



# Federating Infrastructure as a Service cloud computing systems to create a uniform e-infrastructure for research

**David C.H. Wallom**<sup>1</sup>, Matteo Turilli<sup>1,2</sup>, Michel Drescher<sup>3</sup>, Diego Scardaci<sup>3,4</sup> and Steven Newhouse<sup>3,5</sup>

<sup>1</sup>University of Oxford, Oxford, UK

<sup>2</sup>Rutgers University, New Jersey, USA

<sup>3</sup>EGI.eu, Amsterdam, NL

<sup>4</sup>INFN Division of Catania, Catania, Italy

<sup>5</sup>EMBL-EBI, Hinxton, UK



# Overview

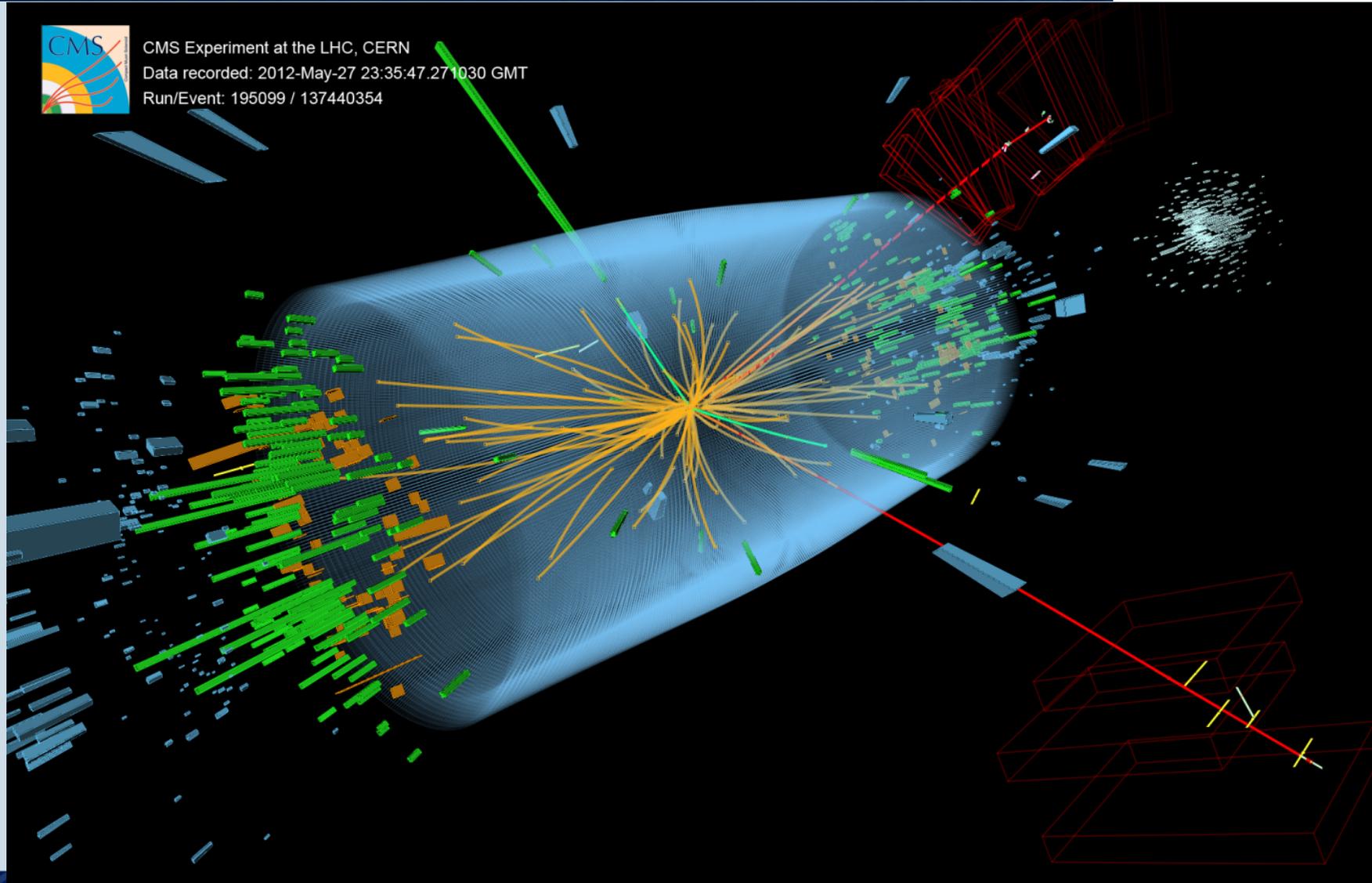


- Why?
- What?
- Who?
- Future?

