Cloud Computing Lightning Talks

What's new in Cloud?

Timetable

5” Welcome

30” Six lightning talks (5” each)

5” Voting

30” Three extended talks (10” each)

20” Discussion & Wrap-up

Technology talks

Elasticluster: Automate Cluster deployment on Clouds

Sergio Maffioletti, University of Zurich
- Project Director GC3
  Grid **C**omputing **C**ompetence **C**enter
- Project Leader of "Swiss Academic Compute Cloud"
- GC3 partners with SwiNG:
  Swiss Academic computational science support platform

COMPSs: Superscalar programming framework providing elasticity & scalability

Daniele Lezzi, CS, Barcelona SC
- PhD in Computer Engineering
- Since 2008 researcher in Computer Science at BSC
- Specialised in programming frameworks for scientific applications
- Focus on Distributed Computing & Interoperability

Federating Clouds: OpenStack based Clouds sharing user identity using VOMS, LDAP, CAS & SAML

Enol Fernandez, IFCA
- PhD on Grid Scheduling
- From 2003 on contributed to many Grid projects
  CrossGrid, int.eu.grid, Euforia, EGEE, EGI-InsPIRE & EMI
- Since 2010 research focus on Clouds

OpenNaaS: Lightweight and open framework to build NaaS applications

Joan A. García-Espín, I2CAT
- M.Sc. in telematics engineering
- I2CAT Director of Distributed Applications and Networks Area
- Involved in EU FP7 projects on network virtualisation & infrastructure mgmt.
- Task Leader JRA2-T2 (NaaS) in GEANT3+

Provider/Consumer talks

KTH Cloud: KTH's Cloud service portfolio embedded in EGI's Federated Cloud infrastructure

Zeeshan Ali Shah, PDC, KTH
- MS in information securtity
- PhD research in Cyber Security at KTH
- Focus on security incidence in large-scale & distributed systems
- System administrator at PDC

EUBrazilOpenBio: Ecological Niche Modelling (ENM) using workflows in the Cloud

Daniele Lezzi, CS, Barcelona SC
- PhD in Computer Engineering
- Since 2008 researcher in Computer Science at BSC
- Specialised in programming frameworks for scientific applications
- Focus on Distributed Computing & Interoperability

Personal Genome Project: Seeking EGI Cloud expertise for worldwide project support

Peter Walgemoed, Carelliance
- Product & Int'l. Business Development Manager at Phillips Medical Systems
- Founder of Carelliance in 2000
- Focus on long term digital data management
- Co-founder of *Dutch Health Hub*

Ready?

Elasticluster

Automated provisioning of computational clusters in the cloud

Sergio Maffioletti <sergio.maffioletti@gc3.uzh.ch>
GC3: Grid Computing Competence Center
University of Zurich.

Do you need to deploy...

a SGE cluster

. . . to cloud-enable your existing workload.

a Matblab cluster

. . . to run Matlab Distributed Computing Server.

an Hadoop cluster

. . . to scale your data processing.

an Ipython cluster

. . . parallelize the execution of your python code.

What issues you may find

Manual deployment and configuration is cumbersome and error prone

Too many home made shell scripts with lot of assumptions on the local infrastructure

Need to migrate deployment from one provider to another

What is elasticlluster

Elasticluster provides a user-friendly command line tool to create, manage and setup computing clusters hosted on cloud infrastructures like Amazon's Elastic Compute Cloud EC2, Google Compute Engine or a private OpenStack cloud).

Its main goal is to get your compute cluster up and running with just a few commands.

How does elasticluster work?

Command line tool

1. creates virtual machines in a cloud
2. installs and configures the software you want
3. add and remove nodes if needed

customization is done by editing text files

elasticluster demo

1. create 5 virtual machines on an OpenStack cloud.
2. install and configure Hadoop on them.
3. connect to the cluster.
4. Run an example.
5. destroy the cluster when done.

show time!

elasticluster demo

1. create 5 virtual machines on an OpenStack cloud.
2. install and configure Hadoop on them.
3. connect to the cluster.
4. Run an example.
5. destroy the cluster when done.

show time!

References

- Elasticluster on PyPI:
  https://pypi.python.org/pypi/elasticluster

```
$ pip install elasticluster
```

- Elasticluster github page:
  https://github.com/gc3-uzh-ch/elasticluster/
- Elasticluster web page:
  http://gc3-uzh-ch.github.io/elasticluster/
- Elasticluster documentation:
  https://elasticluster.readthedocs.org
- GC3 home page: http://www.gc3.uzh.ch
- Ansible home page: http://www.ansibleworks.com

www.bsc.es

Barcelona Supercomputing Center
Centro Nacional de Supercomputación

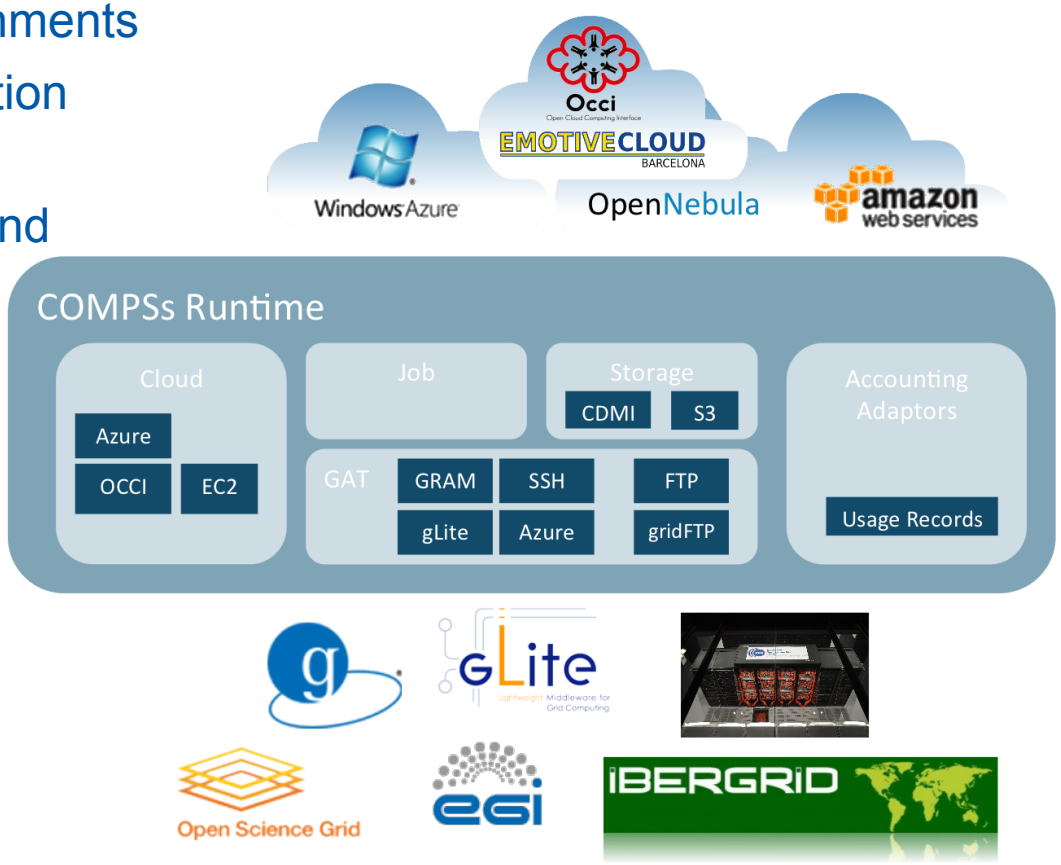
COMPSs in the EGI Federated Cloud

Daniele Lezzi – BSC

EGI Technical Forum 2013 - Madrid

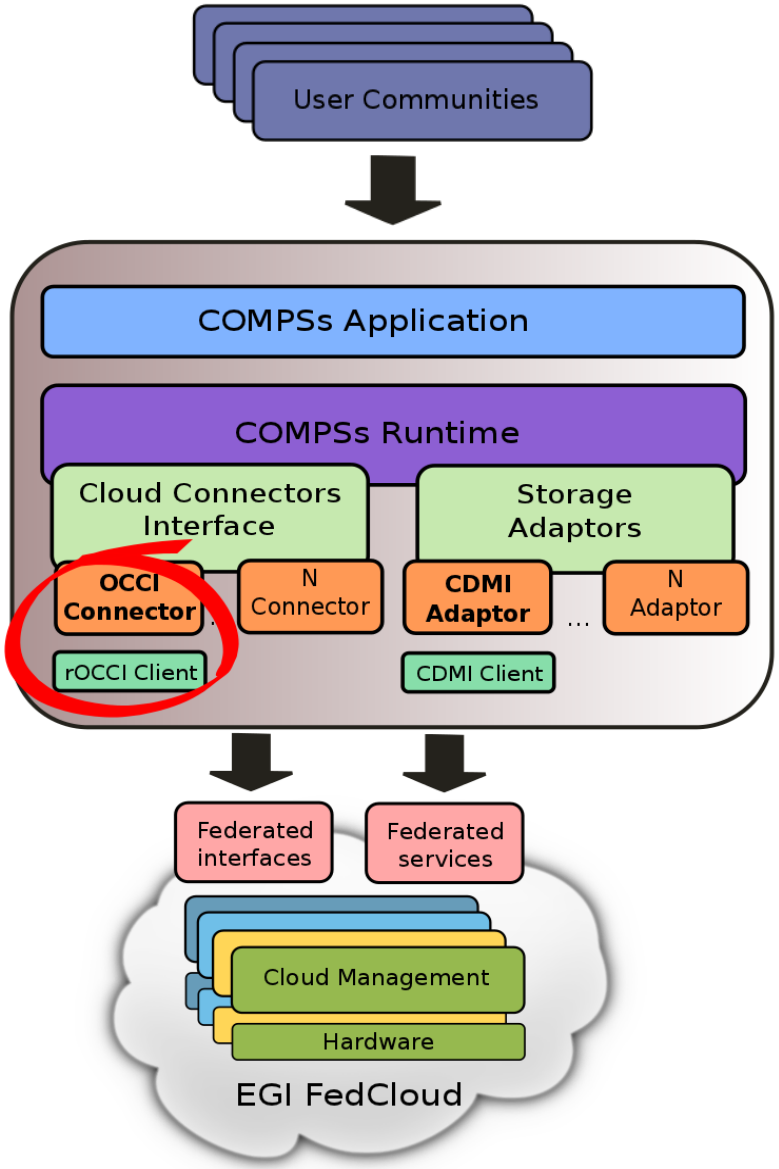
COMPSs: An Interoperable Framework

- Platform unaware programming model that simplifies the development of applications in distributed environments
- Low user intervention for application development
- Transparent data management and remote task execution
- Parallelization at task level
- Based on standards
- Cloud Elasticity and auto scaling



