

# Integrating openModeller with the EGI Federated Cloud

[Boris Parak](#), CESNET, CZ

[Nuno Ferreira](#), TONC EGI.eu

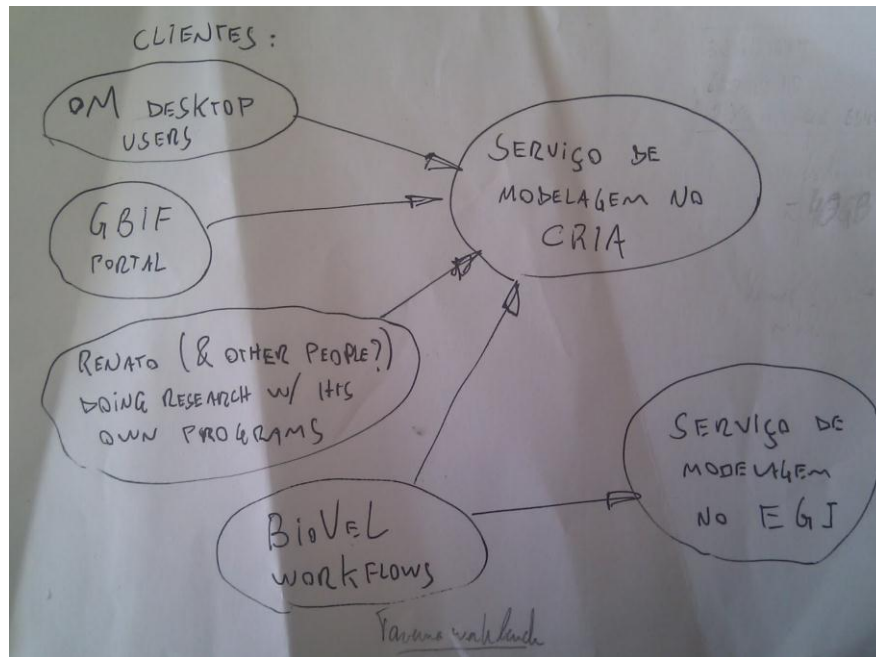
[Renato de Giovanni](#), CRIA, BR

- **FedCloud Wiki page** <http://go.egi.eu/fc.com>
- **Provenance** [BioVeL](#) project (EGI.eu is a partner)
- **Software name** [openModeller](#) (oM)
- **Scientific area** Ecological niche modelling
- **Software goal** Creates a niche model and displays it by projecting it into an environmental scenario.

