

Cloud GARR: L'esperienza d'uso in D4Science

Andrea Dell'Amico
CNR-ISTI



**NET
MAKERS**

D4Science.org

VREs (Virtual Research Environments) support research projects and open science

- Managed services (JupyterHub, Rstudio, Shinyproxy, CKAN-based catalogues, Geoserver, Geonetwork, Hadoop, etc.)
- (mostly) unified storage
- Authentication and authorization
- HA and redundancy wherever is possible

D4Science gateway for OpenAIRE: <https://tools.openaire.eu>

D4SCIENCE LABS

AnalyticsLab



BIOnym



BiodiversityLab



RPrototypingLab



RStudioLab



SUPPORTED PROJECTS AND INITIATIVES

D4STeam



EAGLE



dnet



gCube



Resource Catalogue

Welcome to the D4Science Catalogue!

Here you will find data and other resources hosted by the D4Science.org infrastructure.

The catalogue contains a wealth of resources resulting from several activities, projects and communities including BlueBRIDGE (www.bluebridge-vres.eu), i-Marine (www.i-marine.eu), SoBigData.eu (www.sobigdata.eu), and FAO (www.fao.org).

All the products are accompanied with rich descriptions capturing general attributes, e.g. title and creator(s), as well as usage policies and licences.

Items Search

Insert keywords here

[See All Items](#) [See All Tags](#)

D4Science Catalogue statistics

| | | | |
|-------|---------------|--------|-------|
| 81.3k | 34 | 34 | 30 |
| Items | organisations | groups | types |

Browse by Organisations

| | | | | |
|-----------------------------|---------------------------|-----------------------------------|------------------------|----------------------------|
| iMarine (46753) | D4Science Labs (23332) | French Tropical Tuna Atlas (1556) | FAO (1431) | ORION Knowledge Hub (1101) |
| AGINFRA RAKIP_portal (1000) | AGINFRA RAKIP_trial (886) | IOTC_SS3 (958) | Biodiversity Lab (943) | Global Tuna Atlas (926) |

[See All Organisations](#)

Browse by Groups

| | | | | |
|---------------------------------|---------------------|---|--|------------------------------------|
| biota (48583) | environment (24366) | boundaries (1839) | climatology/Meteorology/Atmosphere (978) | oceans (598) |
| SoBigData City Of Citizens (49) | economy (44) | SoBigData Explainable Machine Learning (40) | SoBigData Societal Debates (39) | SoBigData Ethics and Legality (38) |

OpenAIRE Tools Lab Administration OpenRefine R Studio Production Dev Tools Dev Tools @CNR Social Feed

Notes for the exploitation of this Virtual Research Environment

Development and integration environment for R, Python, and other supported software languages

START

- It is powered by a cluster of RStudio servers, each with 16 cores and 32 GB RAM.
- It is powered by JupyterHub with a maximum of 4 cores and 8 GB RAM per notebook. JupyterHub is provided by [D4Science](https://d4science.org) with the support of EGI.eu

JupyterHub

JupyterHub enables the exploitation of computational environments and resources without burdening users with installation and maintenance tasks.

This JupyterHub environment is (i) preconfigured with libraries and packages to ease the execution of common data analytics tasks, and (ii) provides access to the Workspace enabling sharing of resources with other members much easier.

RStudio

RStudio provides an integrated development environment for R. It includes a console and a syntax-highlighting editor and it enables code execution. Tools for plotting are also included.

This RStudio environment is (i) preconfigured with libraries and packages to ease the execution of common data analytics tasks, and (ii) provides seamless access to the Workspace enabling sharing of resources with other members much easier.

OpenAIRE Lab [Recent](#) [New](#)

| Name | Owner | Last modified |
|----------------------|-------|---------------|
| This folder is empty | | |

Show 5 entries [Previous](#) [Next](#)