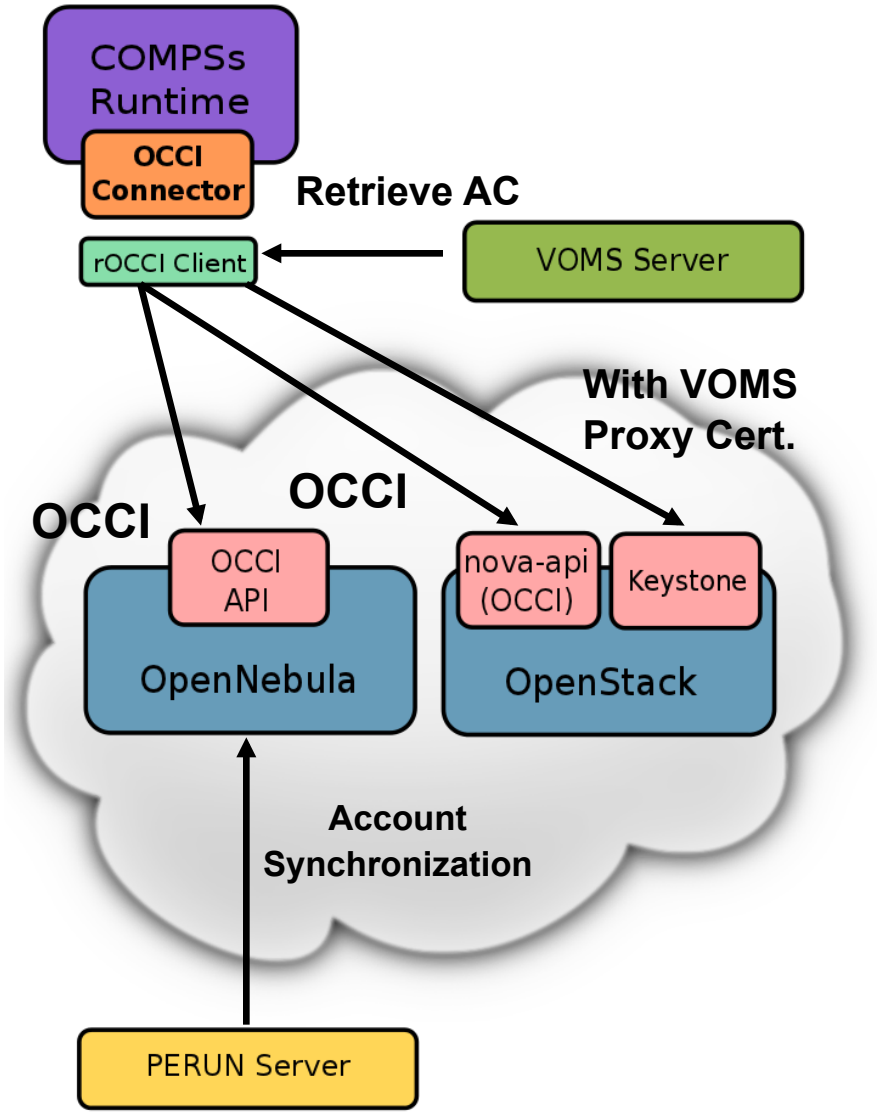
COMPSs integration with EGI FedCloud

- **COMPSs Application**: implementation of the application logic, where some tasks will be instrumented by the COMPSs runtime and executed remotely on EGI FedCloud resources.
- **Cloud Connectors**: implements a common interface allowing the resource management on an specific provider.
    - **OCCI Connector:** translates COMPSs resource management calls to OCCI operations.
- **Configuration**: each provider's available templates and images are set up on COMPSs configuration.



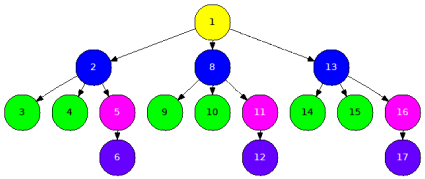
COMPSs integration with EGI FedCloud

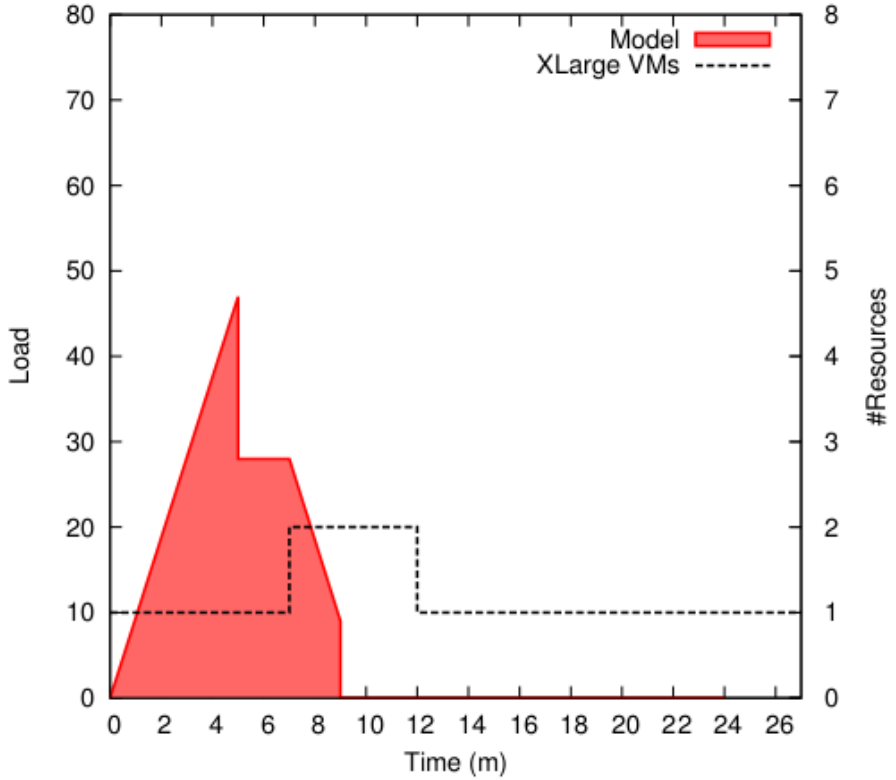
- Configure of COMPSs runtime (available cloud providers, endpoints, etc.)
- Configure of the OCCI connector (VM templates, VOMS credentials).
- Generate of the VOMS proxy certificate.
- Execute the COMPSs application.
- During the execution, some tasks could have hardware constraints (CPU, Mem, Disk, …)
- **OCCI Connector**: maps each task requirements to a suitable template of the available cloud providers, starting new VMs if needed.

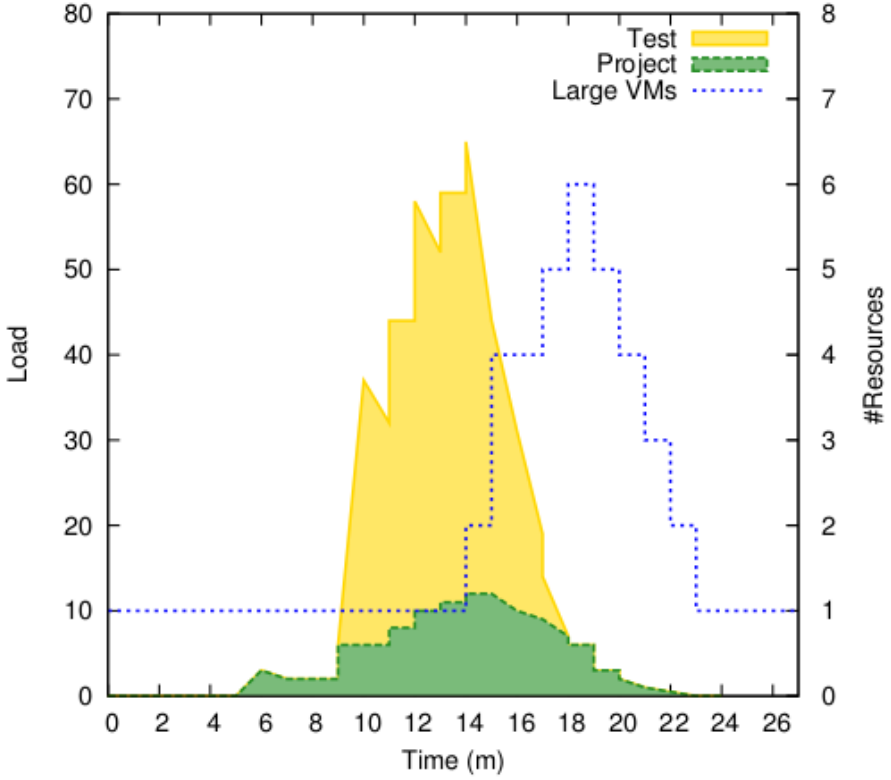


COMPSs integration with EGI FedCloud

Elastic management of resources

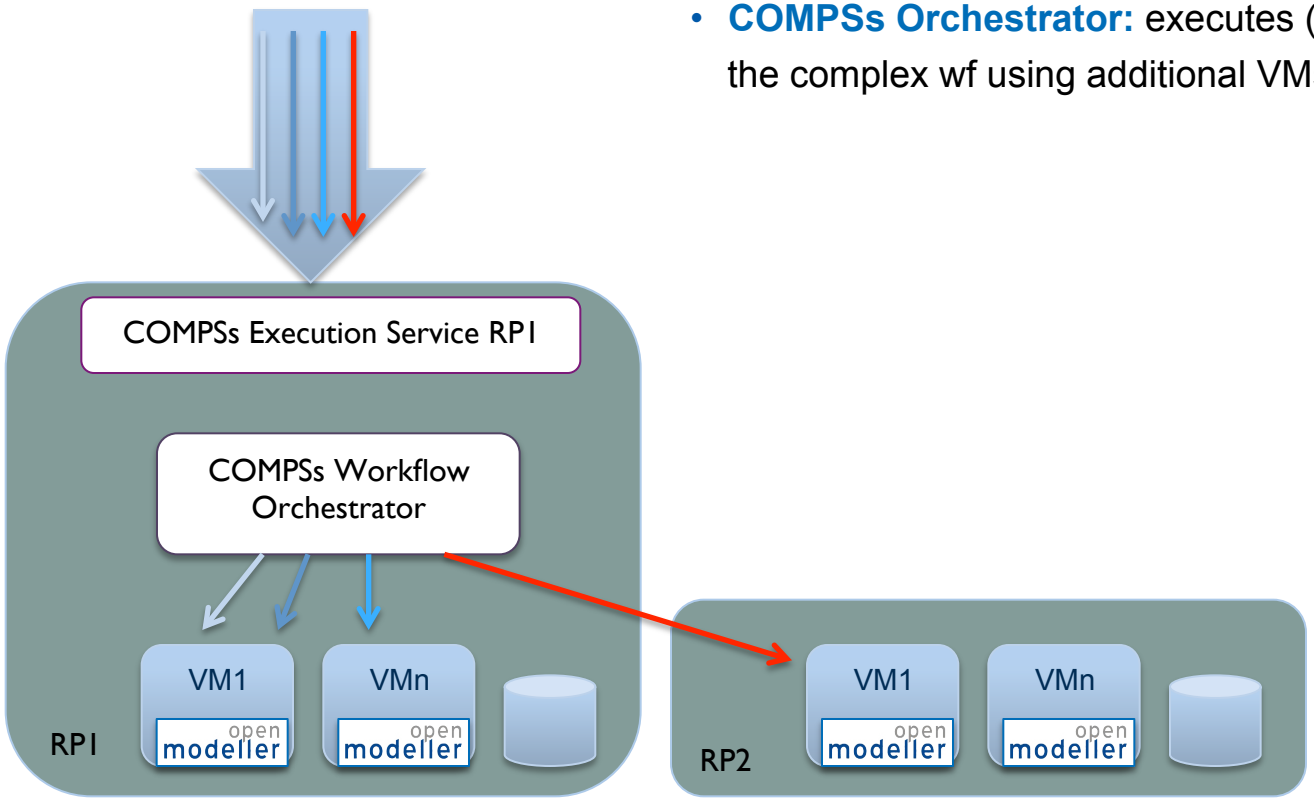






OpenBio in EGI: Hybrid Scenario

- **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.



www.bsc.es

Barcelona Supercomputing Center

Centro Nacional de Supercomputación

Thank you!