# Success with the Grid



CMS Experiment at the LHC, CERN  
Data recorded: 2012-May-27 23:35:47.271030 GMT  
Run/Event: 195099 / 137440354





## Growth of Providers

- Grid Computing
  - Academic resource providers
- Cloud Computing
  - Diversity of resource providers

## Growth of Research Communities



**Tens of 1000's**

Few related use cases  
Single application model

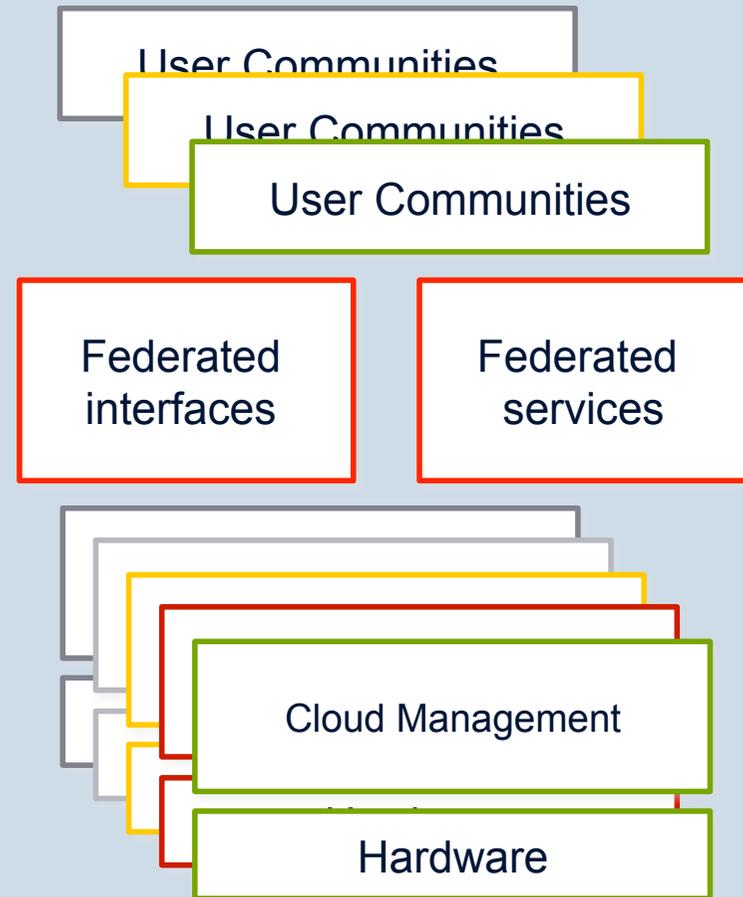


**Millions**

Many diverse use cases  
& application models



- **Provider agnosticism:** the only condition to federate resources is to expose the chosen interfaces and services.
- **Heterogeneous implementation:** no mandate on the cloud technology, protecting resource provider investment
- **Standards and validation:** Use only recommended and common open standards for the interfaces and images.
- **Resource integration:** Cloud Computing to be integrated into the existing production infrastructure, protecting prior EGI investment and increasing ROI

