Helix Nebula – The Science Cloud

Title: Finalised User and Service Requirements Report

Editor: CloudSigma

Work Package: WP3 – Representation of requirements

Submission Date: 19 June 2013

Distribution: Public

Nature: Report

The Helix Nebula project is co-funded by the European Community Seventh Framework Programme (FP7/2007-2013) under Grant Agreement no 312301

Date	Version	Contributor(s)
03/05/2013	0.1	Peter Gray (CloudSigma)
10/05/2013	0.2	Plamen Ganchosov (CloudSigma), Peter Gray (CloudSigma)
28/05/2013	0.3	Plamen Ganchosov (CloudSigma), Peter Gray (CloudSigma), Robert Jenkins (CloudSigma)
06/06/2013	0.4	Plamen Ganchosov (CloudSigma), Peter Gray (CloudSigma)
17/06/2013	1.0	Peter Gray (CloudSigma), Plamen Ganchosov (CloudSigma)
18/06/2013	1.1	Plamen Ganchosov (CloudSigma), Peter Gray (CloudSigma), Robert Jenkins (CloudSigma)
19/06/2013	1.2	Peter Gray (CloudSigma), Rachida Amsaghrou (CERN), Bob Jones (CERN)
20/06/2013	1.3	Peter Gray (CloudSigma), Michel van Adrichem (Atos), Wolfgang Lengert (ESA)
21/06/2013	1.4	Peter Gray (CloudSigma), Michel van Adrichem (Atos), Ramon Medrano (CERN)

The Helix Nebula project is co-funded by the European Community Seventh Framework Programme (FP7/2007-2013) under Grant Agreement no 312301

Table of Contents

The Helix Nebula project is co-funded by the European Community Seventh Framework Programme (FP7/2007-2013) under Grant Agreement no 312301

1. Executive Summary

This report will focus on the activities and outcomes of work package 3 with regard to the gathering of infrastructure and service requirements. We aim to document the learning process and summarise our findings and improvements to the process. The objective was to use the knowledge and experience of both the participating research organisations and commercial cloud service providers to develop a requirements framework that would address the technical and non-technical expectations of the demand-side and the capabilities of the supply-side. In Section 2.0 we focus on the development of the requirements framework, from the requirements gathering template prototype to the design and implementation of the online requirements form and back-end repository. We also outline improvements made to the online form and back-end repository and show how the graphical data can be used to understand at-a-glance the collected data as a whole and how this will enable trend spotting. Section 3.0 explains how the requirements framework has continued to serve its purpose with new flagship candidates, using the example of PICNICC, the first of the new flagship candidates to complete the online requirements form. Section 4 presents the requirements gathered from the respective flagships; EMBL - Genomic Assembly in the Cloud, ESA – Super Sites Exploitation Platform, CERN – ATLAS High Energy Physics and PIC - The Port d'Informació Científica.

The Helix Nebula project is co-funded by the European Community Seventh Framework Programme (FP7/2007-2013) under Grant Agreement no 312301

2. The requirements framework

2.1. The development of a requirements framework

A requirements definition template was prototyped and implemented during the course of 2011. The template underwent subsequent refinement using the existing knowledge base of the original three flagships, CERN ATLAS High Energy Physics, EMBL Genomic Assembly in the Cloud and ESA Super Sites Exploitation Platform. The initial purpose of the requirements framework was to understand the technologies and resources needed to fulfill the requirements of each individual use case and to use these to drive the first Proof of Concept phase. Three Proof of Concept (PoC) environments were created for the flagships and deployed by a number of supply-side participants. Figure 1 shows a section from the original definitions template.

1.1.4.3 Types

Please select the correct server image type based on the technology you use today. Again, please select all applicable technologies used.

Type	Check as appropriate
VirtualBox	☐
KVM	☐
XEN	☐
VMware	☐
ESX	☐
Hyper-V	☐
Others (please specify)	

1.1.4.4 Access methods (upload, download)

Please select the access methods for importing and extracting server images for the flagship to the Helix Nebula cloud framework that you would wish to be made available.

Method	Check as appropriate
FTP (including SFTP)	☐
API via Https	☐
P2P	☐

Figure 1. Requirements Definition Template

The Helix Nebula project is co-funded by the European Community Seventh Framework Programme (FP7/2007-2013) under Grant Agreement no 312301

2.2. The online form and back-end repository

Figure 2 shows the online requirements tool which has been created to allow the demand side to input requirements more easily and efficiently. It also improves the way the supply-side is able to analyse the requirements of each flagship individually and as a group. The requirements of the demand side are then able to be matched accordingly with the appropriate technologies and resources provided by a combination of suppliers. Furthermore, the process of determining a large set of technical and non-technical requirements per flagship has contributed to the development of the Blue Box, bringing us one step closer to a fully operational federated cloud. Based on this information, decisions were made as to which potential Blue Box solution can fulfill the demands set out by the consortium. The entire requirements gathering process has provided a solid foundation for building the Blue Box solutions to bring additional focus on the service aspects of Helix Nebula.