For further information

http://compss.bsc.es

daniele.lezzi@bsc.es

Federating Clouds in OpenStack

E. Fernández, Á. López (IFCA)
G. Borges, J. Gomes (LIP)
R. Valles (BIFI)

IberCloud

Objectives:

Investigate the requirements of scientific users of cloud technologies

Deploy a federated cloud IaaS testbed within the Ibergrid collaboration

Provide a unique user friendly interface for the services

Keystone

AuthN/AuthZ service

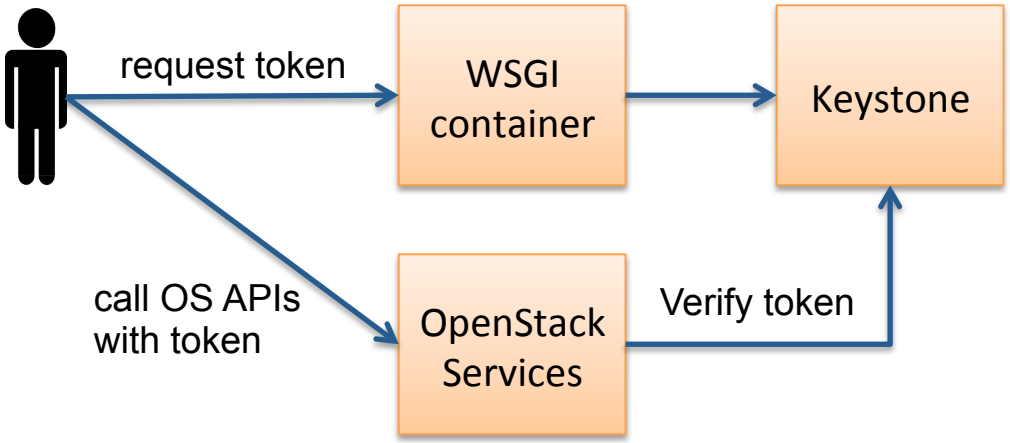
Extensible with WSGI filters

VOMS module:

- used in FedCloud
- intercepts Keystone calls and extracts proxy info mapping it to local users/tenants

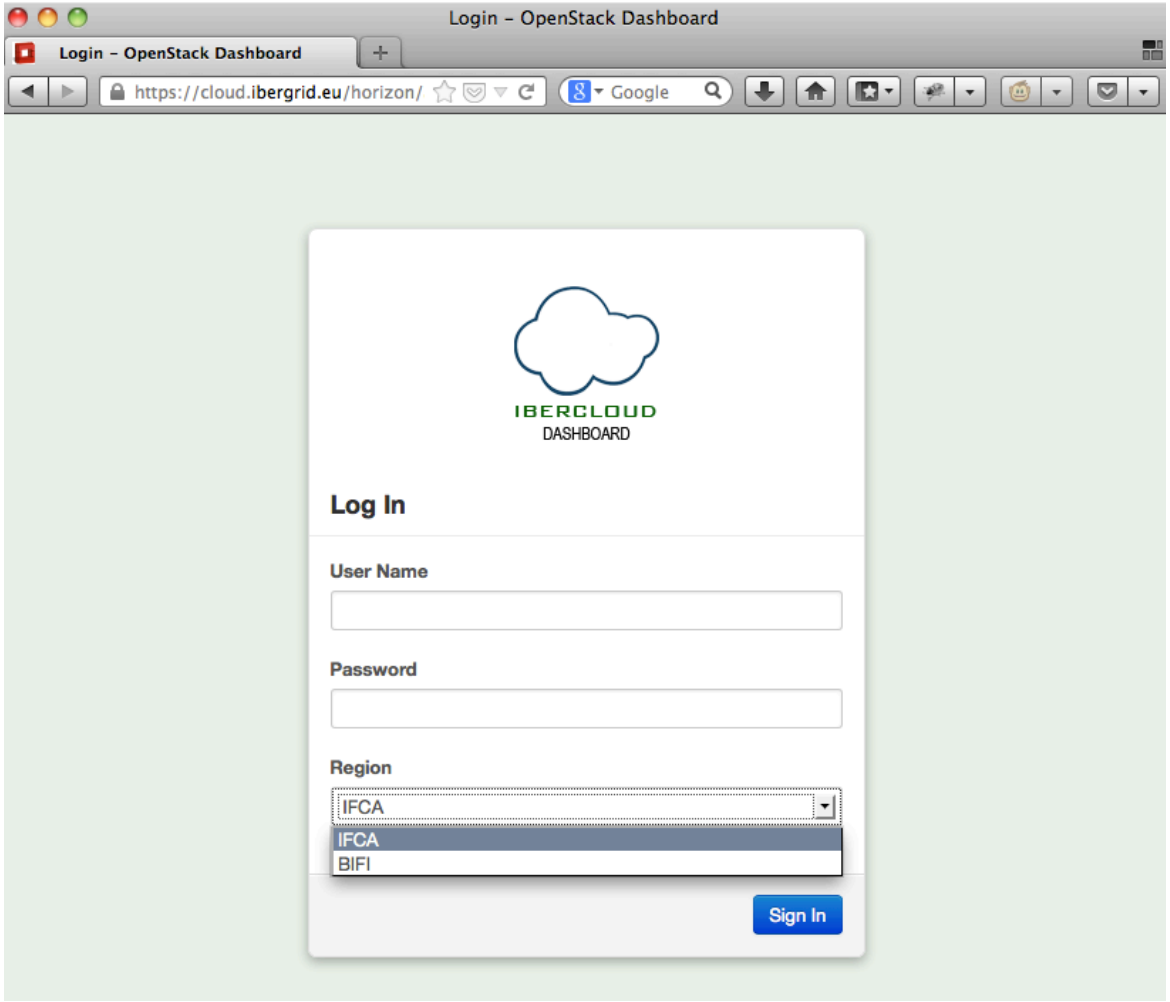
But:

- certificates not user friendly
- not straight forward in the browser



OpenStack LDAP SSO

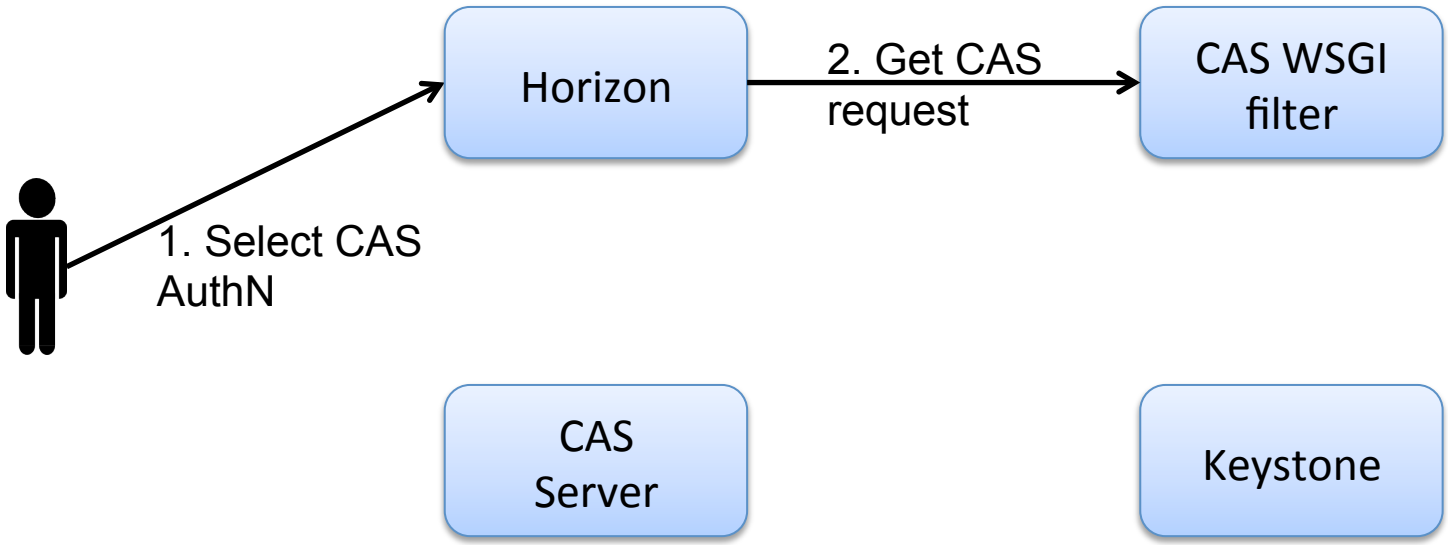
- Read only LDAP WSGI module
- Automatic user creation at the site with a configurable tenant
- Change between sites without re-entering credentials
- Dashboard retrieves unscoped token for all sites and keeps it in user profile
- Leverages region support of OS Horizon