Showing 0 to 0 of 0 entries

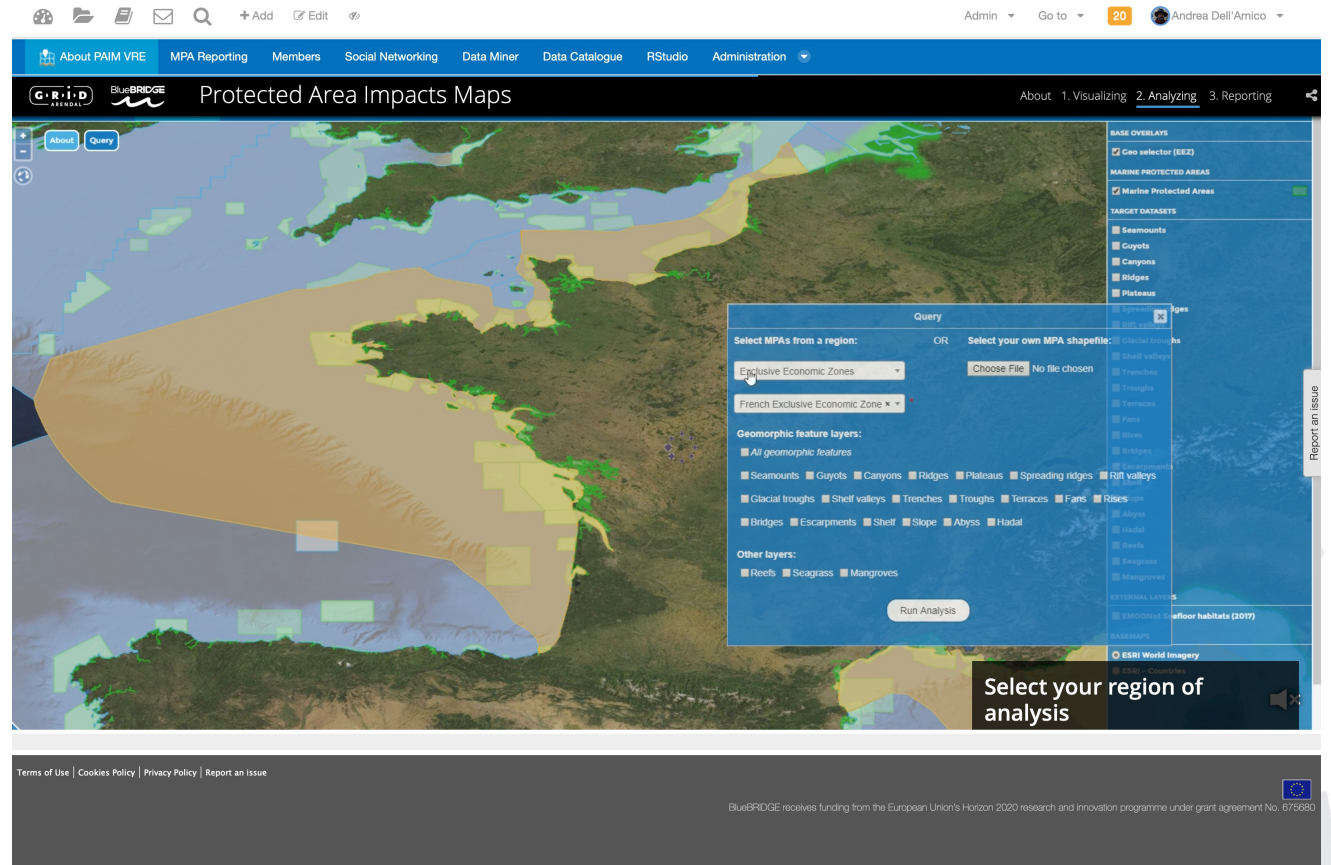
[Go to shared workspace](#)

Use case: Protected Area Impact Maps

<https://i-marine.d4science.org/web/protectedareaimpactmaps>

Services and resources:

- (Authentication)
- (Authorization)
- Static web Application
- Analytics Engine (Dataminer)
- Geoserver
- Workspace
- FAIR



Deployment: everything is automated (mostly)