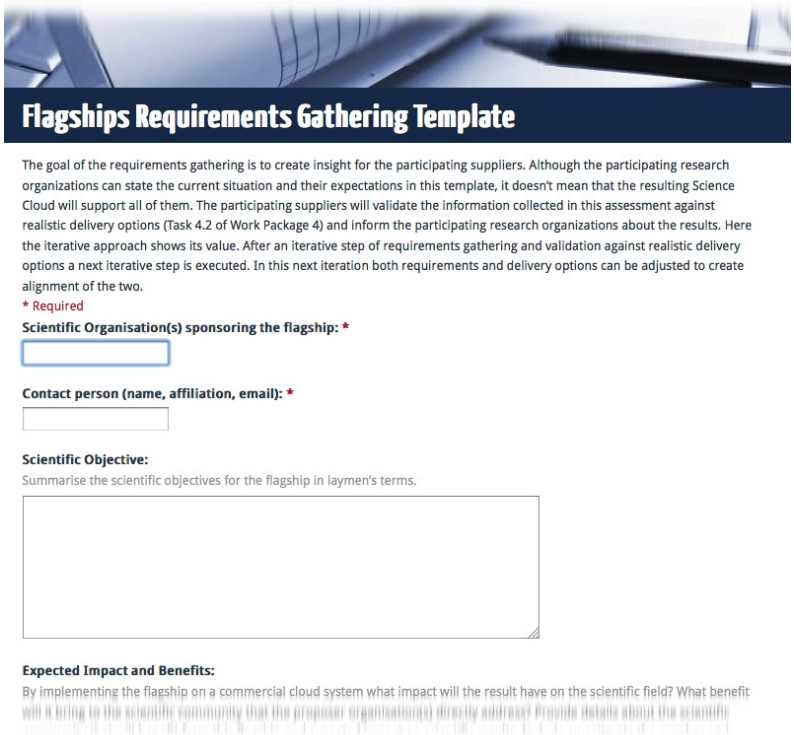

Figure 2. Online Requirements Form

Since the online form was established a number of small amendments have been necessary to meet the needs of the use cases as they develop and reflect the evolving capabilities of the supply side. The online form has been designed to capture and store all submitted information into a single spreadsheet enabling a consolidated view of each flagship which is then made available to the relevant partners from the Helix Nebula consortium. We are also able to extract data in the form of graphical charts which enables the requirements to be viewed at a glance, giving a broad picture of the overall requirements. This helps with the identification of trends developing across the demand side. Figure 3 is a cross-section showing the back-end repository and corresponding graphical data representation.

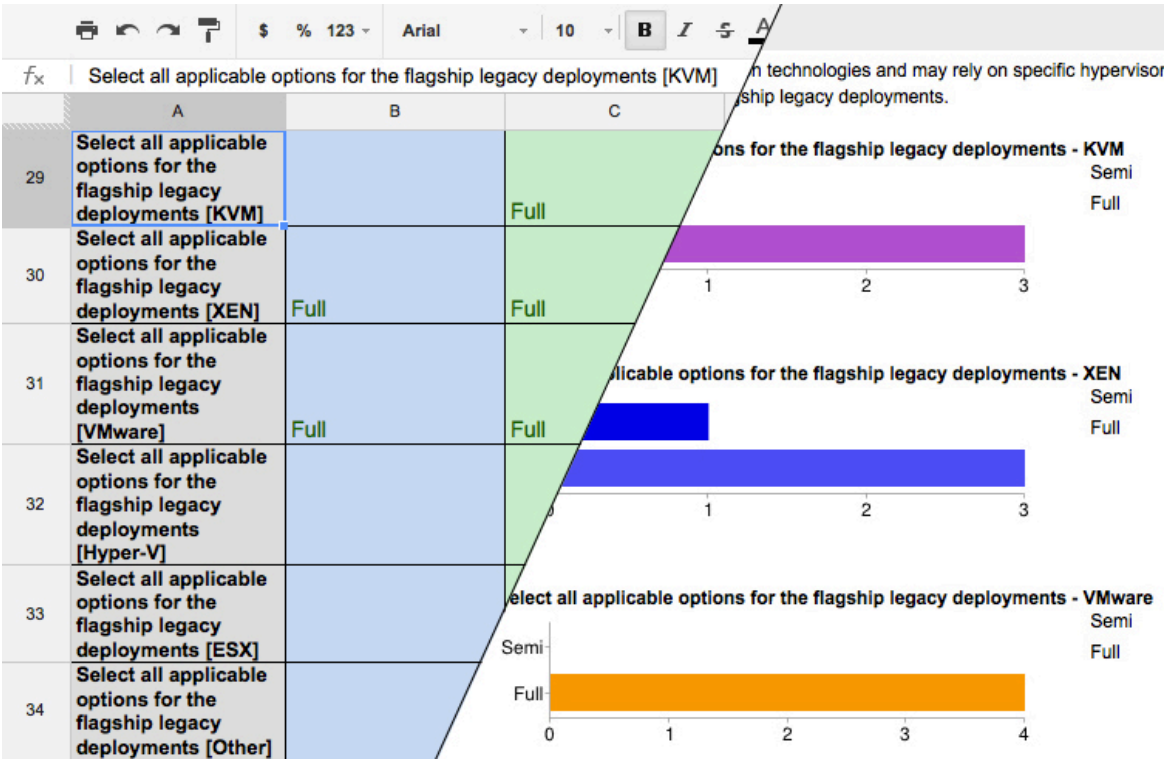

Figure 3. Cross section showing both the back-end repository and graphical representations

2.3. Modifications and improvements

The online requirements tool has been designed in such a way that additional criteria can be included and redundant criteria removed easily and quickly in order to reflect the

current status of the project as other work packages develop. However, small modifications have been made periodically to the requirement framework as anticipated.

The online form has been proven to work very effectively in collecting the requirements of the original three flagships. We have attempted to improve functionality to the online form and back-end repository where necessary.

Some improvements could not be implemented. The absence of a save-icon still prevents multiple users within an organisation from completing the form in stages and saving subsequent sessions. Despite a number of alternative form creation packages having been tested, a reliable and cost effective software solution to this problem has not yet been found. A single-session submit-icon at the end of the current form remains the best option at this point of time. However, this does have one advantage: candidates are forced to think more thoroughly about every aspect of their proposed use case requirements and plan their submission in advance. It is still possible to add extra information at a later stage to the central repository. Any additional information will automatically be updated in the central repository accordingly.