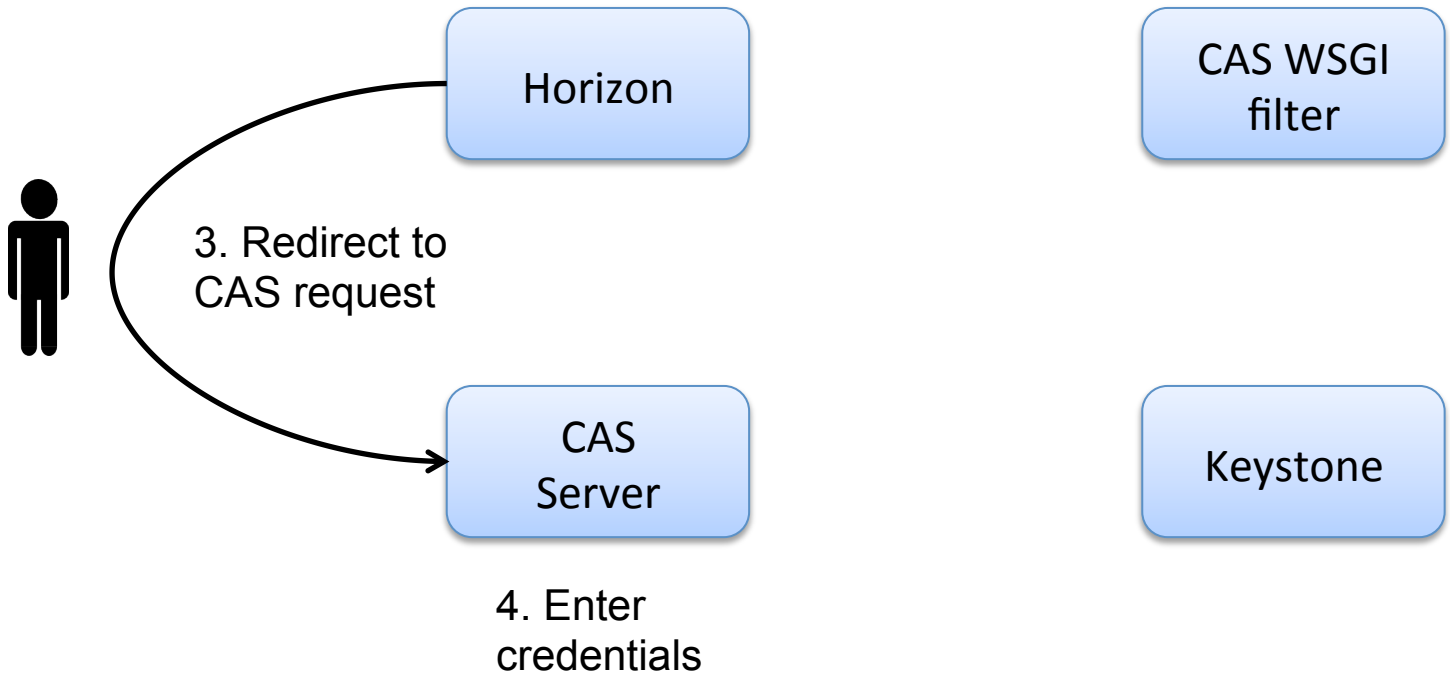
Beyond LDAP and VOMS

- Support for institutional identity:
  - Federated Keystone by University of Kent
    - code not yet officially included in OS
    - not easily deployable in production Keystone
  - Our approach:
    - try not to disrupt site operations: middleware filter checks credentials and creates users in keystone if necessary
    - New CAS (Central Authorization Service) & SAML modules

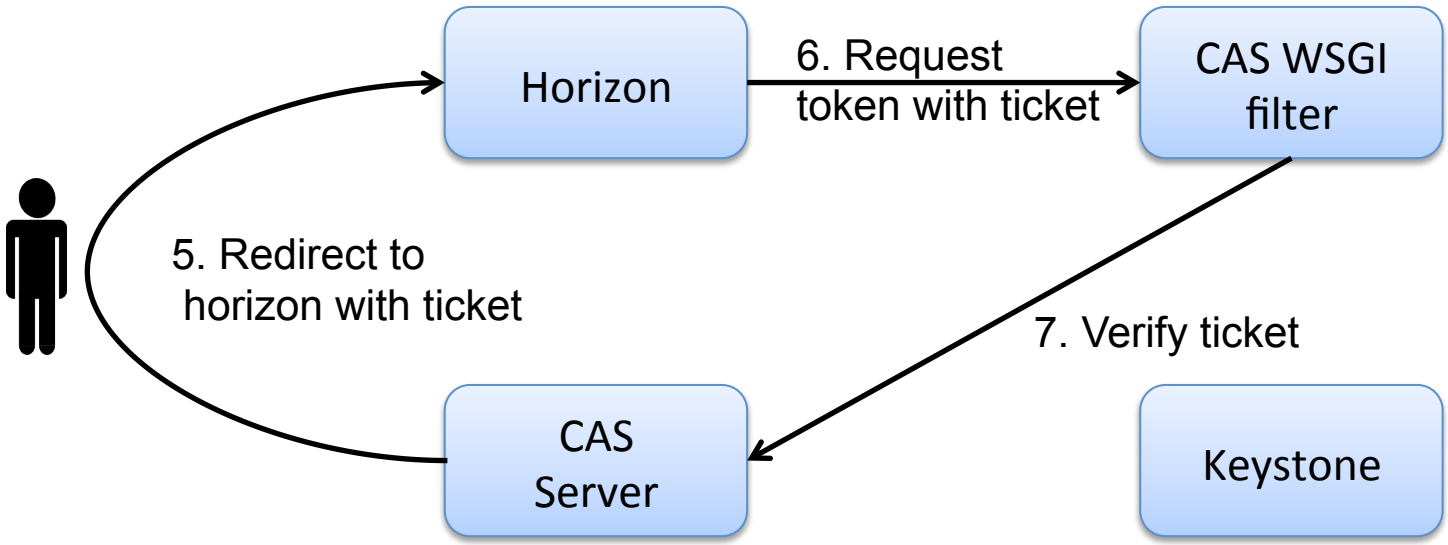
CAS + OpenStack



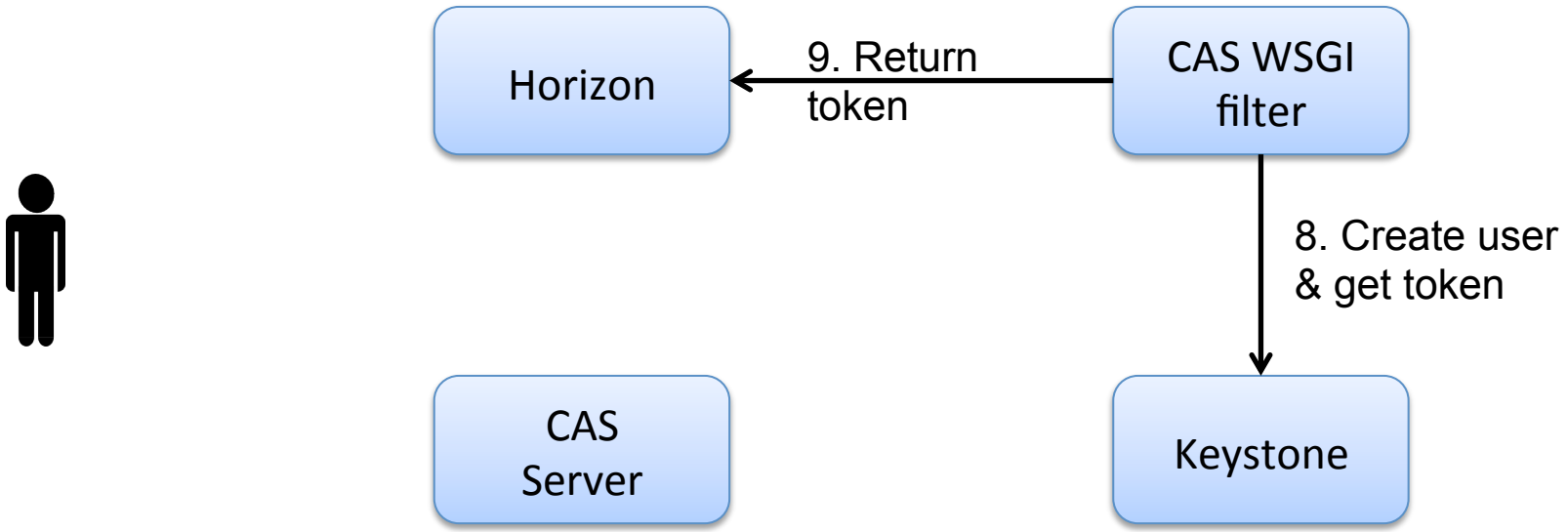
CAS + OpenStack



CAS + OpenStack



CAS + OpenStack



Thanks!

- Ibercloud - https://cloud.ibergrid.eu

- Fork it!
  - https://github.com/enolfc/keystone_cas
  - https://github.com/alvarolopez/keystone-ldap-ro
  - https://github.com/enolfc/django_openstack_auth
  - https://github.com/enolfc/ibercloud

www. OpenNaaS .org

EGI Technical Forum 2013

Joan A. García-Espín – jage@i2cat.net

OpenNaaS – Network as a Service

OpenNaaS is:

- Open-source software framework
- Free to download, personalise and use
- OSGI-based

Useful for:

- Provisioning Networks on Demand
- Deploying Dynamic configurations over the network
- Adapting Networks to different Applications

So… what?

- Seamlessly use and benefit from **network virtualisation**
- Network as "**Resources with Capabilities**"
- **Workflow** and **Job** concepts… in the network!
- **Mapping** between Cloud/App and Network Services

- Smartly manage:
  - Users accessing the network infrastructure
  - Network services together with Cloud/App services
  - Infrastructure modelling and virtual infrastructures
  - Configuration deployment

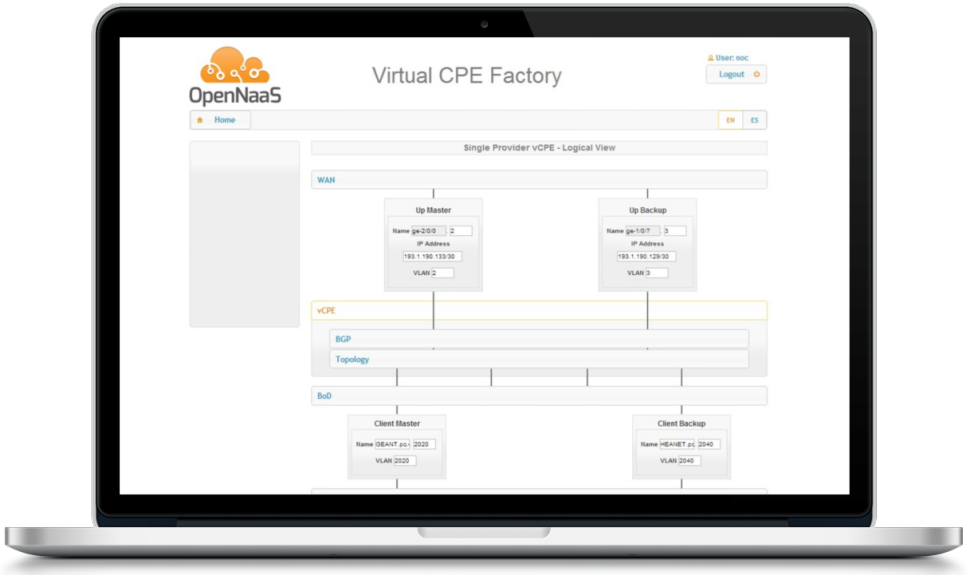
What we did: **Virtual CPE**

Goal

Rationalize the equipment deployed on site of the client by leveraging virtualization so that a single device at the client institution can share the role, or that it can be provided by a small number of aggregating devices at a central location.

Benefits

- Save hardware
- Faster operations
- Delegate to customer
- Gradual deployment
- Flexible management



What we did: **Inter-DC connectivity**

Goal

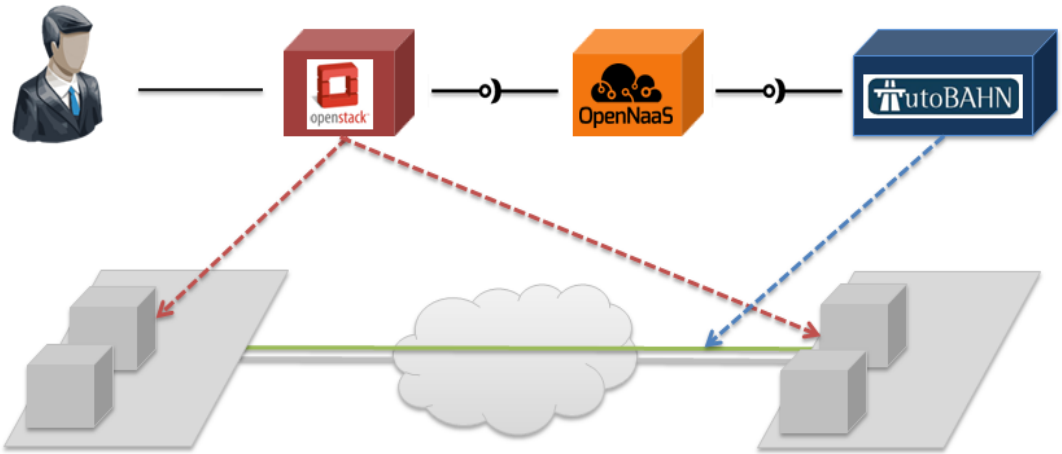
E2E Network service provisioning & management between DCs.