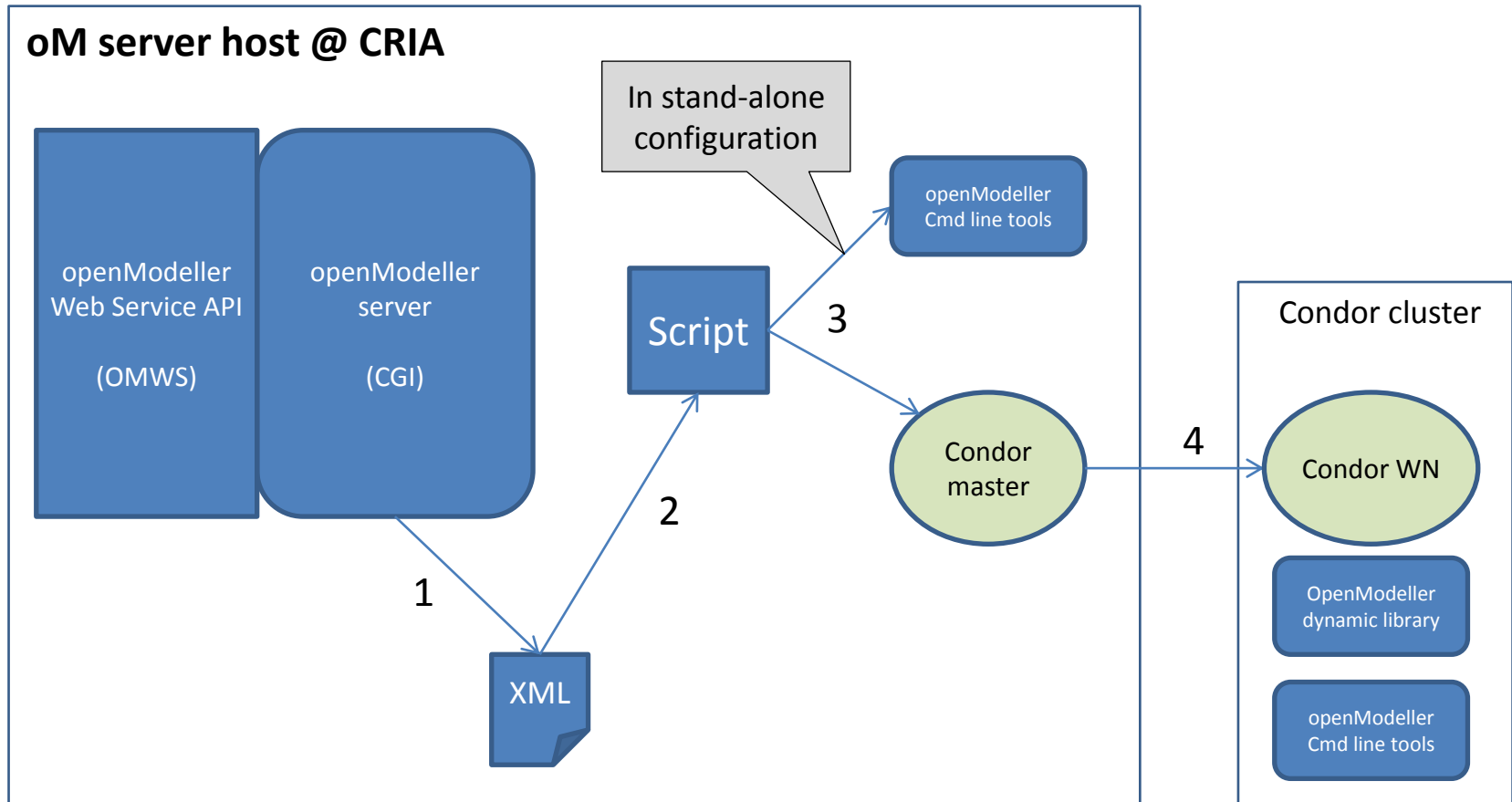


- oM provides a **flexible, user friendly, cross-platform environment** where the entire process of conducting a fundamental **niche modeling experiment** can be carried out.
- The **software includes facilities for reading species occurrence and environmental data**, selection of environmental layers on which the model should be based, creating a fundamental niche model and projecting the model into an environmental scenario.
- **A number of algorithms are provided** as plugins, including GARP, Climate Space Model, Bioclimatic Envelopes, Support Vector Machines and others.
- **Different client interfaces** to interact with oM
- The **project is being developed by** the *Centro de Referência em Informação Ambiental* ([CRIA](#)), *Escola Politécnica da USP* ([Poli](#)), and *Instituto Nacional de Pesquisas Espaciais* ([INPE](#)) as an **open-source initiative**.
- It is **funded by** *Fundação de Amparo à Pesquisa do Estado de São Paulo* ([FAPESP](#)), the [Incofish](#) project, and others.

- CRIA hosts a oM server
- Different clients available
  - CLI
  - Desktop (GUI)
  - Web-based
- Use case goal
  - Virtualize CRIA oM server
  - Operate instance 24/7 @ Europe
  - EGI FedCloud as resource provider (RP)
  - Serve BioVeL users



**The use case aims** to virtualise CRIA openModeller server and then setup and operate an instance of it in Europe using resources from the EGI Federated Cloud and from the EGI Service Availability Monitoring infrastructure. The virtualised instance would serve the Biodiversity Virtual e-Laboratory ([BioVeL](#)) that supports research on biodiversity issues using large amounts of data from cross-disciplinary sources. BioVeL offers the possibility to use 'workflows' (series of data analysis steps) to process data, be that from one's own research and/or from existing sources.



