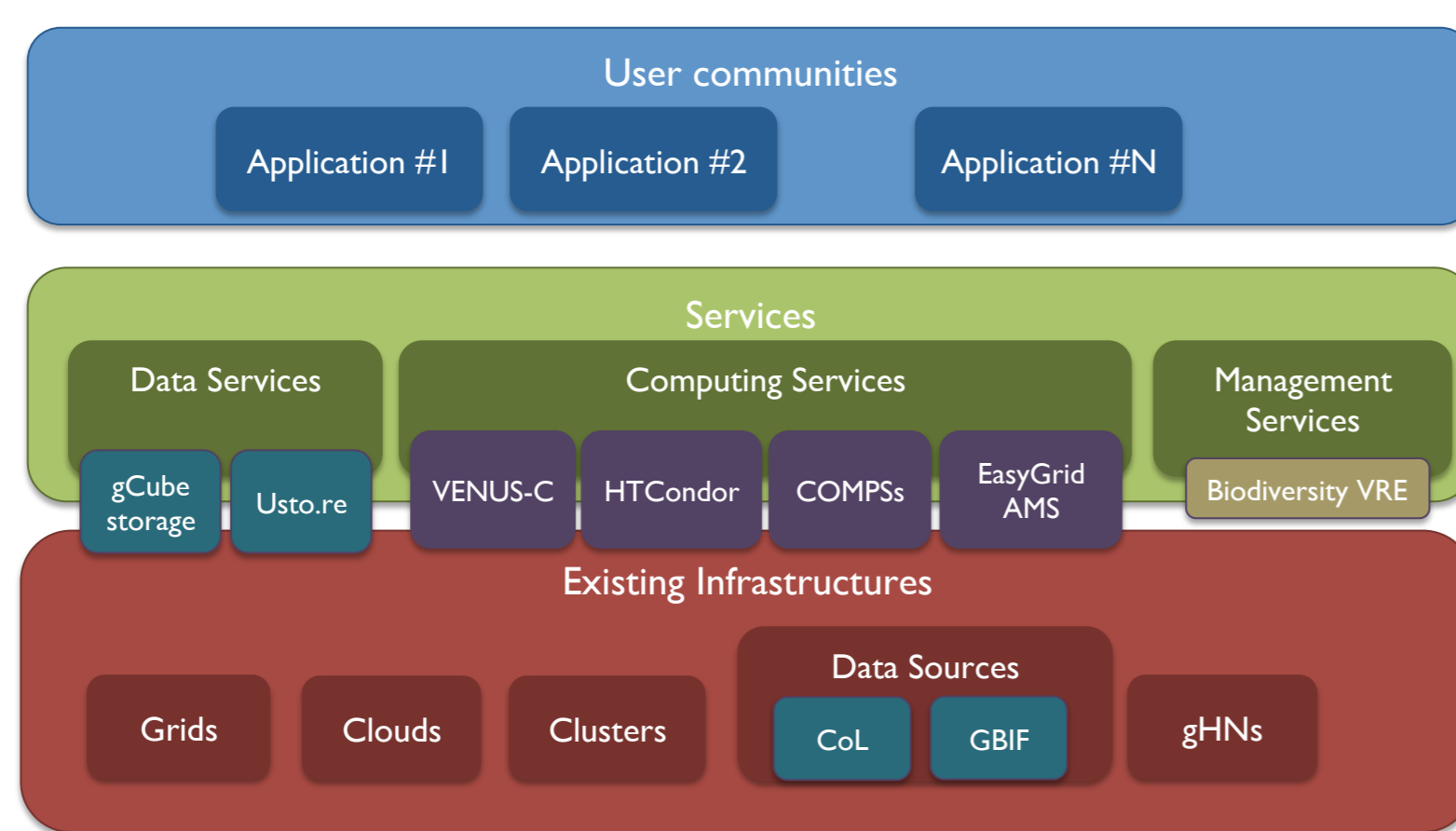
This work describes the collaboration between EU-BrazilOpenBio and BioVeL projects in order to provide biodiversity community with a computational cloud infrastructure able to efficiently process Ecological Niche Modelling (ENM) workflows characterized by intensive computing requirements. The cloud-enabled ENM Service developed in EU-BrazilOpenBio, allows the execution of ENM workflows through the EU-BrazilOpenBio Gateway and through the Workflow Management System available in the BioVeL project. A new application, developed as part of EU-BrazilOpenBio, allows scientists to carry out complex experiments leveraging the openModeller environment and generating models for a large number of species, using complex modelling strategies and involving several algorithms and high-resolution environmental data. The application exploits the interoperability features of the VENUS-C COMPSs middleware aggregating computational resources available to the projects through the EGI Cloud Federation.



An Infrastructure beyond Computing and Storage Resources

EU-BrazilOpenBio is a collaborative initiative addressing strategic barriers in biodiversity research by integrating open access data and user-friendly tools widely available in Brazil and Europe. EU-BrazilOpenBio deploys a joint EU-Brazil cloud-based e-infrastructure that allows the sharing of hardware, software and data on-demand. Biodiversity scientists can use these open access resources and the applications developed by the project to conduct a wide range of conservation and research programmes.