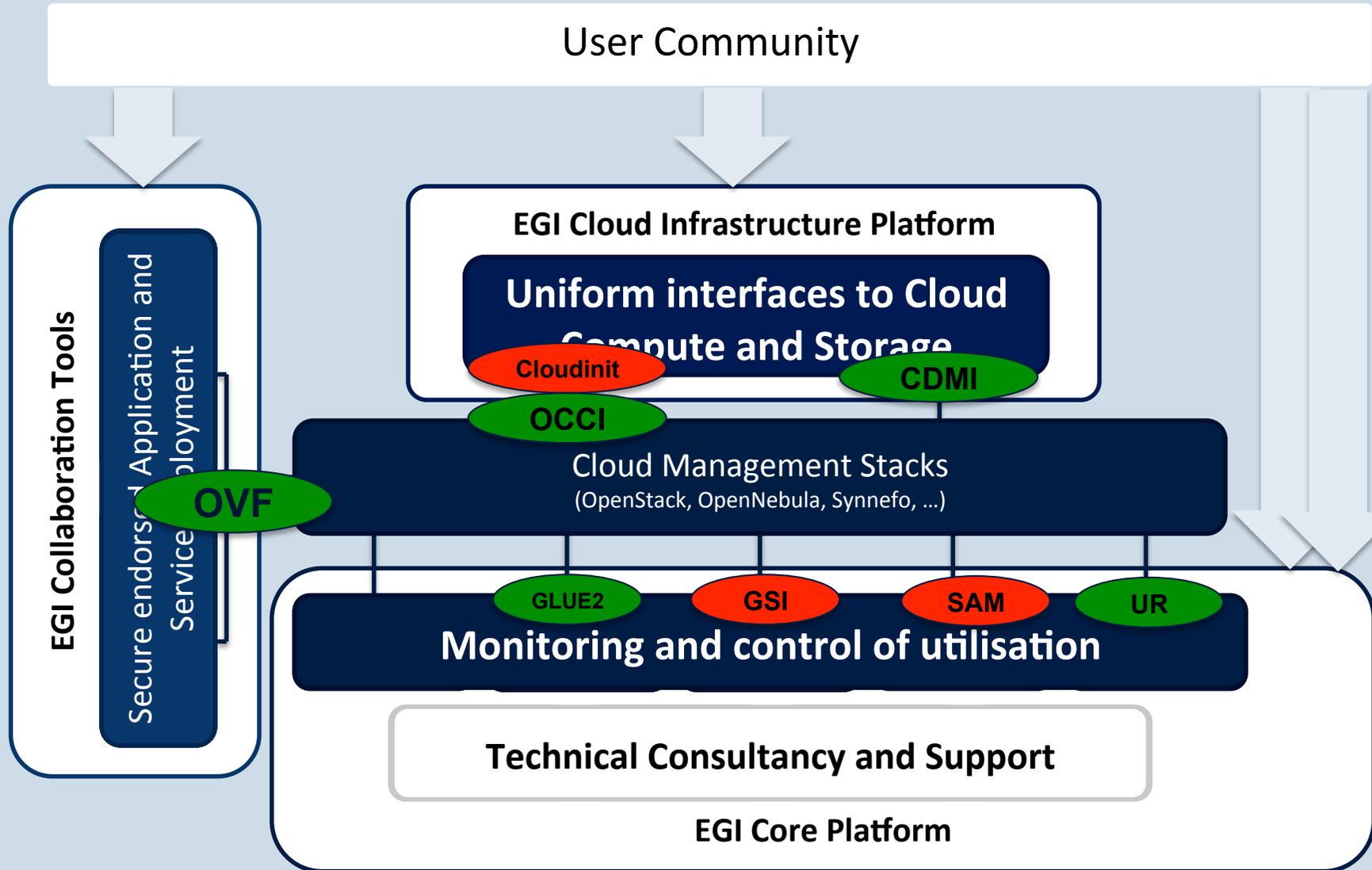




- **Total control** over deployed applications
- **Elastic resource consumption** based on real needs
- **Workloads** processed on-demand
- **Endorsed and accredited applications** available from multiple different communities shared
- **Single sign-on** at multiple, independent providers
- **Centralised access** to service information across multiple providers

#	Capability	User Stories
1	<b>Virtual Machine management</b>	“I want to instantiate a single existing VM image on a remote cloud.”
2	<b>Data management</b>	“I want to instantiate a VM instance from an image that I have created and is not on the cloud I wish to use.” “I want to associate my running VM with a data set in the Cloud.” “I want to take snapshots of my running VM for restart purposes.”
3	<b>Integrated information system</b>	“I want to choose on which resource provider I want to start my single VM.” “I need to know about the Virtual Machine Manager (VMM) capabilities the provider offers.”
4	<b>Accounting</b>	“My usage across different resource providers needs to be recorded and reported to multiple aggregators.”
5	<b>Availability &amp; Reliability</b>	“Information relating to the availability/reliability and current status of the remote virtualised resource needs to be available to me.”
6	<b>VM &amp; Resource state change notification</b>	“When the status of the [VM] instance I am running changes (or will change) I want to be told about it.”
7	<b>Integrated AAI</b>	“I want to use my existing identity, and not re-apply for new credentials to use each component of the service.”
8	<b>VM Image Management</b>	“I want to use a single VM image across multiple different infrastructure providers.”
9	<b>Brokering</b>	“I want my VM instance to run on a resource that is suitable based on a set of policies or requirements rather than my choosing directly which resource will run it.”
10	<b>Contextualisation</b>	“When I deploy a VM instance on a resource I want to give it configuration information for customisation of the default template. This can only happen when it is up and running.”

# Cloud within the EGI Infrastructure Platforms



# Cloud technology Integration Level



Cloud Management Stack	Integration					
	Fed. AAI	Monitoring	Accounting	Img. Mgmt.	OCCI	CDMI
OpenStack	Y	Y	Y	Y	Y	Y
OpenNebula	Y	Y	Y	Y	Y	Y
Synnefo	Y	Y		Y	Y	Y
WNoDeS	Y				Y	
StratusLab*	Y	Y	Y		Y	
Cloudstack					Y	
Emotive	Y					Y

# Resource Allocation



- e-Grant supports;
  - Resource request
  - Resource availability
- 5,000 job slots and 170 TB of storage aside
  