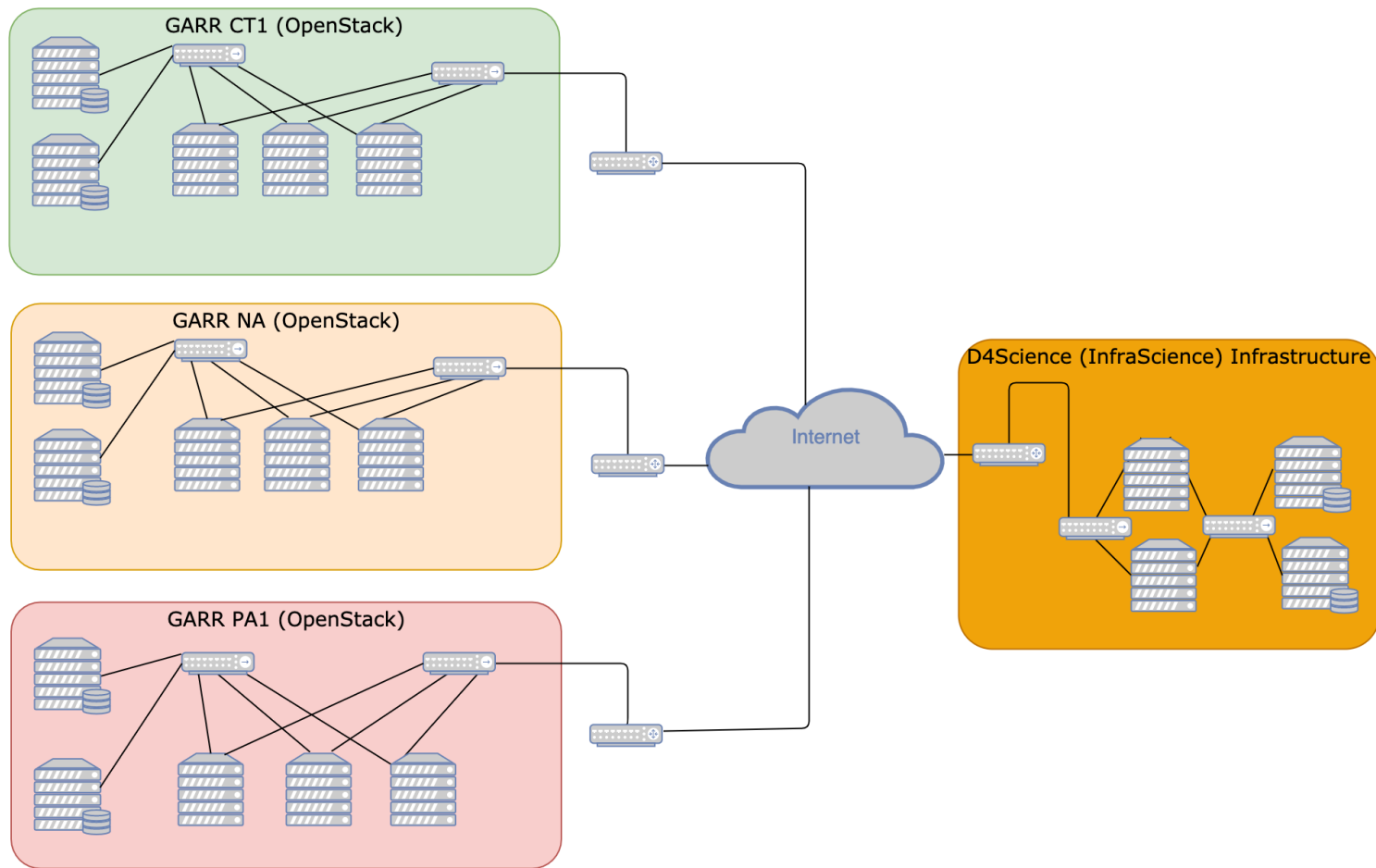
Ansible is the provisioning tool used to configure all the servers and all the services.

- All the servers share a basic configuration set (language, ssh access, timezone, firewall, dns resolver, NTP, ...)
- No manual editing of configuration files
- A new instance of a server can be ready in a matter of minutes

| | |
|--|---------|
| ansible-role-node-js | ★ 0 P 0 |
| Installs node-js from nodesource.com. And, optionally, yarn. | |
| Updated 2 months ago | |
| ansible-role-elasticsearch | ★ 0 P 0 |
| Ansible role that installs the free components of elasticsearch and, optionally, kibana. | |
| Updated 2 months ago | |
| ansible-role-letsencrypt-acme-sh-client | ★ 0 P 0 |
| Ansible role that manages x509 certificates assigned by letsencrypt.org, using the acme.sh client. | |
| Updated 3 months ago | |
| ansible-role-thredds | ★ 0 P 0 |
| Role that installs the THREDDS Data Server. | |
| Updated 3 months ago | |
| ansible-role-python3-environment | ★ 0 P 0 |
| Role that installs python3 and eventually some distribution or pip packages. | |
| Updated 3 months ago | |
| ansible-role-openjdk | ★ 0 P 0 |
| Role that installs openjdk. The Zulu distribution is optionally used (default on Trusty that does not have a valid openjdk repository) | |
| Updated 3 months ago | |
| ansible-role-prometheus-node-exporter | ★ 0 P 0 |
| Installs the prometheus node exporter. | |
| Updated 3 months ago | |
| ansible-role-prometheus | ★ 0 P 0 |
| Role that installs the prometheus server | |
| Updated 4 months ago | |
| ansible-role-dovecot | ★ 0 P 0 |
| Role that installs the dovecot IMAP server. | |
| Updated 4 months ago | |
| ansible-role-java-keystore | ★ 0 P 0 |
| Manages a java keystore | |
| Updated 4 months ago | |
| ansible-role-simplesaml | ★ 0 P 0 |
| This ansible role installs simplesaml. | |
| Updated 4 months ago | |

```
antivirus-servers.yml
authoritative_dns.yml
ca.yml
dhcp-server.yml
freeipa.yml
git-server.yml
haproxy-frontent.yml
imap-director.yml
imap-pop-server.yml
imap-sync.yml
kvm_hosts.yml
mailman.yml
mediawiki.yml
nextcloud-fileserver.yml
postgresql-server.yml
powerdns_admin.yml
radius.yml
redmine.yml
resolvers.yml
roundcube-webmail.yml
san_plan_b.yml
simplesaml.yml
smtp-servers-in.yml
smtp-servers-out.yml
squid.yml
syslog-collector.yml
vm_templates_setup.yml
vpn-service.yml
zabbix-monitoring.yml
```