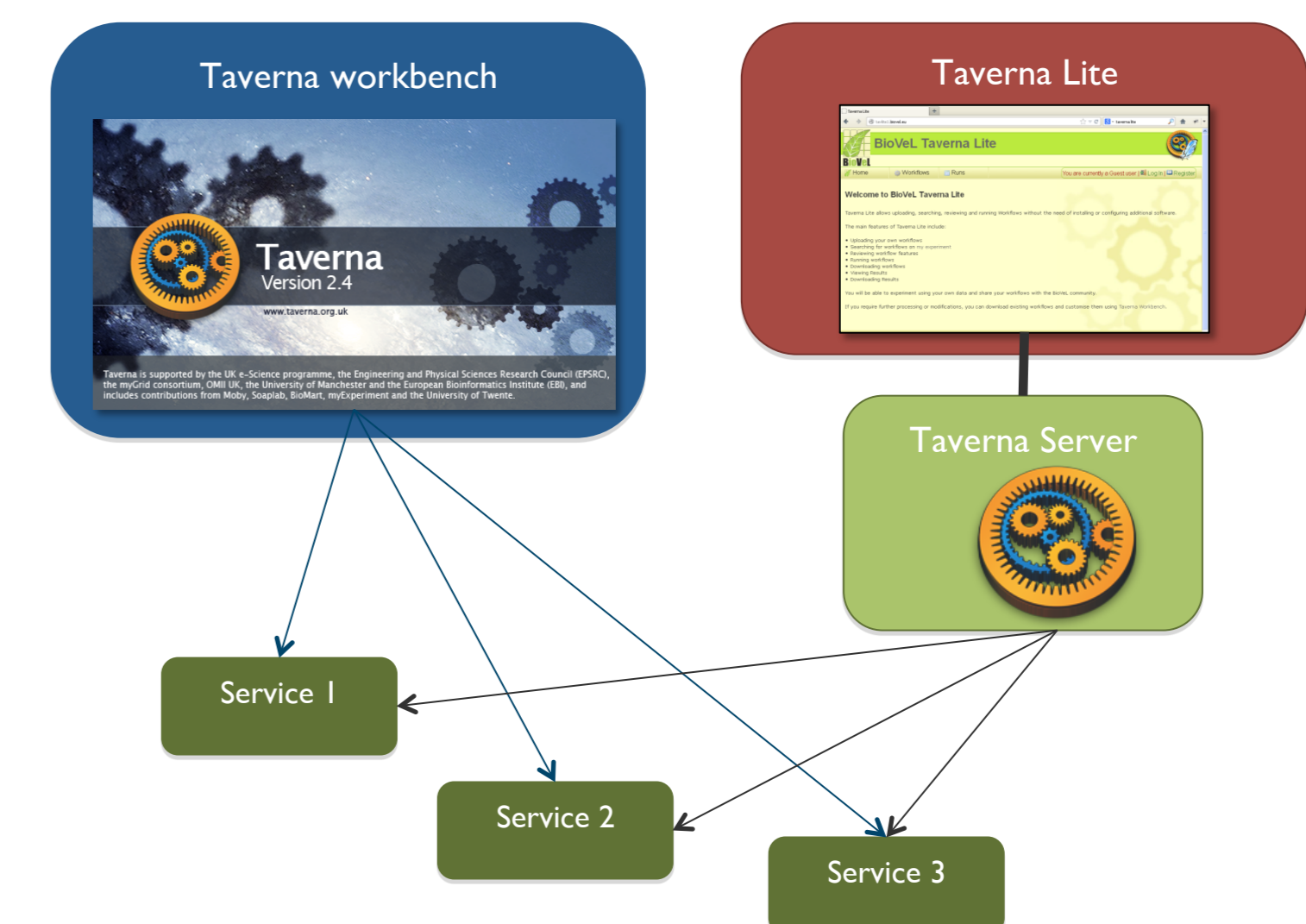
EU-BrazilOpenBio operates a Hybrid Data Infrastructure, i.e. a new type of Data Infrastructure specifically conceived to deal with data-intensive science. Such an infrastructure nicely integrates several technologies, infrastructures and information systems to enable a data-management-capability delivery model in which computing, storage, data and software are made available by the Infrastructure as-a-Service.