- Preliminary Units of Allocation

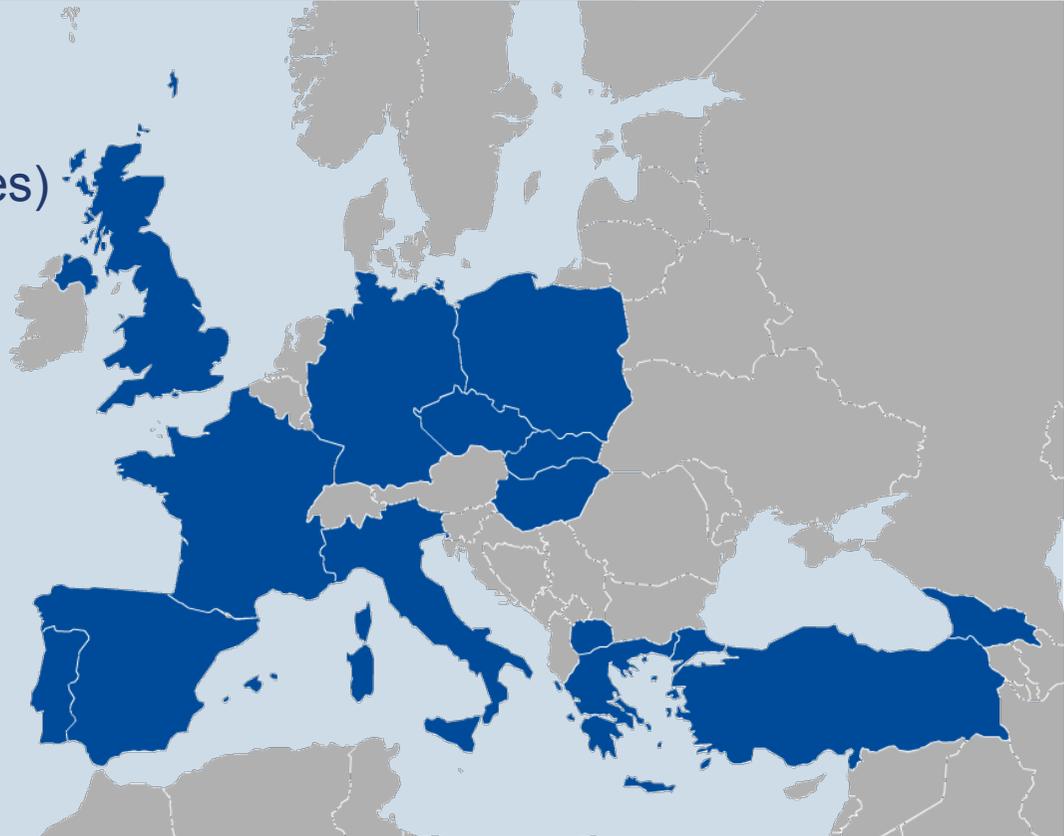


Name	# of vCPU	Mem (GB)	Storage (GB)
Small	1	2	1 x 20
Medium	2	4	1 x 40
Large	4	8	2 x 80
Other	>2	>7.5	n x >40

# FedCloud Infrastructure



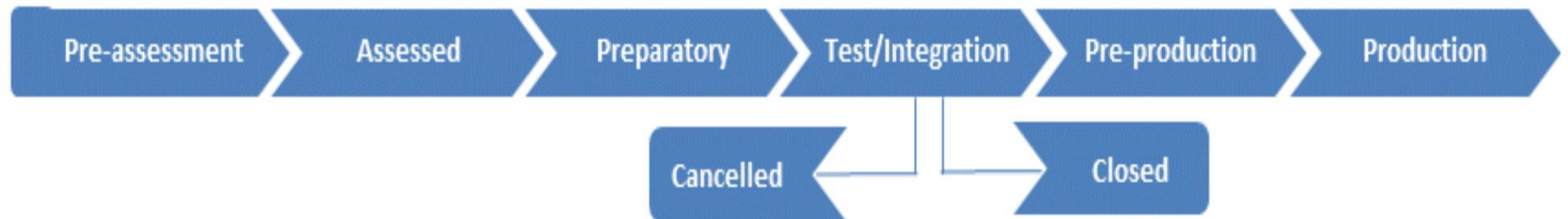
- One year of production
- Resources
  - 21 providers from 14 NGIs (Countries)
    - 55% Openstack, 42% Open Nebula, 3% Syneffo
  - 17 interested in joining from 7 new NGIs
- Usage
  - ~700K VMs
  - ~18M CPU hours





- **Service Hosting**
  - Long-running services (e.g. web, database or application servers)
- **Compute and data intensive workloads**
  - Batch and interactive (e.g. IPython, R, MATLAB) with scalable and customized environments not limited to the traditional job model
- **Datasets repository**
  - Store and manage large datasets for your applications
- **Disposable and testing environments**
  - Host training events, test new developments and applications without overhead