The complete current online requirements form is available at:
https://sites.google.com/a/cloudsigma.com/helix-nebula/

The Helix Nebula project is co-funded by the European Community Seventh Framework Programme (FP7/2007-2013) under Grant Agreement no 312301

2.4. Graphical data

Using the online requirements tool, we are able to extract data from the spreadsheet and automatically provide a summary of each parameter. Some of this information is made available in the form of graphical representations. Below are some examples of consolidated user requirements data represented in a series of graphs.

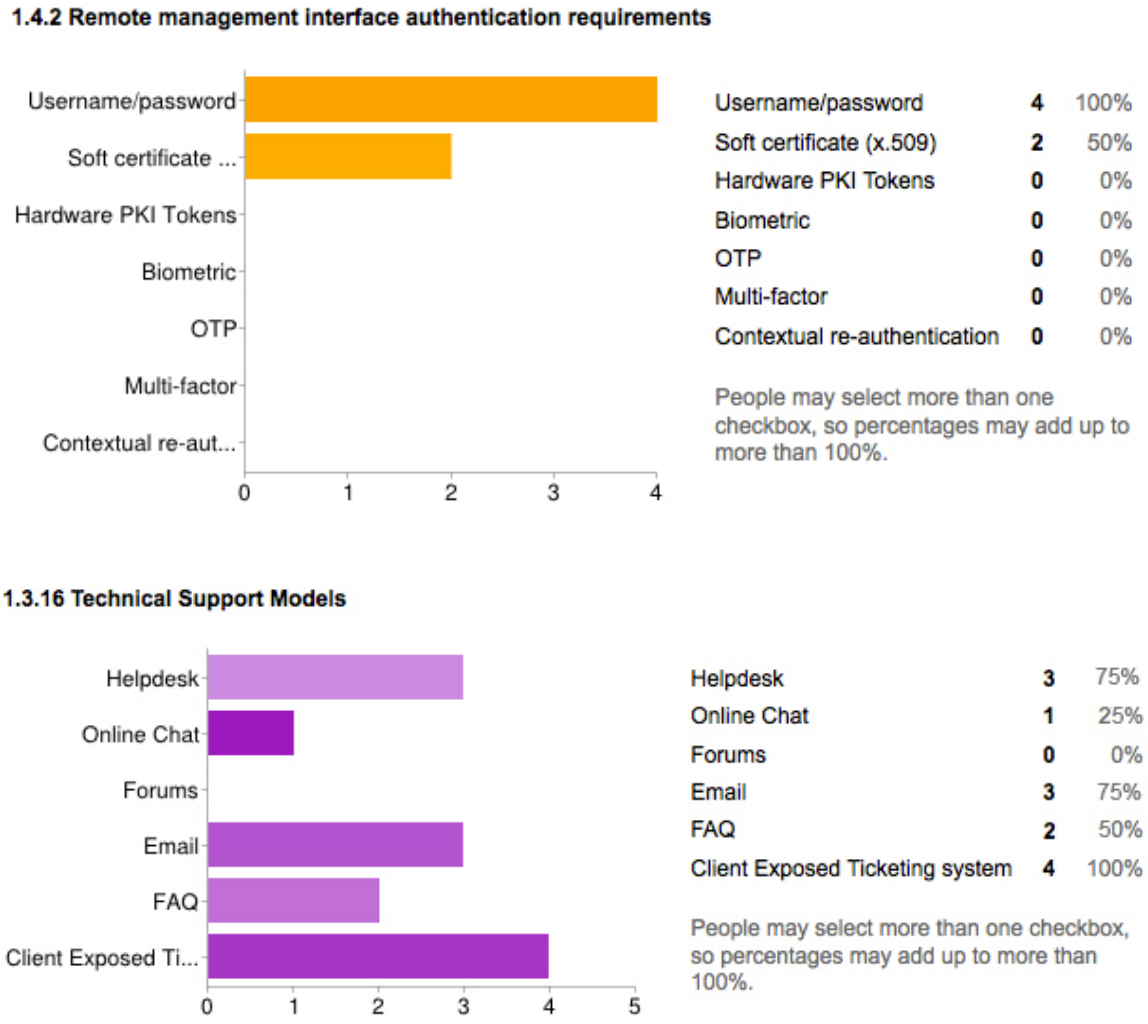

Figure 4. Examples of graphical data representations

The Helix Nebula project is co-funded by the European Community Seventh Framework Programme (FP7/2007-2013) under Grant Agreement no 312301

We can easily see from 1.3.1.6 Technical Support Models in Figure 4, which forms of support are preferred. A Client Exposed Ticketing System is required by all the participating flagships, while Helpdesk and Email support are also considered important services. We can also see that at this point in time there is no demand for Forums to play a part in the technical support model.