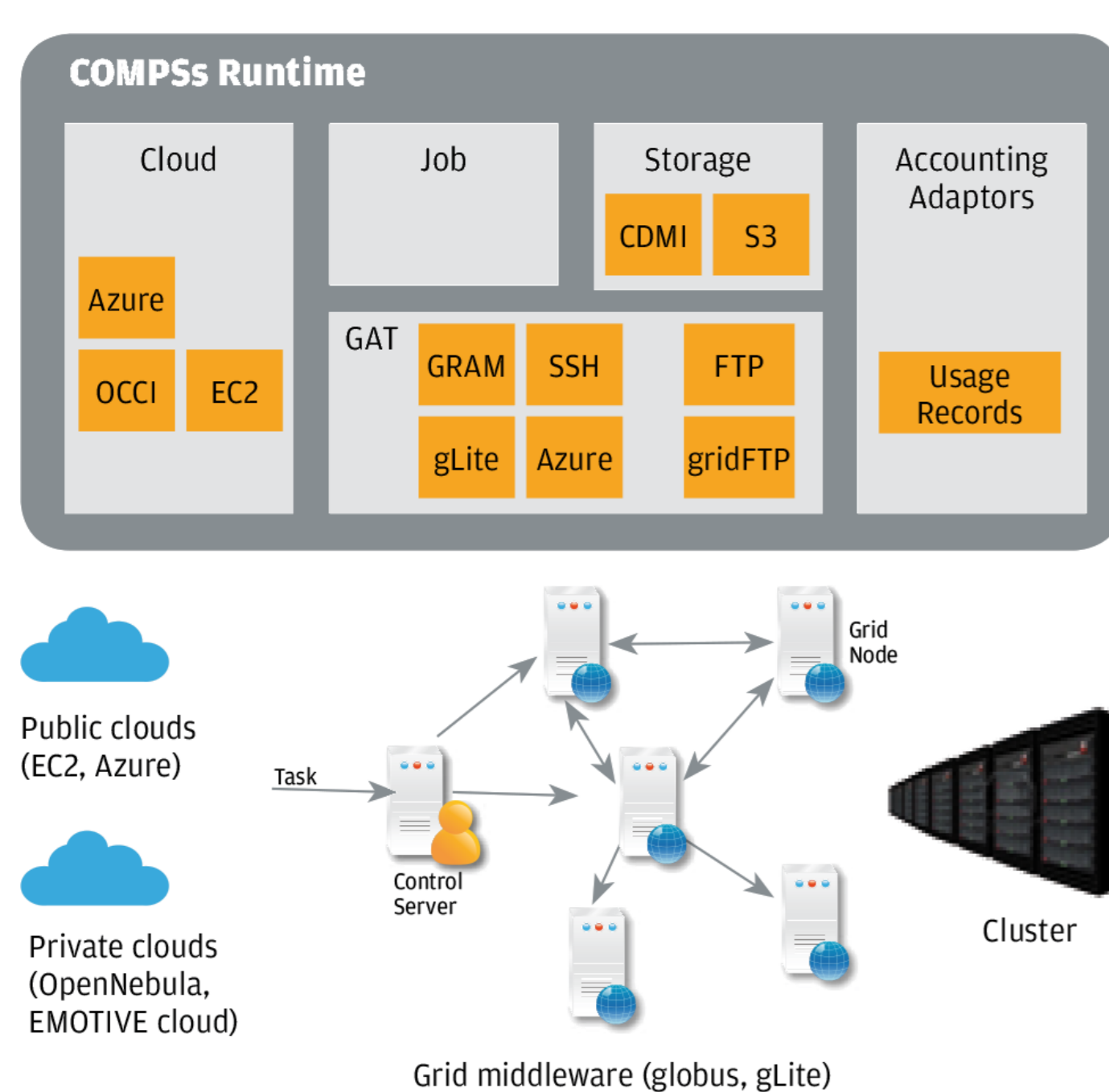
Biodiversity Virtual e-Laboratory

The BioVeL project is creating a set of workflows and Web Services for scientists to perform different analyses in the fields of ecology, taxonomy, phylogenetics and metagenomics. Each workflow is comprised by a number of tasks that can for example communicate with Web Services, interact with the end user or run other specific procedures. Tasks can be organized and connected in order to achieve a desired goal, exploiting data resources and tools and facilitating in-silico experiments. Some of the show cases being carried out as part of the project involve ecological niche modelling to help preventing species invasions and assessing the impact of climate changes.

BioVeL workflows are created and managed with the myGrid suite of tools: Taverna, BioCatalogue and myExperiment. There are different ways in which a workflow can be run, including the Taverna Workbench Desktop environment and the Taverna Lite Web interface.



The COMPSs framework



The COMPSs programming framework, developed by BSC, allows the development of scientific applications and their seamless execution on a wide number of distributed infrastructures. In cloud environments, COMPSs provides scaling and elasticity features allowing to adapt the number of available resources to the actual need of the execution.

This is achieved through the use of connectors for IaaS offerings, as Amazon EC2, OpenNebula and OCCl compliant middlewares. A specific adaptor allows COMPSs to benefit from the Microsoft Azure (PaaS).

An interoperable solution

In order to provide BioVeL with a cloud enabled openModeller endpoint, the EUBrazilOpenBio ENM service is exposed through an extended openModeller Web Service interface (OMWS+ 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.

An Experiment Orchestrator Service acts as dispatcher of user's requests towards different infrastructures. In the case of the EGI Federated Cloud, the VENUS-C middleware is used to instantiate openModeller workflows on cloud resources.

The COMPSs Workflow Orchestrator receives the execution requests and takes care of the execution of the openModeller pipelines on dynamically deployed virtual machines.

An rOCCI connector is used for the VMs management while data management supports CDMI endpoints.