# Support Model



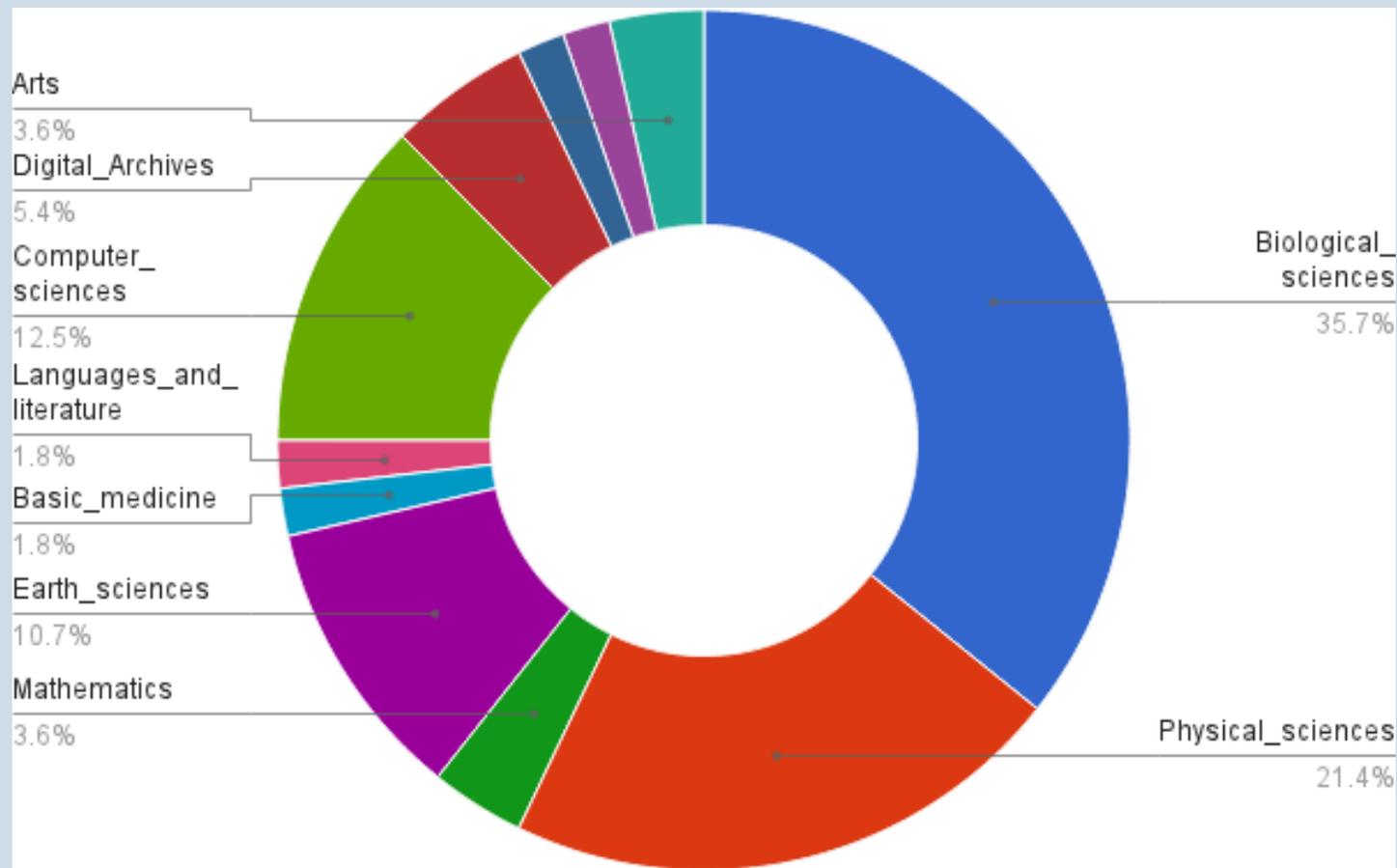
Status	Description
Pre-assessment	FCUS members identified a potential use case that can profit from the EGI Federated Cloud services
Assessed	The use case requirements are assessed with relevant information added to a dedicated wiki page, to manage the full porting & integration process
Preparatory	The user & FCUS team setup the environment to execute tests on the FedCloud (configuration of client environment, identify resource providers, upload vm images, ...)
Test & Integration	Users are accessing the FedCloud through the fedcloud.egi.eu VO to execute tests while integrate their applications to the infrastructure
Pre-production	Test & integration phases were successfully completed. Create a production Virtual Organization (VO) or join an already existing VO
Production	Use case is being used in production. The use case completed all tests and is regularly making use of the Federated Cloud using a production level VO
Closed	Test & integration successfully completed. Use case does not foresee moving into production.
Cancelled	Test & integration did not successfully completed. User cancelled the use case.

# Use Case Discipline Classification

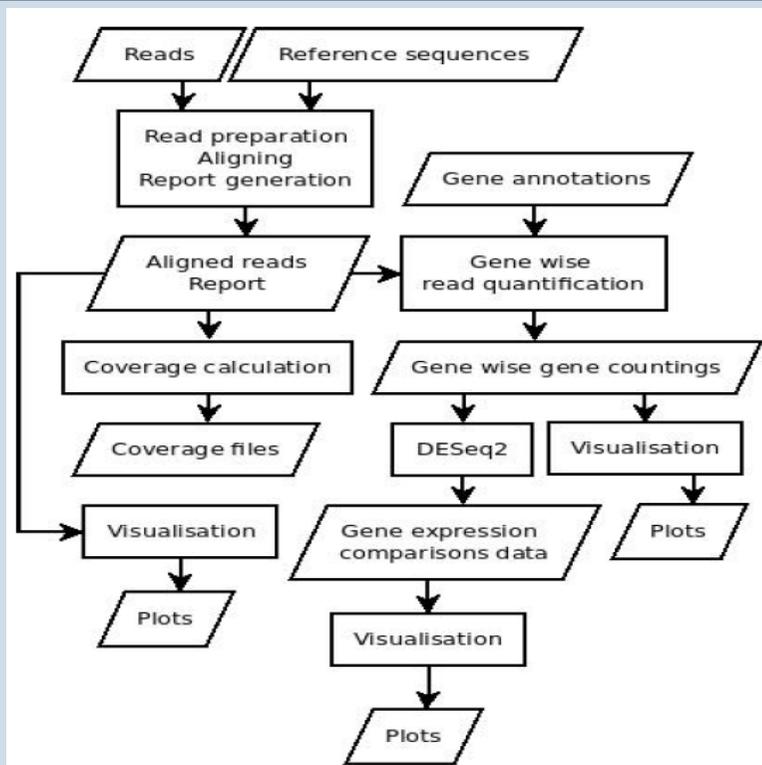


## Usecases

- 12 @ Launch
- 60 to date
  - 15 production



- **Scientific Discipline:** Natural Science, Biological Sciences, **Bioinformatics**
- **Status:** **Production** ([highthroughputseq.egi.eu](http://highthroughputseq.egi.eu) VO)
- **Sites:** GoeGrid, IFCA (Nov-Dec 2014)