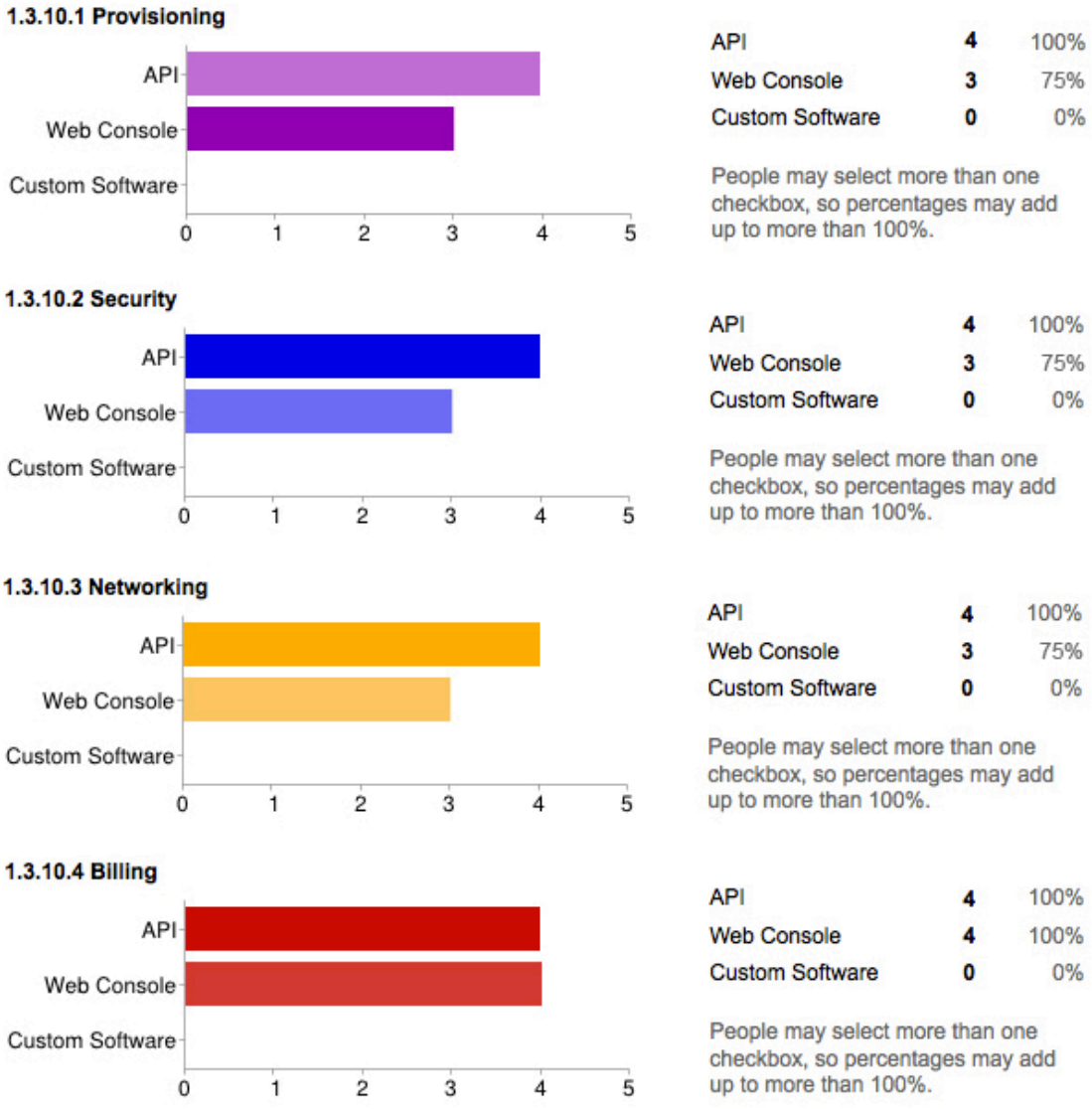

Figure 5. Example of graphical data representations

The Helix Nebula project is co-funded by the European Community Seventh Framework Programme (FP7/2007-2013) under Grant Agreement no 312301

An advantage of graphical representations is that they can provide a "quick view" whereby we can begin to see certain patterns emerging. In Figure 5, 1.3.1.0 Preferred Cloud-Interface, we are able to see that all flagships to date have selected API as the main method of interaction with the cloud. Furthermore, this has been selected across the board for each of the four management aspects; provisioning, security, networking and billing. Some flagships have also selected the option of a web interface, while there have been no specific requests for a custom software interface.

Based on this information, decisions can be made about what types of interfaces the supply-side will continue to make available as part of the overall management model. It is important to note that each individual input is time-coded and user-specific, which allows us to identify trends and monitor requirements or preferences over time. This will become more valuable as more flagships are invited to join Helix Nebula.

3. Prospective flagships

The online form and repository is available to all prospective flagships and has been completed most recently by PICNICC (PIC) Neuroimaging Center on Cloud. The process has been outlined in the following subsection.

3.1. The Port d'Informació Científica (PIC)

PICNICC (PIC) recently made an application for a new use case which outlines their science objectives as well as their predicted technical specifications. This was presented by PIC to the consortium during the 2nd Helix Nebula General Assembly in January, 2013. The active IaaS suppliers provided feedback based on the proposed use case and PIC was given a chance to respond. The table below summarises the comments and questions posed by the supply-side.

T-Systems	T-Systems currently do not support GPU-assisted nodes.

The Helix Nebula project is co-funded by the European Community Seventh Framework Programme (FP7/2007-2013) under Grant Agreement no 312301

	Use of Matlab SW on cloud provider servers need to be clarified (if the current customer license is enabling this). For the long term storage requirement of up to 350 TB, a federated data management solution should be envisaged to optimize the costs of operations.
Atos	Adherence to TechArch and ServArch descriptions needs to be confirmed. A Data Management scheme is required to be established, and a confirmation that the customer will be responsible for the relevant payments. A commitment to initial business volumes, and a (possibly non-committal) capacity planning process to track them, needs to be established. We need to include provision for GPU's in our product mix.
Interoute	How is this to be funded and to what extent? No GPU resources available Data; • If stored externally - where is it - how do we connect? Is it PIC via GEANT? • If bought into the cloud - how much of it or all of it and how is this orchestrated? • Is the data hosted on the FAS2040? Matlab licensing would obviously need investigation
CloudSigma	This is a good example of an existing application which can be improved via a cloud delivery mechanism. Commercial prospects look good and initial data requirements are not onerous. We are also interested in the prospect to test GPUs for their applications also.

A conference call was held between representatives from both PIC and CloudSigma on behalf of the supply-side to address the specific supplier feedback. A common understanding was reached and PIC was invited to submit their requirements using the online form.

The Helix Nebula project is co-funded by the European Community Seventh Framework Programme (FP7/2007-2013) under Grant Agreement no 312301

As PIC was not involved in the initial development of the requirements framework, a subsequent conference call was arranged by CloudSigma to cover points on the questionnaire that were not fully understood. PIC was then able to submit their requirements to the online form and the back-end repository was automatically updated. Consequently, the PIC requirements were automatically consolidated with the results from the existing flagships.