Software-based management solution leveraging virtualization technologies, applied on top of network and Data Center infrastructures. Components: OpenNaaS and OpenStack.

Benefits

Cloud and network services integration (IaaS & NaaS)

User-driven provisioning



get & compile & install

- Built it from scratch:

```
$ git clone git://github.com/dana-i2cat/opennaas.git
$ cd opennaas
$ git checkout develop # optional
$ mvn install
```

Fetch code

Fetch unstable
Built it

- Update from source:

```
$ cd opennaas
$ mvn clean
$ git pull git://github.com/dana-i2cat/opennaas.git
$ mvn install
```

Clean past built
Fetch updates
Built it

- Run it:

```
$ cp -r platform/target/opennaas-0.10 /srv
$ cd /srv/opennaas-0.10
$ ./bin/opennaas.sh
```

Enjoy!

Joan A. García-Espín
jage@i2cat.net

THANK YOU
MOLTES GRÀCIES

KTH-CLOUD

Resource Provider at EGI-Federated Cloud Task force

PDC Center for High Performance Computing

e-infrastructure

Where and What ?

Hardware:

- AMD 384 cores with 768 GB RAM
- 20 TB as NAS
- 1Gbps Ethernet
- Low latency Infiniband network

PDC Center for
High Performance Computing

Software:

- Currently Opennebula
- OpenStack *in transition*

Federation

VM Mangement

- OCCI 1.1. via rOCCI server egi.cloud.pdc.kth.se with support of x.509 based authn

Accounting

- via ossm/apel

InfoSys :

- via bdii on bdii.cloud.pdc.kth.se

VOMS federation with ssh keys contextualization

Experience

2009 First cloud deployment with Eucalyptus

Sept 2010 Part of EGI Fed cloud Task force .

2010 .. Provided services for several Nordic/European cloud projects like Venus-C , NEON, SNIC-Cloud and etc.

Value added services

- Dedicated research group for Data-intensive computing
  - Spark, Hadoop ..
  - iRods
  - Image analysis
- Specialized in Domains of
  - Life sciences*
  - Health sciences*

* *https://documents.egi.eu/public/ShowDocument?docid=1514*

Value added services

- Support for :
  - Galaxy http://galaxyproject.org/
- Central PDC service desk

Basically we love to go extra mile with our users ☺

Questions ?