Source: Konrad U. Förstner

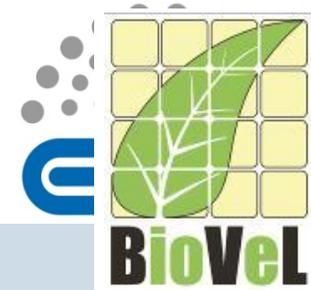
## Next Generation Sequencing

- Implemented in Python 3
- Needs:
  - Third-Party libraries (numpy, scipy, matplotlib, pysam)
  - Short Read mapper ‘Segemehl’
  - R
- Usual runtime on a multi-core machine several hours to some days

## Implementation on EGI Federated Cloud

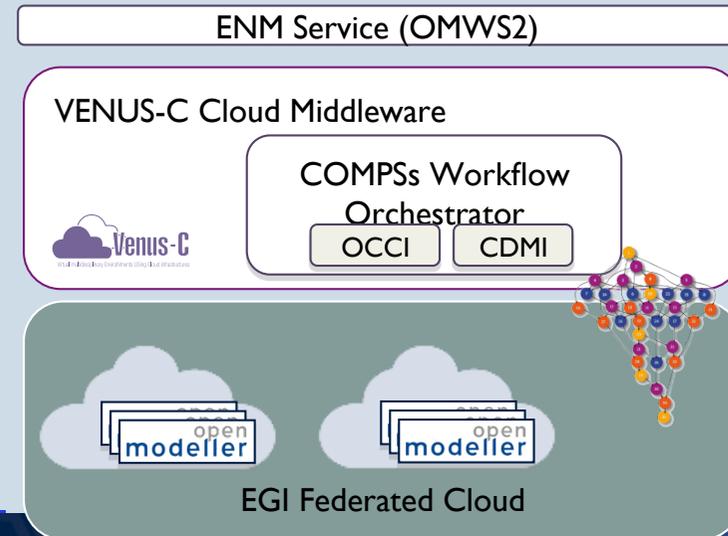
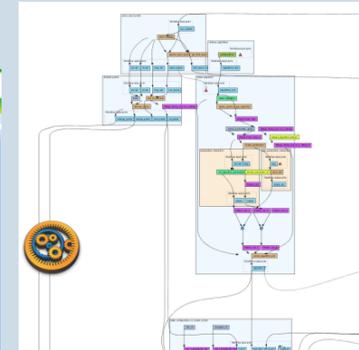
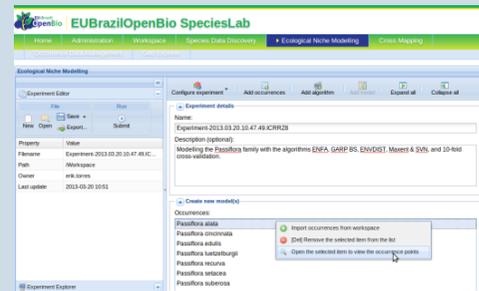
- VO supported by GWDG, IFCA and CESNET sites
- VM with 24 cores and 128 GB of RAM
- Block storage up to 3TB
- To serve peak loads

# OpenModeller on the Biovel Portal



The Ecological Niche Modeling (ENM) Workflow takes as input a file containing species occurrence points to create a model with the [openModeller](#) Web Service.

- The EUBrazilOpenBio ENM service is exposed through an extended openModeller Web Service interface
- Multi-staging and multi-parametric oM experiments are implemented through [COMPSs](#) that dynamically creates the virtual resources to execute the operations.
- An OCCI connector is used for the VMs management while data management supports CDMI endpoints.