4. Requirements

4.1. EMBL – Genomic Assembly in the Cloud

Contact person (name, affiliation, email):	Rupert Lück (rupert.lueck@embl.de), Christian Boulin (Christian.boulin@embl.de), Vladimir Benes (benes@embl.de), Jonathon Blake (blake@embl.de), Tobias Rausch (rausch@embl.de), Jürgen Zimmermann (zimmerma@embl.de) EMBL-EBI: Paul Flicek (flicek@ebi.ac.uk)
Scientific Objective:	The Next Generation Sequencing technologies, such as the Illumina Sequencing platform have had a huge impact on how we now perform research in biology, massively increasing the amount of sequence data that can be produced. Previously it took 10 years for the Human Genome Project and a consortium of large sequencing facilities to produce sufficient sequence data for the human genome to be assembled. Now, with massively parallel sequencing technologies a human genome can be sequenced in 14 days on a single machine. This opens up many possibilities for studying biodiversity. It no longer requires a massive consortium to sequence a large eukaryotic genome. From bacterial to insect, research model organism to agricultural species, even human genome sequencing can now be done quickly in a few lanes of Illumina sequencing. This has resulted in many large projects with the aim of sequencing large numbers of genomes such as the Genome 10K project (www.genome10k.org) and the 1000 genomes project (www.1000genomes.org) With the capability for even small academic groups to sequence their organism of choice, the focus changes from sequence production to data analysis and assembly of the organism's genome. This is technically difficult due to the type of data produced, 10's of millions of short sequence reads around 100 bases long, and the data sizes generated, 300 GB – 1TB per genome depending on number of sequencing runs used. In cases where a reference genome already exists, a reference based assembly can be done. In the case of the majority of species where no reference genome exists one has to be assembled de novo requiring high performance computer infrastructure.

The Helix Nebula project is co-funded by the European Community Seventh Framework Programme (FP7/2007-2013) under Grant Agreement no 312301

	The option also exists to produce de novo assemblies of already sequenced organisms (e.g. human), or to combine de novo assembly and reference based assembly techniques providing the potential to compare assemblies. The infrastructure required to perform de novo assembly of a large eukaryotic genome has been a hurdle to scientific groups wishing to perform these studies. Once the assembly is completed, an annotation of the genome to identify the locations of the protein coding genes and other features will enable detailed study of both species-specific biology and leverage the power of genomic analysis across thousands of species. We believe that the availability of a de novo assembly and annotation cloud service will open up new possibilities for basic research by removing the computational resource hurdles involved in performing de novo genomic assembly. Furthermore we believe that by making it possible for research groups to create genome assemblies and their associated annotation in the cloud we are supporting innovative research on a wide basis, thus providing a richer picture of the Earth's genomic diversity. Based on our experience we see a clear potential for this pilot to become a success. Once live this service should be very attractive to users from a wide range of NGS domains, hence, serving as a basis for future extension e.g. to clinical applications or research in the agricultural sector.
Expected Impact and Benefits:	EMBL is situated in the centre of Europe and serves the European scientific community both inside EMBL and beyond (figure 2). By implementing de novo assembly and annotation in the cloud we will be providing a large service to basic research. The service should make it possible to researchers to obtain genomic assemblies of their model organism without smaller laboratories having to make large capital investments in computing infrastructure. With over 200 Illumina HiSeq2000 sequencers in Europe there is the potential to produce enough sequence data for 1600 large eukaryotic genomes per month. Taking only 1% of this devoted to de novo genome assembly we are looking at 16 x 5000 = 80,000 CPU hours' worth of genomic assembly data processing per month in Europe. As mentioned above, large projects exist where many thousands of organisms are planned to be sequenced and as these projects develop we expect many more species to join the sequencing list. During the pilot phase we expect that up to 500 genomes per year may be assembled with this service. As additional data is generated the requirements for annotation and re-annotation with the new data sources will increase. This step is less memory intensive, but more CPU intensive and is expected to be run multiple times for each assembled genome.
Existing or potential partnership:	ATOS, CloudSigma, T-Systems, SixSQ
Proposer Motivation:	Part of EMBL's mission is to provide services to the scientific community in Europe. We are clearly motivated to doing so and in the pilot phase our major

The Helix Nebula project is co-funded by the European Community Seventh Framework Programme (FP7/2007-2013) under Grant Agreement no 312301

	focus is on the scalability of resources required for this challenging project.
Proposer Long-term Objectives:	Our long term objective is to make de novo assembly and genome annotation widely available to the scientific community maximize the scientific and social benefits we believe this service can bring. We also propose to expand into comparative assemblies of human genomes given that the confidentiality and legal issues of dealing with potential clinical data can be satisfied. We believe the provision of this service to human genome re-sequencing has great potential for benefiting the medical community. The cloud computing model provides the power and scalability to allow us to make this service available to the widest possible user base.
1.1.1 Please define the connectivity method expected to be utilised. Please select all methods that are compatible with a successful deployment.	Internet
1.1.4.1 Please check the formats of legacy server and/or drive images that will be provided for the proof of concept environment.	QCOW2, VMDK
1.1.4.2 Please select whether server images being provided are private or publicly available images.	Private
1.1.4.3 Please select the correct server image type based on the technology you use today.	XEN, VMware, Amazon
1.1.4.4 Please select the access methods for importing and extracting server images for the flagship to the Helix Nebula	SCP

The Helix Nebula project is co-funded by the European Community Seventh Framework Programme (FP7/2007-2013) under Grant Agreement no 312301

cloud framework that you would wish to be made available.	
1.1 Resource Sizing	**Please outline different server node types required for the proposed flagship proof of concept deployment. Be especially sure to include server configuration types that represent the minimum and maximum requirements even if their aggregate number is small. Additionally, providing a % guide of the overall importance of each server type is also useful. Fill in as many tables as you deem appropriate with regards to covering the majority of your expected flagship deployment.**
1.1.5.1 Example Server Type A [CPU/Core]	
1.1.5.1 Example Server Type A [RAM]	
1.1.5.1 Example Server Type A [HDD]	
1.1.5.1 Example Server Type A [SSD]	
1.1.5.1 Example Server Type A [NETWORK]	
CPU (Overall, Max.)	2.4
RAM	64
HDD	4TB
SDD	N/A
Network	4TB
CPU	2.4
RAM	32
SDD	N/A
HDD	4TB