Zeeshan Ali Shah,

zashah@pdc.kth.se

@zeeshanalishah

www.bsc.es

Barcelona Supercomputing Center
Centro Nacional de Supercomputación

The EUBrazilOpenBio-BioVeL Use Case in EGI

Daniele Lezzi – BSC

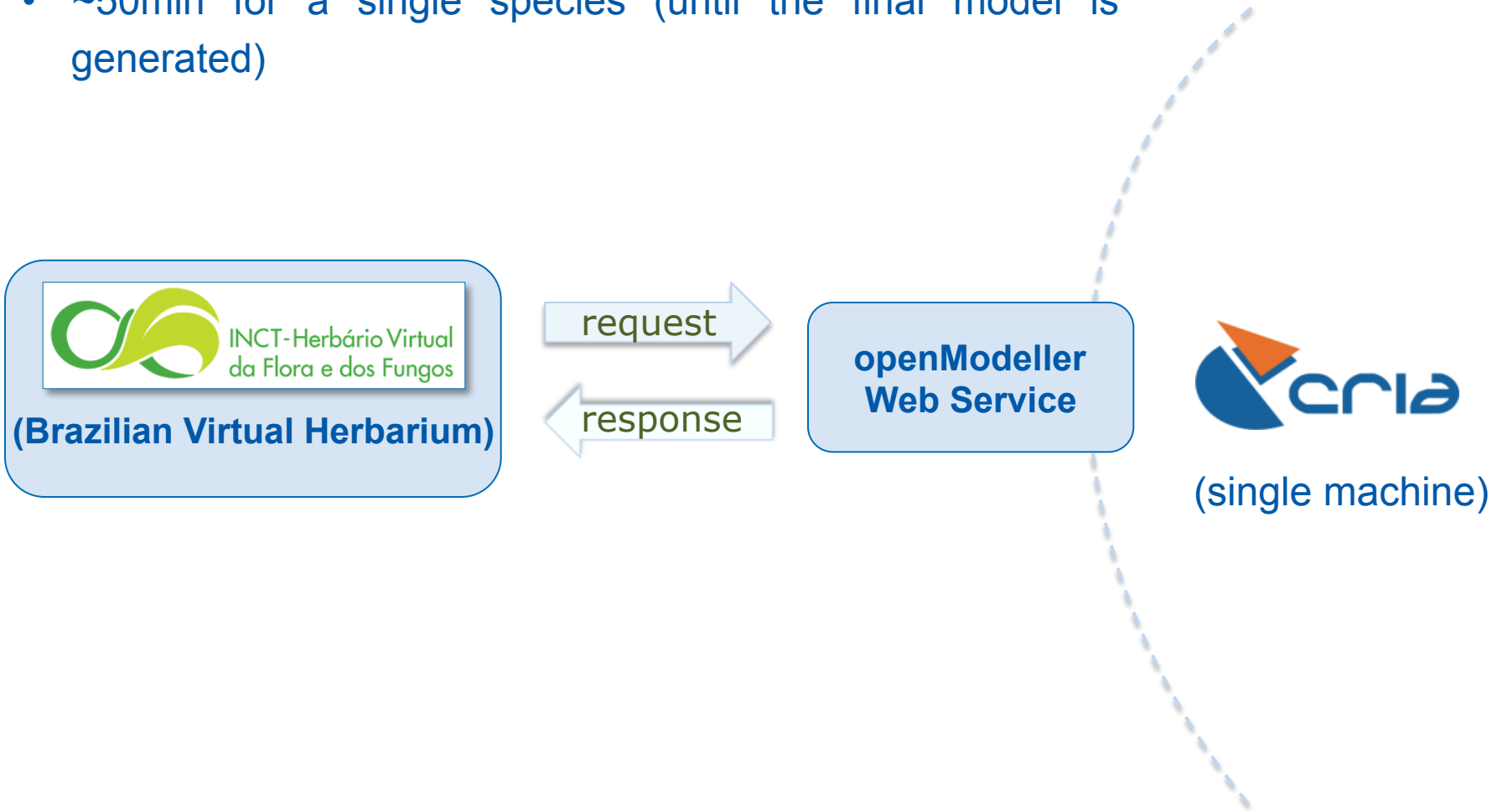
EGI Technical Forum 2013 - Madrid

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

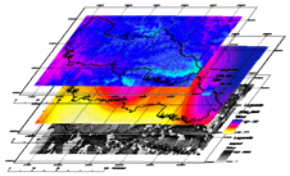
BioVeL

The Use Case: Ecological Niche Modelling

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



The Use Case: Ecological Niche Modelling

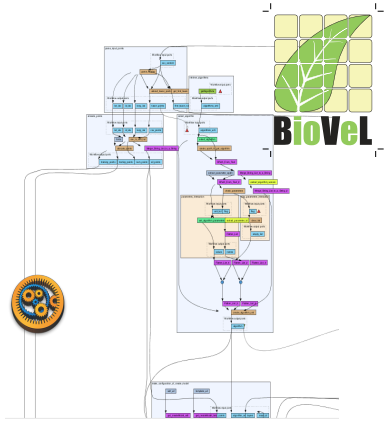


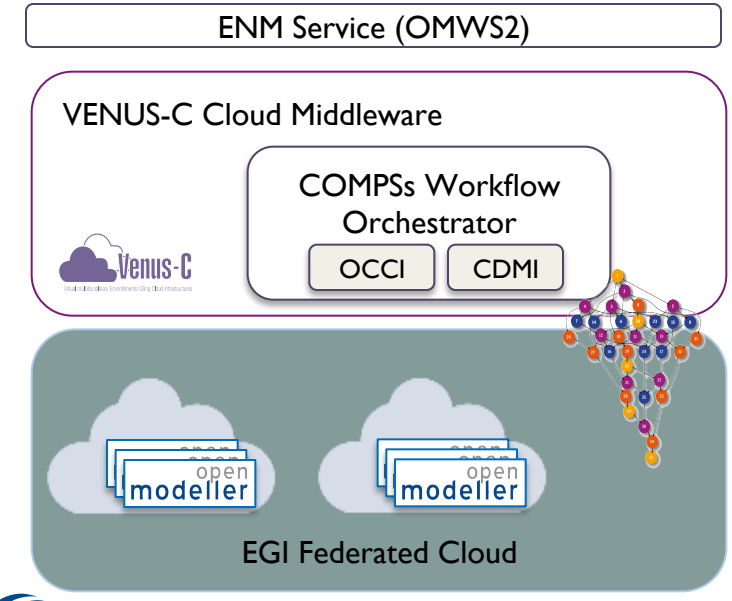
2012 lista de espécies flora do brasil

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

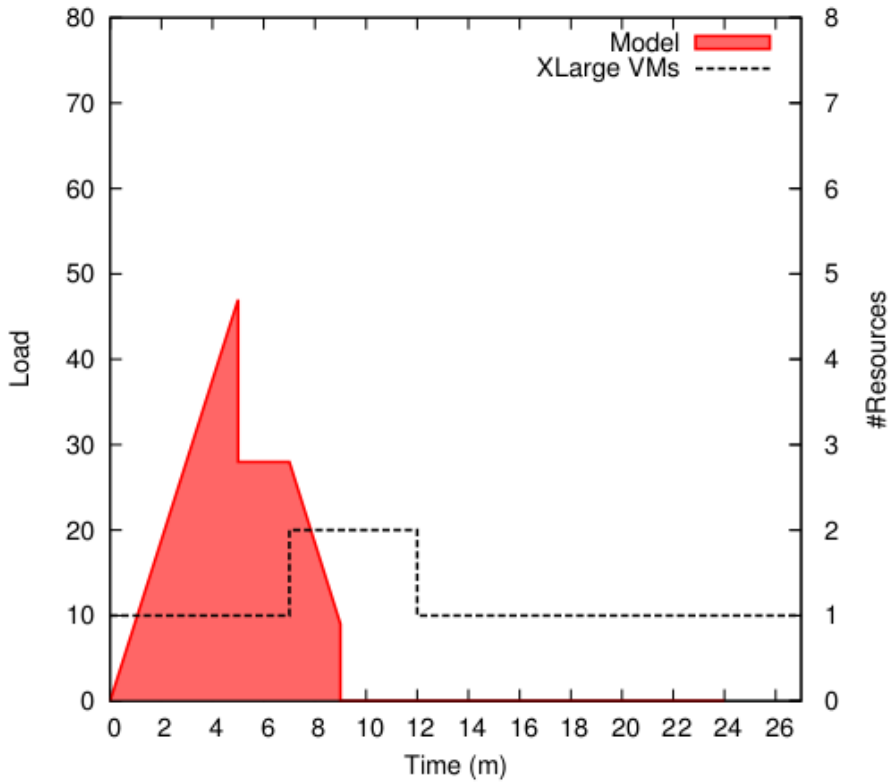


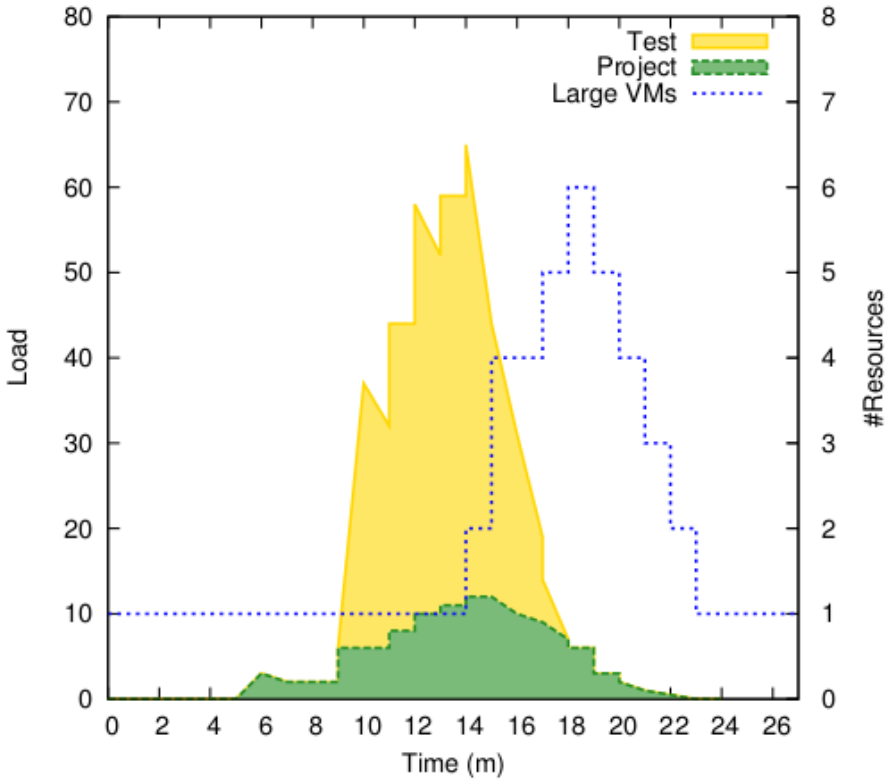


- **Shared requirements** between EUBrazilOpenBio and BioVeL
- The **EUBrazilOpenBio ENM service** is exposed through an extended openModeller Web Service interface (**OMWS2** in the picture).
- Such interface in EUBrazilOpenBio supports multi-staging and multiparametric experiments implemented through **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.

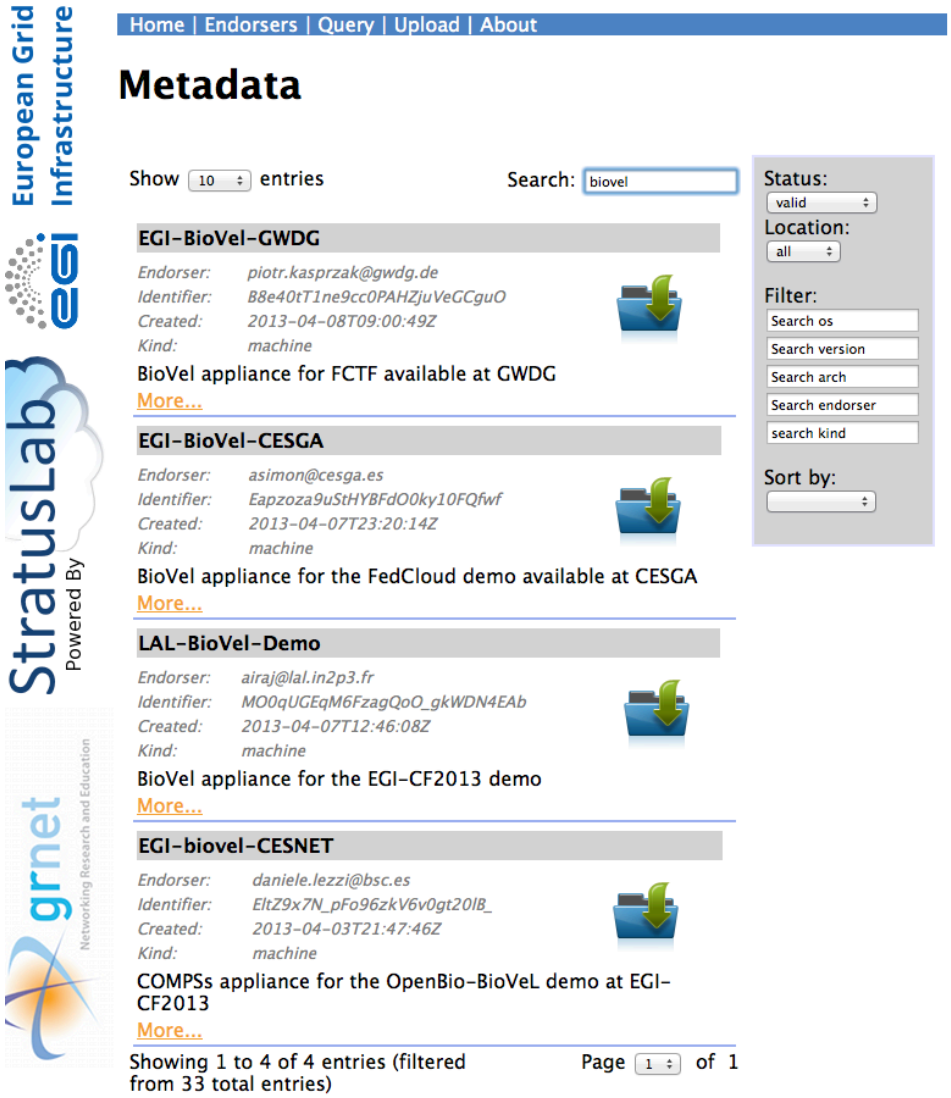
Integration in EGI: An interoperable and scalable solution

- Execution of oM workflows optimized by COMPSs
- Serve multiple independent requests properly managing the pool of resources





Integration in EGI: An interoperable and scalable solution



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

Personal Genomes Project
as a potential EGI community

Next Generation Federated HPC infrastructure to drive international genome discovery

Peter Walgemoed
Carelliance & Dutch Health Hub

http://DHH-IPC.nl
peterwalgemoed@gmail.com

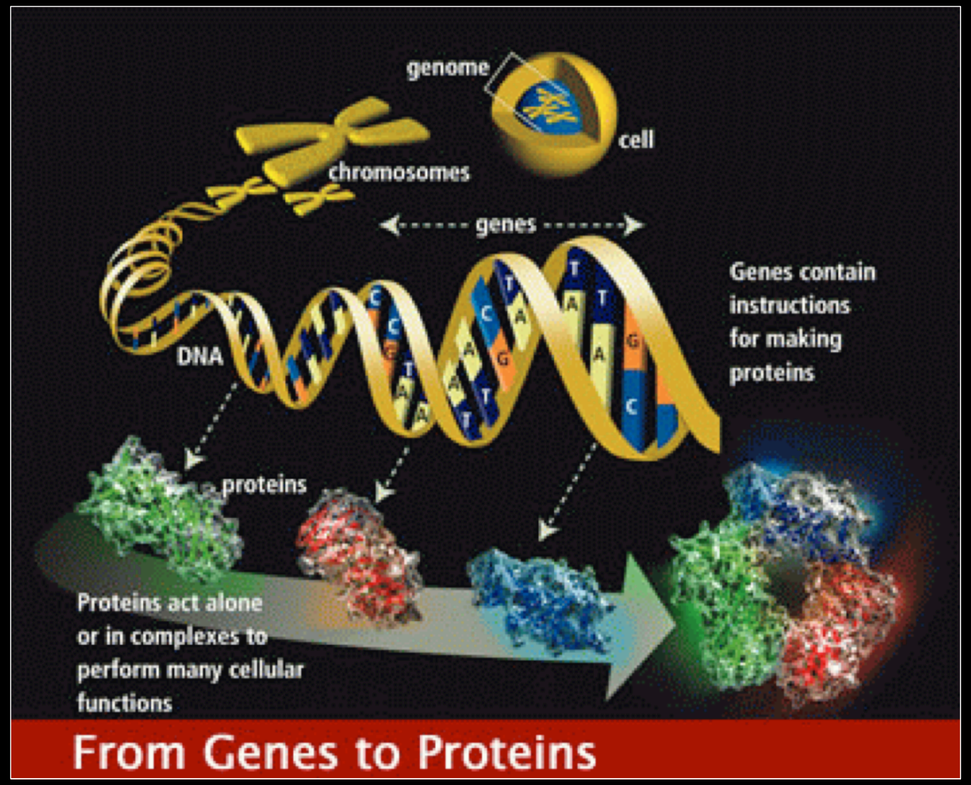
Presented by
Ad Emmen
Dutch Health Hub & Contrail