D4Science – GARR: schema



GARR – D4Science: how it started

- One Region (GARR-CT1) to *test the service*
- Some dev service
- Provisioning of the Vms
- Instances activated manually from the dashboard

GARR – D4Science: now

- Three regions
- Production services
 - Rstudio
 - Analytics Engine
 - Hadoop Cluster
 - Elasticsearch Clusters
 - JupyterHub
 - Kubernetes
 - ...
- Instances managed by ansible exploiting the OpenStack API

GARR – D4Science: how we use the infras

- @GARR: 135 VM, 20+ TB of disk, 3.3 TB of RAM
- @CNR: 463 VM, 800 TB of disk, 6+ TB of RAM
- In the GARR regions:
 - Bigger instances
 - Instances have a shorter lifespan
 - Cluster of services to ensure reliability (analytics engine, JupyterHub, Rstudio)
- We still have SPoF (some services cannot have remote replicas, DNS shortcomings, etc.)

GARR – D4Science: how it looks

cloudusers • isti • garr-ct1 ▾

Project / Compute / Overview

Overview

Limit Summary

Compute



Instances
Used 58 of 325



VCPUs
Used 626 of 650



RAM
Used 1.2TB of 1.2TB

Volume



Volumes
Used 67 of 650



Volume Snapshots
Used 5 of 100



Volume Storage
Used 8.6TB of 10TB

Network



Floating IPs
Allocated 43 of 50



Security Groups
Used 19 of 100



Security Group Rules
Used 88 (No Limit)

Network



Floating IPs
Allocated 21 of 50



Security Groups
Used 15 of 100



Security Group Rules
Used 50 (No Limit)

Usage Summary

cloudusers • isti • garr-na ▾

Project / Compute / Overview

Overview

Limit Summary

Compute



Instances
Used 27 of 325



VCPUs
Used 448 of 650



RAM
Used 1.2TB of 1.2TB

Volume



Volumes
Used 71 of 650



Volume Snapshots
Used 0 of 100



Volume Storage
Used 7.5TB of 10TB

Network



Floating IPs
Allocated 27 of 0



Security Groups
Used 14 of 100



Security Group Rules
Used 67 (No Limit)

Limit Summary

Compute



Instances
Used 50 of 325



VCPUs
Used 627 of 650



RAM
Used 1.2TB of 1.2TB



Volumes
Used 71 of 650



Volume Snapshots
Used 0 of 100



Volume Storage
Used 7.5TB of 10TB



Floating IPs
Allocated 27 of 0



Security Groups
Used 14 of 100



Security Group Rules
Used 67 (No Limit)

GARR cloud: what gives us

- Quite a lot of resources
- Reliability
- The opportunity of offloading some workload when we need to
- APIs

GARR cloud: what we would like to have

- Better storage options: true S3, remote Posix FS (*Data Lakes* is the buzzword of the second half of 2021)
- Availability Zones
- VPN as a service
- Load balancer as a service

GARR cloud future: federation?

- Give to the federation a subset of the internal resources?
- OpenStack constraints?
 - Software version
 - Installation method?
- What (legal) terms?
- Other?

Federated cloud: opportunities!

- Build competences: cloud, distributed storage, monitoring and observability of complex infrastructures.
- Develop solutions tailored to the research community
- Avoid the vendor lock-ins (every commercial cloud provider has its own services, with their APIs and functionalities. Moving out of them is challenging and expensive)