**Further information:** [http://openmodeller.sourceforge.net/om\\_server.html](http://openmodeller.sourceforge.net/om_server.html)

Integration	Disk space	Memory	VCPU	IP
Test phase	30 GB	1 GB	1 or 2	public
Production phase	Circa 100 GB	Under assessment	Under assessment	public

## ➤ Relevant oM facts:

- CRIA's oM server in Brazil running with following specs:

2 Processors Intel® Xeon® Six-Core E5649 2.53GHz, 12M cache, 5.86 GT/s \* 32GB Memory (8x4GB), 1333MHz Dual Ranked LV RDIMMs \* 4 Hard Disks of 600GB 10K RPM SA SCSI 6Gbps 2.5in Hotplug

- Job run time ranges from seconds to hours
- Not possible to know *apriori* what it will be the job run time based on the input parameters

## 1. Server in the cloud

- Start up an oM stand-alone server in the EGI Federated cloud
- Reconfigure the service to use Condor or other workload management system (e.g. [ToPoS](#))

## 2. Static VM pool

- Start a fixed number of job execution hosts in the EGI cloud

## 3. Monitored VM pool

- Monitoring the job execution image instances (and the server instance?), in case of any failure start a new instance

## 4. Dynamic VM pool

- Monitoring the load on the service, adding and removing job execution VM instances according to this load

## 5. Dynamic, monitored VM pool

- Integration of scenarios 3 & 4

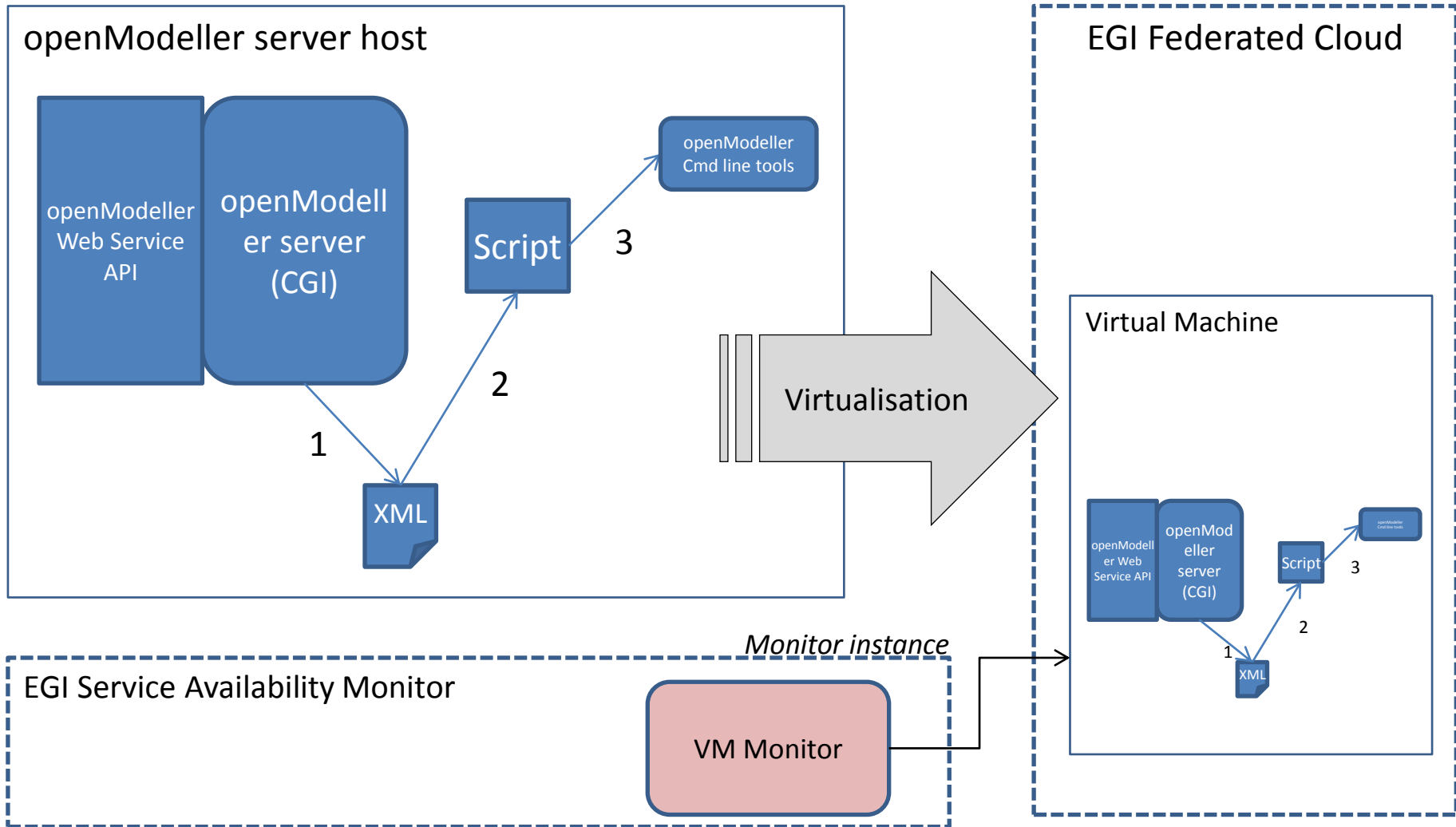
### Source of info:

Skype meeting on the 17<sup>th</sup> July 2012 (EGI.eu TONC and oM SOAP interface developer)

- The first assessment of oM use case requirements (via Skype) produced a **set of 5 scenarios of increasing complexity** to implement the use case in the EGI FedCloud.
- After a **re-evaluation of how the oM server works** in conjunction with the developer Renato (BioVeL meeting 20120904), **scenario 1 described earlier seems to be enough** to accomplish the goals for the time being.
- And! no need to have a workload management system in place if oM VM has VCPUS => 2.



# Scenario 1: server in the cloud



Steps	Remark	Action on	Status
1 . Create a VM image from the openModeller server in stand-alone configuration		Renato (CRIA)	Done
2 . Send oM HW specs to EGI		Renato	On-going
3. Find a provider; start a VM instance in an EGI cloud; test the web-service.	Use the EGI FedCloud client side scripts (requires customisation)	Nuno (TONC)	Done
4. Create a monitor that opens trouble ticket if the server is down	EGI Service Availability Monitor	EGI Operations	Not started
5. Register the instance in the BioVel catalog	Biocatalog (may require modification to handle service duplications)	Renato	Not started

## Source of info:

Skype meeting with EGI.eu TONC and oM SOAP interface developer (20120717)

- oM developers encapsulated a minimal server in a VirtualBox VM
  - Debian Squeeze 6.0.5
  - Kernel 2.6.32-5
- VirtualBox VM exported as .ova
  - .ova contains the .ovf descriptor and the disk images in VMDK format
  - Available @ <ftp://ftp1.cria.org.br/...>
- Final hardware specifications to run the VM still under assessment

## ➤ How we choose a Resource Provider?

1. Check [EGI FedCloud resource status](#) ... target RPs with green status

User registration	User access	VM availability	Elastic IPs	Object Storage	Persistent Storage
-------------------	-------------	-----------------	-------------	----------------	--------------------

2. CESNET and SARA were contacted for preliminary assessment
3. Query the FedCloud RPs about acceptance of use cases
  - ([FedCloud private mailing list](#), 20120814)
  - Only CESNET replied to EGI.eu
4. Engage with the RP (CESNET)

## ➤ Info sent to the RP via email (20120813)

- **Name:** Nuno Ferreira
- **Organization:** EGI.eu
- **email:** [nuno.ferreira@egi.eu](mailto:nuno.ferreira@egi.eu)
- **DN:** /O=dutchgrid/O=users/O=universiteit-utrecht/OU=chem/CN=Nuno Loureiro Ferreira
- **Pub SSH key**
- **Use case heads up**

➤ **CESNET RP contacts**

- [Miroslav Ruda](#)
- [Boris Parak](#)

➤ **Access provided to TONC member to the following end-points (20120814)**

- OCA (OpenNebula's native API)
- Sunstone (GUI)
- OCCl 0.8 (OpenNebula's OCCl draft implementation).