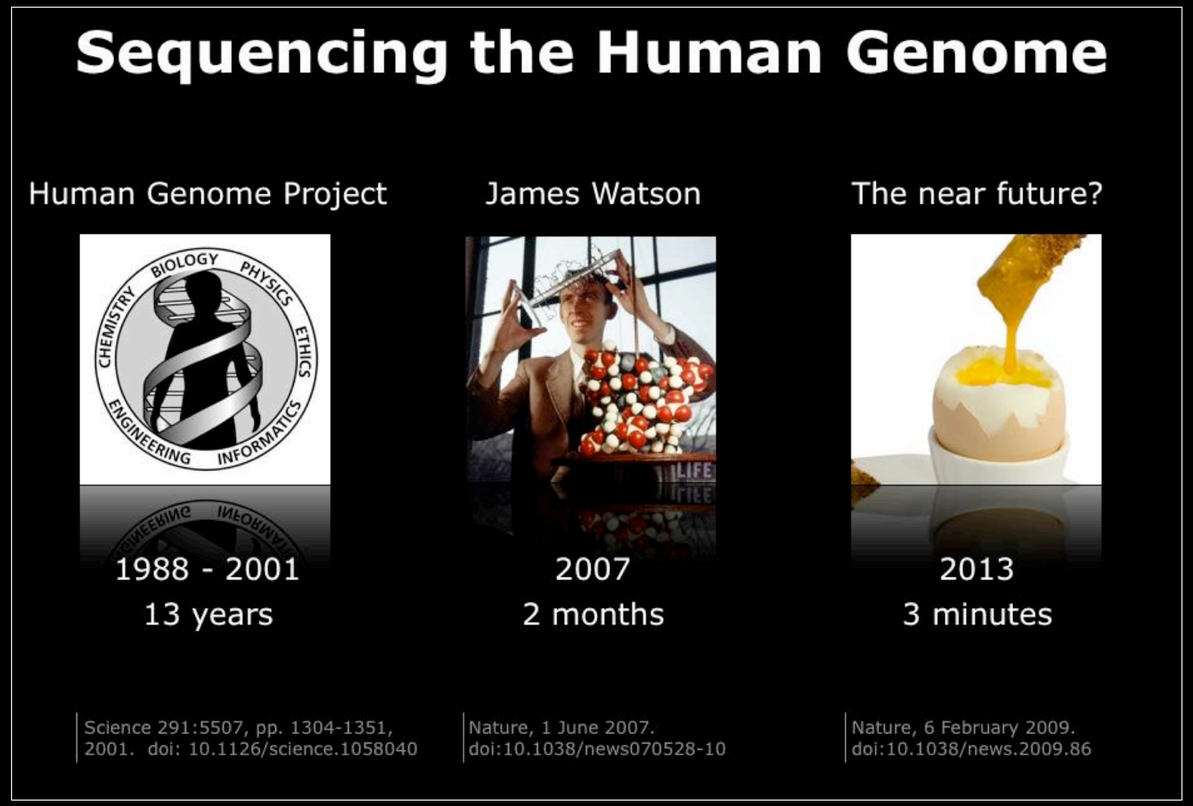
Dutch HealthHub
smart data solutions

contrail
open computing infrastructures for elastic services

carelliance

Community: Personal Genomes Project





http://www.personalgenomes.org

ARVADOS

Open source platform for managing and analyzing biomedical big data

Usage catching on in genome community

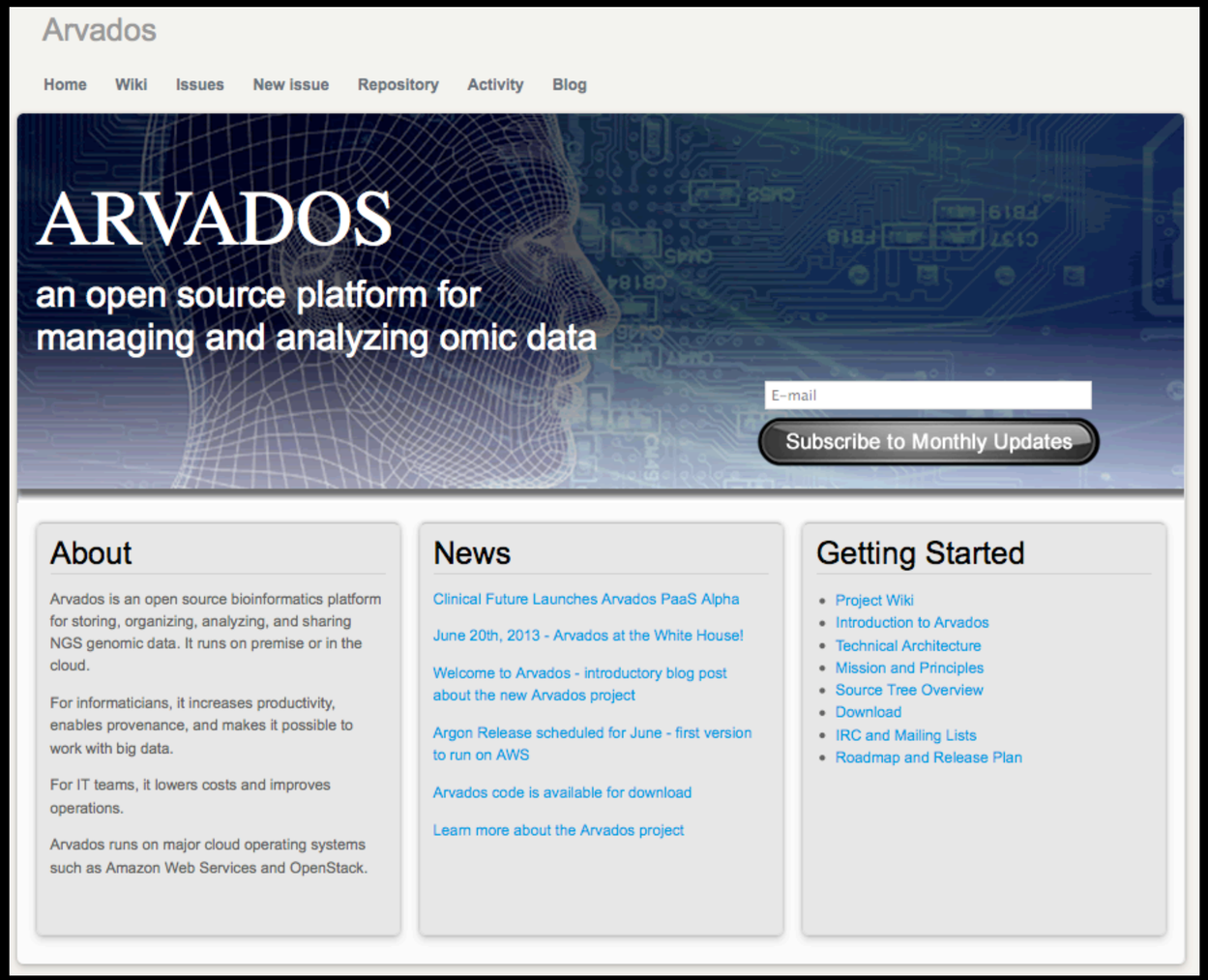
http://arvados.org

ARVADOS

Open source platform for managing and analyzing biomedical big data

Usage catching on in genome community

http://arvados.org



Challenges

1. Store and organize 100's of TB's of large files with multiple meta-data schema
2. Run informatics analyses that do distributed computations on very large datasets
3. Do real-time high-performance queries on compact genome data (e.g. variants)
4. Ensure validity and maintain provenance on all data in the system over time
5. Make it easy to reproduce pipelines exactly as they were done in the past
6. Protect all data with flexible access control rules and strong encryption
7. Share large data sets between data centers and organizations without physically moving data

Arvados technology

Apps

Cancer Diagnostics

Variant Visualization

etc.

Application Framework (APIs and SDKs)

Arvados

Sharing

Real-Time Analysis of Genomic Data "Lightning"

Governance

Distributed Computation "Crunch"

Provenance

Analysis

"Keep" Data Storage & Management

Cloud Operating System

Cloud

Arvados Cloud

Private Cloud

Public Cloud

Arvados technology

Apps

Cancer Diagnostics

Variant Visualization

etc.

Application Framework (APIs and SDKs)

Arvados

"Keep" Data Storage & Management

Analysis

Provenance

Distributed Computation "Crunch"

Governance

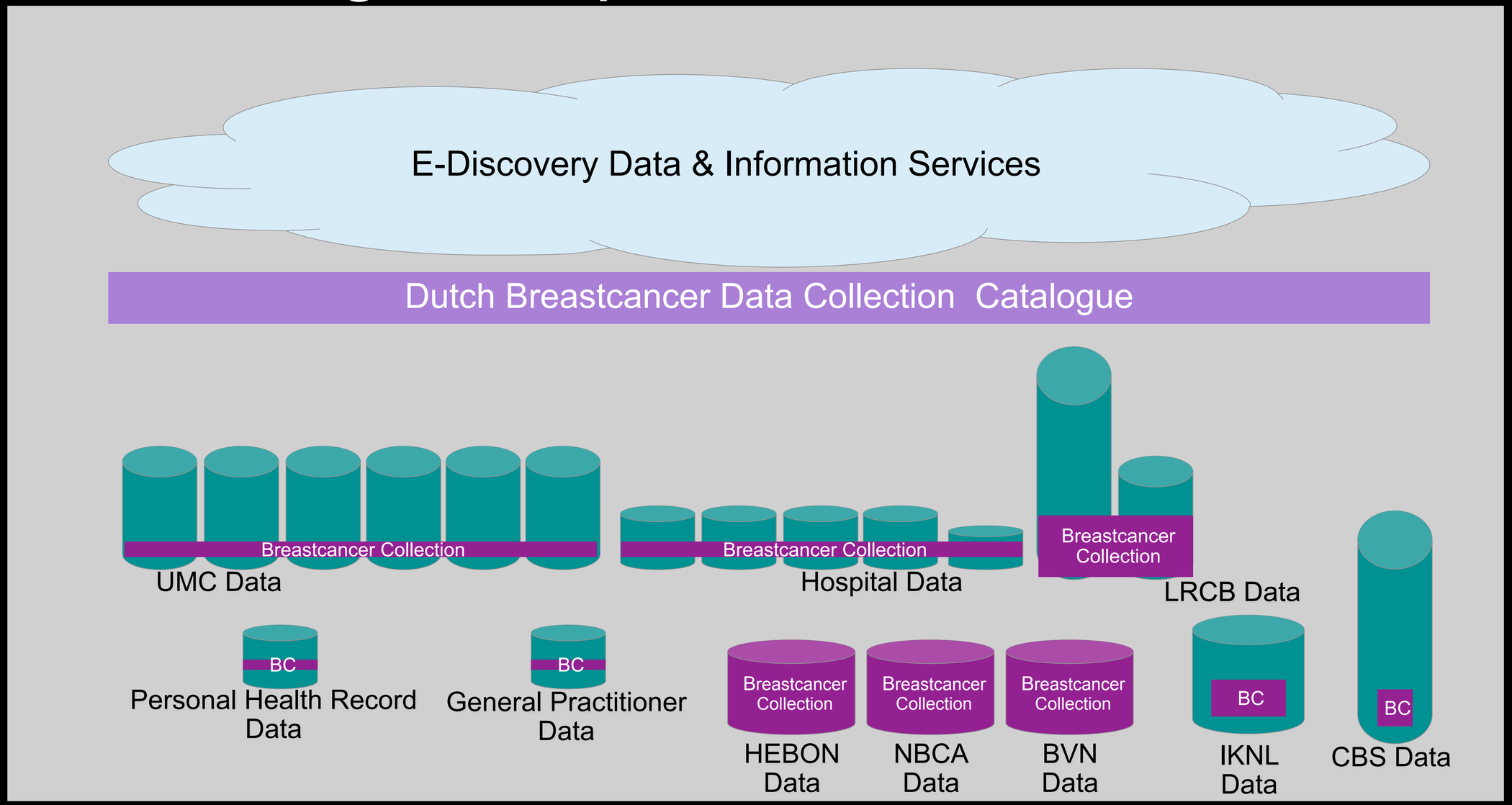
Real-Time Analysis of Genomic Data "Lightning"

