



EGI Proposal of Collaboration for the Pancancer Analysis of Whole Genomes

24 July 2015

This document describes the EGI proposal in reply to the request for collaboration to support the Pancancer Analysis of Whole Genomes project submitted by the Steering Committee of the International Cancer Genome Consortium (ICGC).

1 About EGI

EGI is a federation of shared computing, storage and data resources from national and intergovernmental resource providers that delivers sustainable, integrated and secure distributed computing services to European researchers and their international partners. In this context, EGI operates a hybrid cloud federation with 19 cloud providers across 12 European countries offering IaaS cloud and storage services, and more than 750,000 Virtual Machines run since January 2014. The EGI Cloud Federation aggregates resources by defining a set of standard open-source interfaces and protocols to access the different cloud functions - such as resource discovery, user authentication, compute and data access services - in a uniform way at all the sites, enabling workloads to span and seamlessly migrate across resource centers.

EGI.eu is the non-profit organisation based in Amsterdam established to coordinate and manage EGI on behalf of its participants such as National Grid Initiatives (NGIs) and European Intergovernmental Research Organisations (EIROs) including EMBL-EBI.

2 EGI proposal for collaboration

In order to support the Pancancer Analysis of Whole Genomes collaboration, EGI can offer the following total capacity meeting the technical specifications defined in the call:

- 1984 cores
- 280 TB

The offer could further increased with the contribution of additional cloud providers who expressed interest but could not finalize an offer in time because of the summer period. The possibility of an extension of the current will be investigated if of interest to ICGC.

The resource pledges offered provide the following technical capabilities.

Specifications		
Group 1 Provider: SURFSara (The Netherlands)	Number of cores	256
	Maximum VM size that the cloud infrastructure offered can support (memory, number of virtual cores and disk)	32 vCPU, 125GB RAM, 1.5 TB SSD (local, not shared) disk per virtual machine
	Total storage capacity offered [TB]	40TB block storage (Ceph) + 12 TB local disk (SSD) = 52 TB
	Block storage capacity supported (total and maximum per single resource) [TB]	5 TB per virtual machine
	Supported interfaces and protocols	OpenNebula 4.12.1 with web interface for VM management
Group 2 Provider: SWITCH (Switzerland)	Number of cores	1600, Expanding to around 4800 by October 2015
	Maximum VM size that the cloud infrastructure offered can support (memory, number of virtual cores and disk)	16 vCPU, 64 GB Ram, any size disk (this can be increased to 24 vCPUs, 96 GB RAM) but that would make the scheduling inflexible
	Block storage capacity supported (total and maximum per single resource) [TB]	Ceph based block storage. Thinly provisioned up to 100TB per disk (but a total of around 300 TB all in all) - also expanding by October
	Total storage capacity offered [TB]	200 TB
	Supported interfaces and protocols	OpenStack APIs
Group 3 Provider: CIEMAT (Spain)	Number of cores	128
	Maximum VM size that the cloud infrastructure offered can support (memory, number of virtual cores and disk)	First stage 16 vCPUs, 64 GB RAM, 350 GB disk / Second stage 16 cores, 64 GB RAM, 1 TB disk
	Block storage capacity supported (total and maximum per single resource) [TB]	First stage: 2 TB / Second stage: 20TB
	Total storage capacity offered [TB]	28 TB
	Supported interfaces and protocols	OCCI, OpenStack APIs

OpenStack resources will be accessible through the launcher of the pancancer workflow system. OpenNebula cloud resources will be accessible through a local account and via manual scheduling of the containers.

Additional cloud services can be offered if useful for downstream analysis or other tasks of the pancancer computation workflows. These provide the following capabilities:

Additional cloud services: specifications		
Group 4 Provider: MARGI (Rep. of Macedonia)	Number of cores	768
	Maximum VM size that the cloud infrastructure offered can support	24 GB, 24 vCPUs, 0.5 TB

	(memory, number of virtual cores and disk)	
	Total storage capacity offered [TB]	17 TB
	Block storage capacity supported (total and maximum per single resource) [TB]	14 TB, 1 TB
	Supported interfaces and protocols	OpenNebula, OCCl
Group 5 Provider: CYFRONET (Poland)	Number of cores	512
	Maximum VM size that the cloud infrastructure offered can support (memory, number of virtual cores and disk)	12 GB, 8 vCPUs, 10 GB
	Block storage capacity supported (total and maximum per single resource) [TB]	1 TB
	Total storage capacity offered [TB]	20 TB
	Supported interfaces and protocols	OpenStack, OCCl

In addition to the cloud services described above EGI will:

- Provide expertise in application porting to cloud (creation of VMs and application containers)
- Assist with the running of the framework for executing the Pancancer workload in the EGI Cloud Federation by providing effort for the “Cloud Shepherd” role requested by ICGC.
- Offer technical support to the Pancancer collaboration via a the EGI helpdesk: <http://helpdesk.egi.eu/>
- Collaborate for the redistribution of raw and analyzed data produced by the Pancancer project.

3 EGI Requirements

The mission of EGI is to “create and deliver open solutions for science and research infrastructures by federating digital capabilities, resources and expertise between communities and across national boundaries”, as such the support of the PanCancer Analysis of Whole Genomes project is of great interest. For the donation offered, we ask EGI and the providers of the cloud services to be recognized in publications and project communication activities through the following acknowledgement statement:

This work used EGI through resources from from Country_1, Country_2,

EGI is willing to establish a long term Service Level Agreements to support the Pancancer project as well as engage with ICGC through a MoU.

4 Contact

For proceeding further with the proposed collaboration please contact: support@egi.eu

For technical information about the present offer: Enol Fernández enol.fernandez@egi.eu