The Helix Nebula project is co-funded by the European Community Seventh Framework Programme (FP7/2007-2013) under Grant Agreement no 312301

Network	4TB
1.1.6 Virtualisation	**If you currently use virtualisation technologies and may rely on specific hypervisors please outline your usage currently. Select all applicable options for the flagship legacy deployments.**
[KVM]	
[XEN]	Full
[VMware]	Full
[Hyper-V]	
[ESX]	
[Other]	N/A
1.2 Relevant Software Systems	
Please outline any relevant software systems in use for the flagship today and what systems might be required going forward as part of the initial flagship proof of concept.	We need to use MIT Starcluster (star.mit.edu/cluster) as the provisioning software. It allows ● To dynamically launch a number of instances that form a SUN / Oracle Grid Engine (workload scheduler) based cluster ● To attach block storage to VMs to form a high performance GusterFS based shared file system across the storage volumes that are attached to individual VMs. Starcluster currently requires the Amazon EC2 API. All other (assembly and annotation) software we use will be most likely open source tools.
1.2.1 Please outline the current situation, possible restrictions and any requirements on the portability of applications used by the flagship.	The portability is constrained by the availability of the EC2 API to satisfy the Starcluster requirements. There is, however, the possibility to extend Starcluster to work with different APIs.
1.2.2 Please outline the current practices, considerations and requirements on the test and go live strategy	N/A

The Helix Nebula project is co-funded by the European Community Seventh Framework Programme (FP7/2007-2013) under Grant Agreement no 312301

used by the flagship	
1.3 Operational	**Please outline the capacity requirements and nature with regards to the flagship**
1.3.1.1 Minimum; most likely; maximum number of servers in a request	Server type A: 1;50;100 / Server type B: 1;1;1
1.3.1.2 Minimum; most likely; maximum amount of storage in a request	Min: 2.0TB; Mean: 2.4 TB; Max: 3.0 TB of total block type storage available to GlusterFS; equal portions of the total storage needs to be individually attached to 4-6 GlusterFS VMs (i.e. 2.4 TB raw storage capacity = 4 x GlusterFS nodes; each GlusterFS node has 600 GB (e.g. 4x150 GB) block storage attached; 2.4 TB raw = 1.2 net GlusterFS capacity). On top of GlusterFS which is per each genomic analysis we need about 500 GB of persistent block storage
1.3.1.3 Minimum; most likely; maximum external network bandwith in a request	100 Mbps; 1 Gbps; 10 Gbps
1.3.1.4 Minimum; most likely; maximum lead time to provide the requested cloud service	seconds; 30 sec; 2 min
1.3.1.5 Minimum; most likely; maximum service provision to be supported	seconds; 5 min; 120 min
1.3.1.6 Minimum; most likely; maximum service provision you are willing to commit to	seconds; 30 min; 60 min
1.3.1.7 Minimum; most likely; maximum period of extension of service provisoning	30 min; 24 h ; 24 h
1.3.1.8 Minimum; most likely; maximum period for service provision	5 days; 14 days; 6 weeks

1.3.1.9 Minimum; most likely; maximum period for termination of service provision	minutes
1.3.1.10 Maximum allowed notification period for termination service provision	1 minute
1.3.2.1 Specify the types of users	Customer users (use flagship resource); Technical users who are authorized to order new services and who facilitate the provisioning of resources; administrators
1.3.2.2 Which party should own the information about the users	EMBL
1.3.2.3 Which party should maintain the information about the user	EMBL ideally would be able to use an existing Helix Nebula identity management system that allows to administer the user base for the flagship
1.3.2.4 How should information about the users be exchanged (e.g. API, Web Console, Custom software)	(Web service) API; Web Console
1.3.3.1 How should the service catalogues from the multiple suppliers be presented to the users?	There is a single integrated service catalogue that presents the services of all suppliers; EMBL selects which service catalogue information we present to our users
1.3.4 Provisioning	**Please outline in the sub-sections below the resource provisioning models desired or required as part of the flagship deployment. Key areas to include are if automatic provisioning via API is required, aspects of any API functionality that may be required, expecting scaling requirements and models, required resource lead times, etc.**
1.3.4.1 Please outline the resource usage profile for deploying	The EMBL flagship requires automatic provisioning via Starcluster and EC2 API

resources. Please indicate the variance of resource requirements over time, the length resources are expected to run for and the preferred method of usage.	EMBL Flagship high level architecture and workflow: Genomic assembly and annotation: ========== The EMBL genomic assembly and annotation consists of the following use case: 1. Customer uploads data for a single genome which needs to be processed to persistent Cloud storage which will be created on-demand a. Storage will be built using GlusterFS file system which is a fast shared file system that is accessible to Cloud based HPC cluster. GlusterFS will aggregate ~4 VM instances and the attached storage (~2.5 TB of total raw disk space) to create a shared file system of ~1.2 TB net capacity) b. Customer uploads genomic sequencing data (several 100 GBs; few hours to 3 days) 2. Create HPC cluster of ~50 nodes on-demand (in the order of minutes) 3. Run the sequence of computational steps to process the genomic input data: (min: 5 days; mean: 10 days; max: 6 weeks) a. QA step to verify the quality of the input genomic sequencing data (from 1.b) b. Assembly stage i. Data Input: Output from 3.a ii. Multi-step sequential process iii. Each step has many jobs which should be distributed to a batch compute farm / cluster; these jobs are independent from each other iv. Output results to internal Cloud storage c. Annotation stage i. Data input: Output from 3.b ii. Multi-step sequential process iii. Each step has many jobs (some have 100.000s of jobs) which should be distributed to a batch compute farm / cluster; these jobs are independent from each other iv. Output results to internal cloud storage 4. Download of (several 10 GBs) result data (from 3.c) by customer (hours) 5. Deprovision Cluster and Shared file system (minutes)
1.3.4.2 Please outline acceptable lead times for new long term capacity requirements	Long term capacity: 12h; short term capacity extensions: real time