Sharing

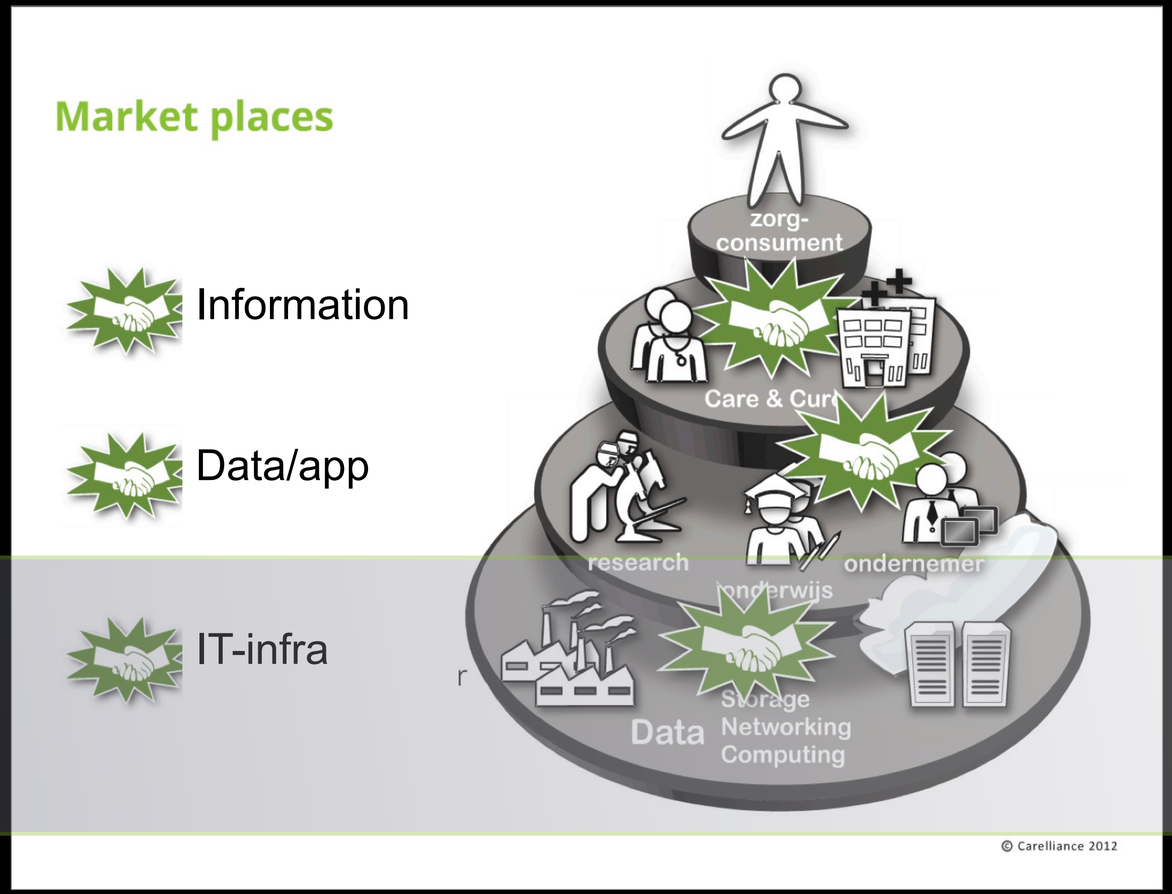
EGI Fed Cloud

Investigate
Integration with EGI federated Cloud

Trusted Digital Repositories

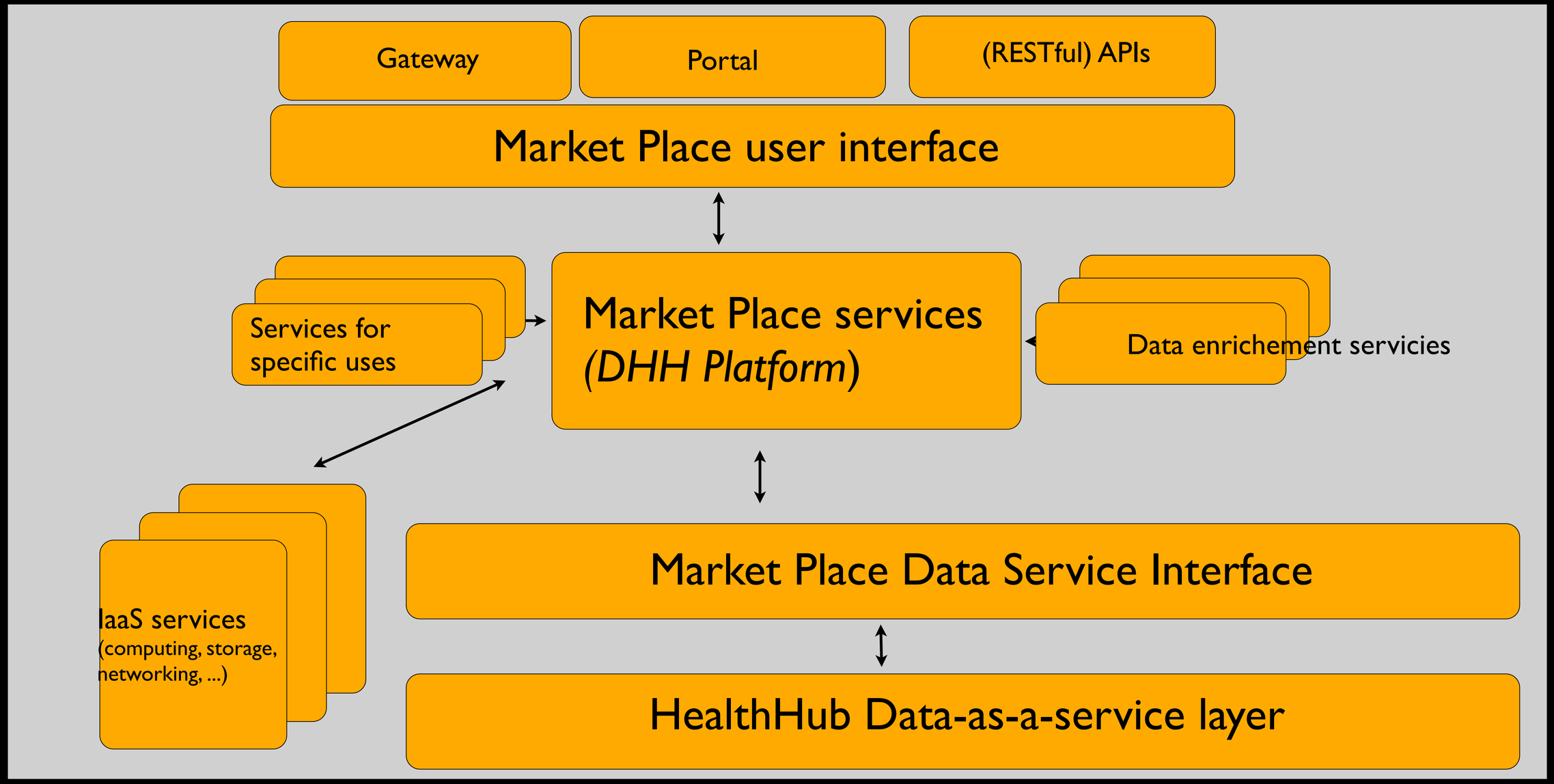


Service Marketplaces
Dutch Health Hub start

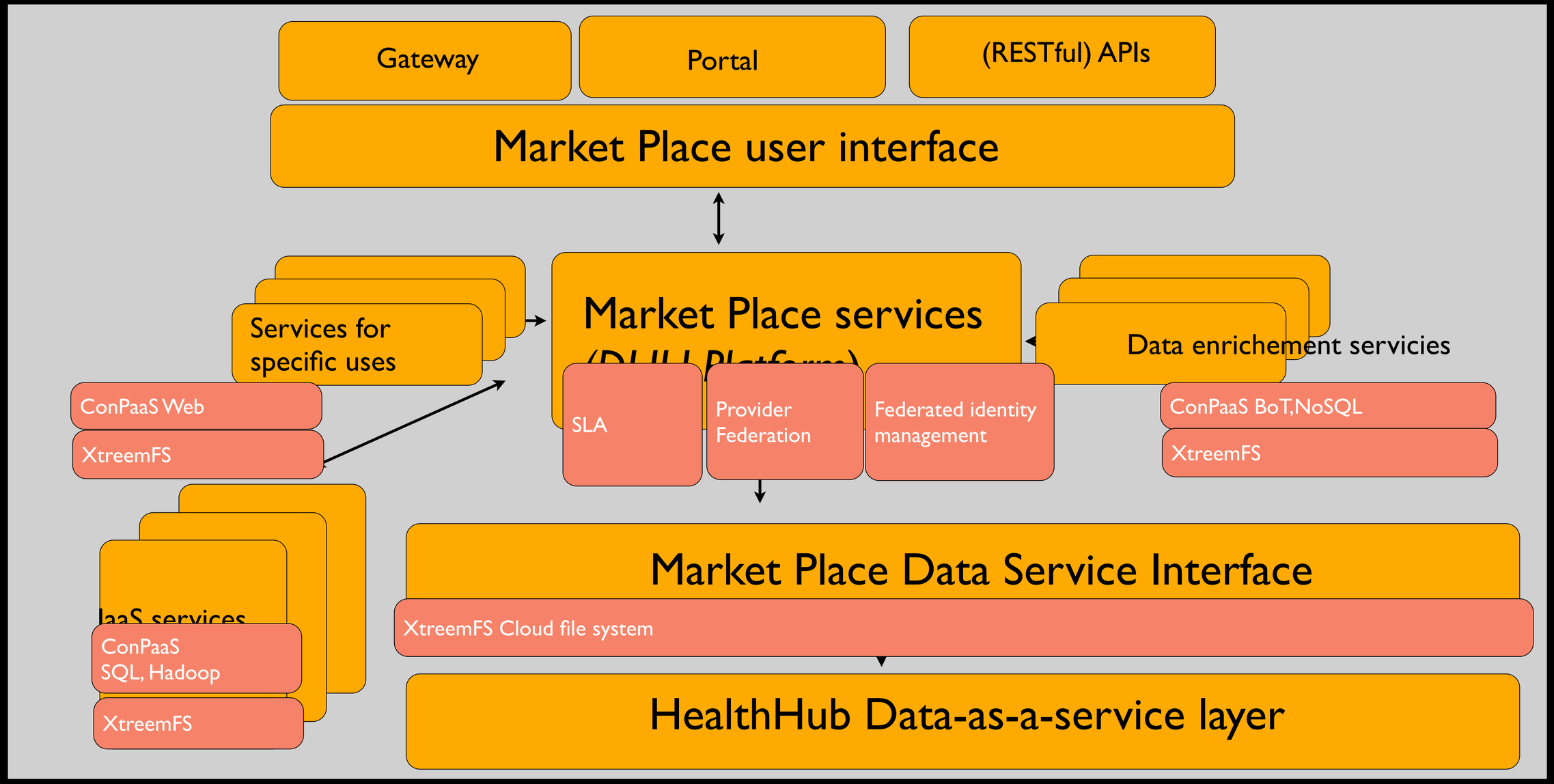


Infrastructure-as-a-Service
(IaaS) marketplace

Some Contrail tools for Dutch Health Hub Market place



Some Contrail tools for Dutch Health Hub Market place



Next steps

1. Organise interest in community in Europe
2. Implement test environment as part of Dutch Health Hub
3. Investigate integration with EGI Federated Cloud

End

Let’s vote!