Service available at <https://portal.biovel.eu/>

# Peachnote



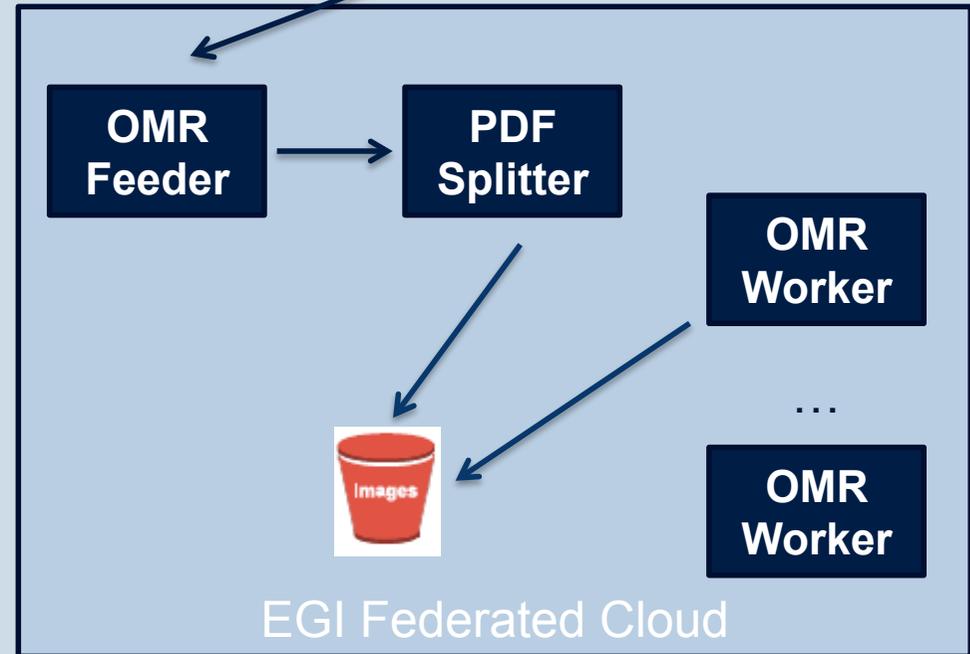
- **Scientific Discipline:** Humanities, Arts, **Musicology**
- **Status:** **Production** (peachnote.com VO)
- **Sites:** CESNET, FZJ (Nov-Dec 2014)

Peachnote is a **music score search engine and analysis platform.**

Hundreds of thousands of music scores are being digitized by libraries all over the world. In contrast to books, they generally remain inaccessible for content-based retrieval and algorithmic analysis.

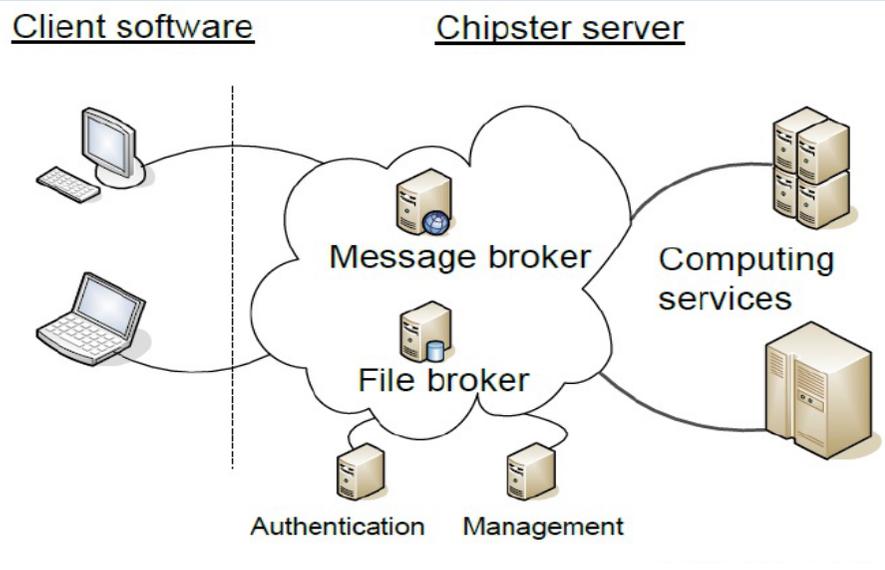
There is no **analogue to Google Books for music scores**, and no large corpora exists that can empower advanced analysis on music scores.

Peachnote **want to help change that** providing visitors and researchers access to a massive amount of symbolic music



- **Scientific Discipline:** Natural Science, Biological Sciences, **Bioinformatics**
- **Status:** **Test & Integration** (fedcloud.egi.eu VO)
- **Sites:** PRISMA-INFN-Bari