The Helix Nebula project is co-funded by the European Community Seventh Framework Programme (FP7/2007-2013) under Grant Agreement no 312301

as well as short term ad hoc variable requirements as appropriate	
1.3.4.3 Please outline any requirements regarding availability and quality of service aspects of any provisioning tools you intend to use (such as API, web console etc.)	The EMBL flagship requires automatic provisioning via Starcluster and EC2 API. Starcluster monitors the dynamic load of the provisioned SUN Grid Engine cluster. Based on the load it can dynamically (burst) provision more or less resources (cluster node VMs) to cope with the cluster load.
1.3.4.5 Please outline any requirements regarding the delivery of content during the provisioning like the method of adding the content to the provided resource and any content delivery tools you intend to use (such as API, web console etc.)	Apsera fsap transfer protocol; FDT transfer protocol / tool
1.3.4.6 Please outline any requirements regarding the deployment of software during the provisioning like the method of adding / including the software to the image and any software deployment tools you intend to use (such as API, web console etc.)	N/A
1.3.5 External Management / Compatibility	N/A
1.3.6 Monitoring	**Please outline any tools and requirements for monitoring of your system. If you have specific tools used already please outline them. Please also**

	specify the aspects of computing being monitored, metrics, etc.
1.3.6.1 System Monitoring	Nagios; Nimsoft
1.3.6.2 Performance Monitoring	Ganglia; Nimsoft
1.3.6.3 Network Monitoring	Nagios; Nimsoft
1.3.6.4 Website Monitoring	Nagios; Nimsoft
1.3.6.5 Security Monitoring	Logging; Alarms
1.3.7 Service Level Reporting	**Please outline any requirements for service level reporting of the services.**
1.3.7.1 How should the service levels from the multiple suppliers be presented to the users?	There is a single integrated service level reporting system that presents the services levels of all suppliers for the delivered services: EMBL assembles the service level reports we present to our users from information from the suppliers
1.3.7.2 Who should have access to the service level reports?	Technical users; administrators
1.3.7.3 How do you want to access to the service level reports? (e.g. API, Web Console, Custom software)	(Web Service) API
1.3.7.4 Other requirements for service level reporting	N/A
1.3.8.1 How should the contract information of the services from the multiple suppliers be presented?	There is a single integrated contract information system that presents the contract information of all suppliers for the delivered services

1.3.8.2 Who should have access to the contract information of the service?	Specific persons / roles within EMBL
1.3.8.3 How do you want to access to the contract information (e.g. API, Web Console, Custom software)	API, Web Console
1.3.8.4 Other requirements for service level reporting	N/A currently
1.3.9.1 Specify the tooling (mention supplier, software name and version) used to collect assets / software used by the flagships?	N/A currently
1.3.9.2 Specify the tooling (mention supplier, software name and version) used to administrate assets / software used by the flagships?	N/A currently
1.3.9.3 Who should collect assets / software used by the flagships running on, but not part of the cloud services?	N/A currently
1.3.9.4 How do you want to access to the asset / software license information?	N/A currently
1.3.9.5 Other requirements for service level reporting	N/A currently

1.3.10 Preferred Cloud-Infrastructure	Please outline any expected infrastructure management models for the various management aspects outlined below.
1.3.10.1 Provisioning	API, Web Console
1.3.10.2 Security	API, Web Console
1.3.10.3 Networking	API, Web Console
1.3.10.4 Billing	API, Web Console
1.3.11 Please outline the operating systems you intend to utilise as part of the flagship deployment including specific versions.	Centos, Ubuntu
1.3.11.1 Other OS not listed above	None
Please outline your requirements with regards to software licensing. If you are expecting to host licensed commercial software as part of your flagship deployment please fill in this section.	N/A currently
Do you anticipate cloud infrastructure providers to have access to running cloud virtual machines? If so please specify acceptable access levels	It would be beneficial if the provider would have full access during the implementation and testing of the flagship. Later during the production the provider ideally should have no access by default. In case of problems we would/could still provide access via SSH
Please outline the remote access methods you intend to utilise to manage flagship cloud infrastructure	VNC, SSH

1.3.15.1 Please outline uptime requirements in relation to availability of cloud infrastructure.	MTBF: >6-12 months RTO: <1 business day RPO: <1 day
1.3.15.2 If any minimum performance guarantees and service level agreements may be required please outline them in as much detail as possible here.	The latency between the storage and the compute nodes for a provisioned single analysis cluster needs to be of the order of a local data center network.
1.3.15.3 Please outline any aspects relating to reliability here. If you have requirements for data resilience/durability, recovery and/or retention, data integrity; this is the correct section to outline such requirements.	Data integrity / durability during 4 weeks for a local GlusterFS filesystem (across 4-6 nodes) need to be maintained.
1.3.15.4 Please outline any aspects relating to accessibility here. Accessibility differs from availability in the sense that accessibility reflects functionality (systems) needed to be able to access the Cloud service but is not managed by the supplier.	MTBF: >4-8 months RTO: <1 business day
1.3.15.5 Please outline any aspects relating to disaster recovery here. Disaster recovery differs from availability in the sense that	MTBF: >6-12 months RTO: <1 business day RPO: <1 day