➤ **RP converted VM to be xen aware**

- `qemu-img convert -f vmdk DISK.vmdk -O raw DISK.img`
- Kernel update 2.6.32-5-xen (supports pv-ops)
- [CESNET FedCloud manual](#) updated

➤ **RP created a oM template (ONE) & booted VM (20120926)**

➤ **SUNSTSONE (GUI):**

- <https://carach5.ics.muni.cz/>

➤ **oM client**

- [openModeller Desktop](#)

**Note:** oM GUI was used to test the webservice

- **EGI FedCloud endpoint:** <http://took21.ics.muni.cz/ws1/om>

**Note:** The web service endpoint is the only bit of information BioVeL users (and oM users in general) need to know to make use of this service.

➤ **SSH client**

- usage of ssh pub key for VM remote access
- oM VM timezone changed
- Web service configuration file edited:

```
[root@took21 ~]# cat /var/www/vhosts/modeller/ws1/config/server.conf | grep BASE_URL
#BASE_URL=http://200.144.120.70/maps/
BASE_URL=http://took21.ics.muni.cz/maps/
```

```
[root@took21 ~]# hostname
took21.ics.muni.cz
[root@took21 ~]# uname -a
Linux took21.ics.muni.cz 2.6.32-5-xen-686 #1 SMP Sun May 6 07:43:49 UTC 2012 i686 GNU/Linux
[root@took21 ~]# cat /etc/debian_version
6.0.5
[root@took21 ~]# free -m
              total          used          free   shared    buffers     cached
Mem:           1019           429           590         0         208         154
-/+ buffers/cache:           66           953
Swap:          1905             0          1905
[root@took21 ~]# cat /proc/cpuinfo | grep "model name"
model name      : Intel(R) Xeon(R) CPU           E5649  @ 2.53GHz
model name      : Intel(R) Xeon(R) CPU           E5649  @ 2.53GHz
[root@took21 ~]# df -h
Filesystem      Size  Used Avail Use% Mounted on
/dev/xvda1      938M  226M  666M  26% /
tmpfs           510M   0  510M   0% /lib/init/rw
udev            494M   84K  494M   1% /dev
tmpfs           510M   0  510M   0% /dev/shm
/dev/xvda9      54G   181M   51G   1% /home
/dev/xvda6      1.9G   35M   1.8G   2% /tmp
/dev/xvda8      15G  951M   13G   7% /usr
/dev/xvda7      180G  544M  171G   1% /var
[root@took21 ~]# ls -lrt /var/www/vhosts/modeller/ws1/tickets/ | tail -5
-rw-r--r-- 1 modeller modeller 11221 Oct 11 15:03 test_proc.11z434
-rw-r--r-- 1 modeller modeller   717 Oct 11 15:03 test_resp.11z434
-rw-r--r-- 1 modeller modeller     3 Oct 11 15:03 prog.11z434
-rw----- 1 modeller modeller   271 Oct 11 15:03 11z434
-rw-r--r-- 1 modeller modeller     0 Oct 11 15:03 done.11z434
[root@took21 ~]# uptime
12:10:34 up 21 days, 19:37,  1 user,  load average: 0.00, 0.00, 0.00
```

Ticket jobs



Action	Responsibility
<b>1. Define the HW specs of oM server for production mode</b> <ul style="list-style-type: none"> <li>• A few VCPUs ( <math>\geq 2</math> )</li> <li>• Memory, disk space</li> </ul> <p>Note: "CESNET can provide up to 22 VCPUs and 90 GB of RAM for a single" (Boris)</p>	Renato
<b>2. Prepare oM test suite (aims to test the web service)</b>	Renato
<b>3. Define final structure of oM VM data partitions</b> <ul style="list-style-type: none"> <li>• Persistent disk for static data (environment layers)</li> <li>• Scratch/non-persistent data (job outputs, ...)</li> </ul>	Boris (after input from Nuno/Renato)
<b>4. Turn on data sync in the VM (pulls all environmental layers from CRIA)</b>	Nuno
<b>5. How to re-direct BioVeL Taverna based workflows to the new end-point?</b>	Renato
<b>6. Register oM in EGI AppDB and BioVeL catalog</b>	Nuno & Renato
<b>7. How to implement High Availability of the oM VM?</b>	?
<b>8. How to monitor the oM VM?</b>	?
<b>9. Integration of OCCI/CDMI standards</b>	?
<b>10. How will this web service be upgraded? Sustainability plans? Marketplace?</b>	?