## ELIXIR Pilot Action Proposal: Using virtual machines and clouds in bioinformatics training



### User-friendly analysis software for high-throughput data:

- NGS
- Microarray
- Proteomics
- sequence data

### Chipster in the EGI FedCloud:

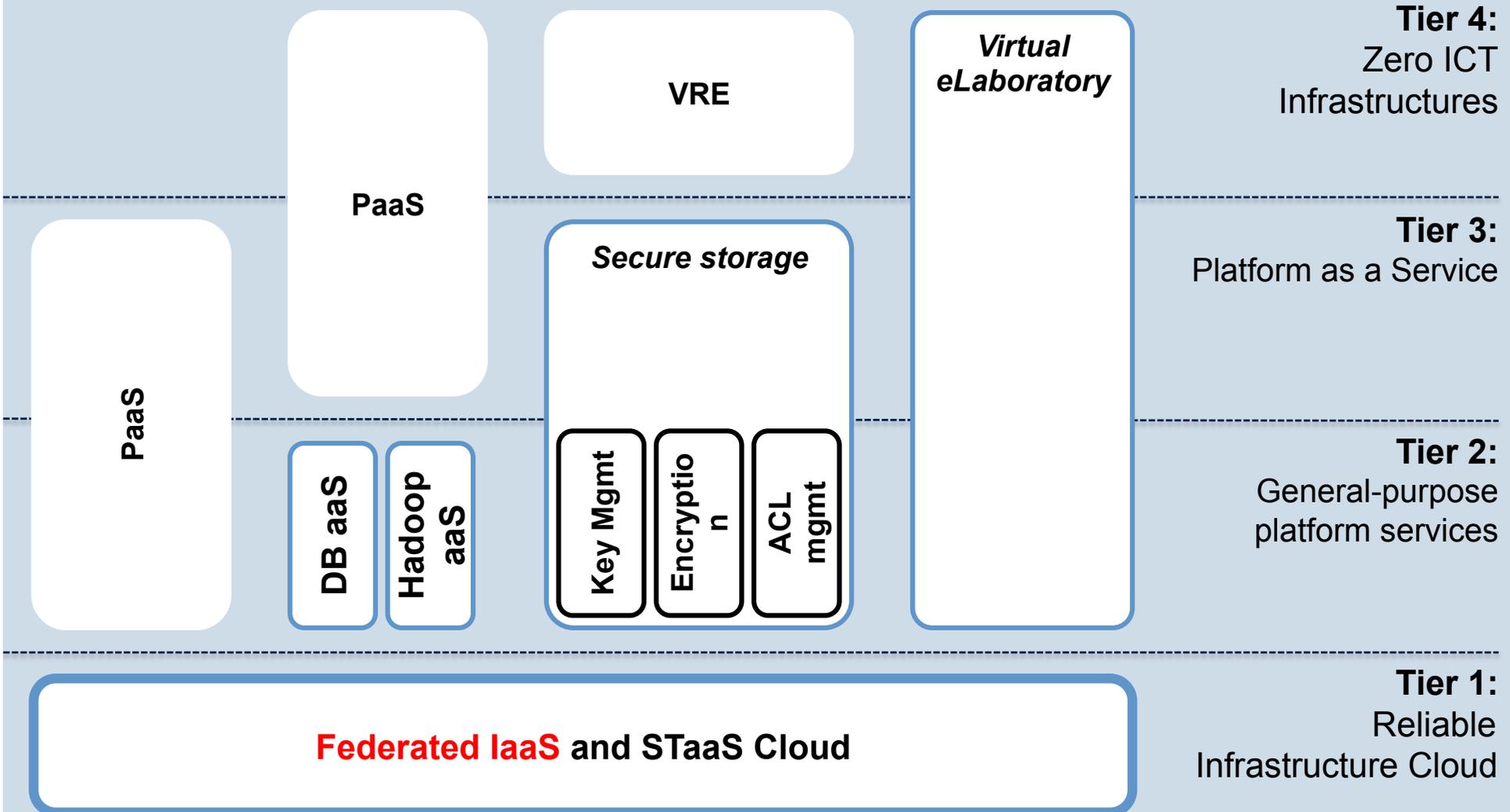
- 'light' VM (datasets removed)
- Chipster VM configured through contextualisation
- shared block storage exported as NFS for tools (500 GB)
- block storage for output (500 GB)

# Future Work



- Inter and Intra cloud networking,
  - Using standards to support creation of custom multi resource private networks for cloud resources
- Development of multi provider and multi consumer business models for federated cloud operation,
- Expansion in higher level tools and services building on concrete foundations.
- Further development and release of updated standards in cloud and other related areas
  - In development progress of complete EGI profile which will fully document all interfaces to the CMF
- Connect resources from other commercial cloud providers to showcase deployment of standards

# Federated Cloud Services



# Conclusion



The EGI Federated Cloud is a federation of institutional private Clouds, offering Cloud Services to researchers in Europe and worldwide

A single cloud system able to

- Scale to user needs – beyond test and development instantiations to production private clouds
- Integrate multiple different providers to give resilience
- Prevent vendor lock-in
- Enable resource provision targeted towards the research community

Thanks!



Alison Packer, Álvaro López García, Binh Minh Nguyen, Björn Hagemeyer, Boris Parak, Boro Jakimovski, Cal Loomis, Carlos Gimeno, Christos Loverdos, Daniele Cesini, Daniele Lezzi, David Blundell, Diego Scardaci, Elisabetta Ronchieri, Emir Imamagic, Enol Fernandez, Esteban Freire, Feyza Eryol, Florian Feldhaus, Gergely Sipos, Ignacio Blanquer, Iván Díaz, Jan Meizner, Jerome Pansanel, John Gordon, Kostas Koumantaros, Linda Cornwall, Luis Alves, Malgorzata Krakowian, Marios Chatziangelou, Marco Verlato, Marica Antonacci, Mattieu Puel, Matteo Turilli, Michel Jouvin, Michel Drescher, Miguel A. Díaz, Miroslav Ruda, Mischa Salle, Nuno L. Ferreira, Owen Synge, Paul Millar, Peter Solagna, Piotr Kasprzak, Roberto Barbera, Ruben Valles, Sándor Ács, Salvatore Pinto, Silvio Spardi, Soonwook Hwang, Steven Newhouse, Stuart Pullinger, Sven Gabriel, Thijs Metsch, Tomasz Szeplieniec, Víctor Mendez, Viet Tran, Vincenzo Spinozo, Zeeshan Ali Shah, Zdenek Sustr

# Thanks & Questions?