disasters affect entire sites and therefore affect multiple customers and maybe even other suppliers.	
1.3.16 Please outline any support models you require for the flagship deployment.	Helpdesk, Email, FAQ, Client Exposed Ticketing system
1.3.17 Please outline any requirements and considerations to integrate the local IT support organization with the suppliers support organization for the flagship deployment.	A multi-tiered and integrated support organization will be desirable
1.3.18 Please outline any relevant aspects regarding enterprise application operations that are required for successful deployment.	N/A
1.3.19 Please outline any relevant aspects regarding enterprise application operations that are required for successful deployment.	Data that will be uploaded by EMBL or other customers will traverse a firewall and will be copies to the GlusterFS file system in a private cluster network. The compute nodes of the cluster will be located on the private network only. All access of the compute cluster nodes to the data will occur through the private network. After the analysis the result data set will again traverse the firewall to be downloaded by the customer.
1.4 Security	**Please use this section to define any security requirements not mentioned elsewhere in the document; using the following subheadings as a guide (not all headings are required).**
1.4.1 Authentication, Authorisation and Accountability (AAA)	Identity enrollment requirements – how identity is initially verified, Single sign-on requirements for management interface, RBAC management requirements – e.g. for remote management interface. Accountability requirements – security-related logging, signed time stamping, WORM functionality.

The Helix Nebula project is co-funded by the European Community Seventh Framework Programme (FP7/2007-2013) under Grant Agreement no 312301

1.4.1.1 Notes	None
1.4.2 Remote management interface authentication requirements	Username/password, Soft certificate (x.509)
1.4.2 Remote management interface authentication requirements	Username/password, Soft certificate (x.509)
1.4.3.1 LAN security (for internal transfers) – e.g. IPSec	Initially no specific additional transport security requirements; at later stages IP Sec could be required
1.4.3.2 Network segregation	By firewall on the master VM to bridge the public network from the private cluster and storage network
1.4.3.2 Network access control (client health check)	N/A
1.4.4.1 Secure de-provisioning/ deletion requirements	All data need to be physically deleted after a certain run of a client's genomic analysis
1.4.4.2 Hardware decommissioning requirements (degauss etc…)	Initially none; at a later stage physical destruction of disk drives might be necessary
1.4.4.3 Specific data export/portability requirements (formats, time limits)	N/A
1.4.5.1 Minimum SSH key length policy for remote access	1024
1.4.5.2 Key management	Integrated key management
1.4.5.3 At-rest	Currently none; eventually at a later time point

The Helix Nebula project is co-funded by the European Community Seventh Framework Programme (FP7/2007-2013) under Grant Agreement no 312301

encryption (e.g. encryption gateway)	
1.4.5.4 Crypto hardware/acceleration	Currently none; eventually at a later time point
1.4.5.5 Entropy/randomness sources.	Currently none; eventually at a later time point
1.4.6.1 Incident response services and service levels	Will not be managed by EMBL
1.4.6.2 Incident/vulnerability severity classification used, if any	Will not be managed by EMBL
1.4.6.3 Incident reporting (to/by demand side)	Will not be managed by EMBL
1.4.6.4 Vulnerability reporting and management (to/by demand side)	Will not be managed by EMBL
1.4.6.5 Testing requirements (e.g. external pen-testing)	Will not be managed by EMBL
1.4.6.6 Third party security services used, interfaces required.	Will not be managed by EMBL
1.4.7.1 Certifications required	None currently; at a later stage the analysis of patient data might require more
1.4.7.2 Right to audit	Yes at later stages of the flagship
1.4.7.3 Any other procedural security policy requirements that would have to be	Might be applicable at a later time point

complied with (e.g. around personnel clearance, subcontracting, jurisdiction).	
1.4.8.1 Location/jurisdiction-limitations	Inside EU only
1.4.8.2 Third parties/subcontractor	Initially not allowed
1.4.8.3 Breach notification	If a breach happens EMBL need to be notified
1.4.8.4 Maximum, minimum data retention	10 years
1.4.8.4 Access and rectification	Should be possible to administrators and service user
1.4.8.5 Purpose limitation	N/A
1.5 Networking	**Please outline in more detail the expected networking requirements for the flagship. Include internal and external connectivity, availability requirements, private and public networking needs and an overall expected deployment topology.**
1.5.0.1 Estimate Necessary Capacity (peak bandwidth, 95th percentile bandwidth)	10 GBit inside the flagship infrastructure
1.5.0.2 Topology	We need a service such as Aspera's FASP protocol that allows customers who want to upload large data files to maximize the bandwidth
1.5.0.3 Number of Nodes	20-100
1.5.0.4 Multiple Interfaces	One interface for private cluster network; one for public access
1.5.0.5 Fail Over Plan	None

1.5.1 Private Networking	Jumbo Frame
1.5.2 Public Networking	Jumbo Frame, Multiple IP addresses per Interface, Host Level Firewall, Reverse DNS Management
1.5.3 DNS	Hostname should be addressable; custom hostname resolution; reverse DNS
1.6 Storage	**Please outline overall storage requirements for the flagship with respect to total capacity, availability and performance metrics**
1.6.0.1 Capacity	see 1.6.1.1
1.6.0.2 Availability	see 1.6.1.1
1.6.0.3 Performance	R/W from/to Glusterfs at 7+ Gbps
1.6.0.4 Usage Profile	mixture of long reads and a high number of short reads / writes
1.6.1.1 Block Device Storage	As outlined above, for every new EMBL genomic assembly and annotation (steps 1-5 above in overall architecture description) a new GlusterFS based fast shared file system will be build. GlusterFS will aggregate ~4 VM instances and the attached storage (~2.5 TB of total block based storage) to create a shared file system of ~1.2 TB net capacity) In addition to the GlusterFS needs and on top of what is specified below, we will need a permanent data store of 500 GB block based storage, which is persistent across the lifetime of the whole project to allow storing bulk data that continuously needs to be accessible in the Cloud.
1.6.1.2 Volume Storage	If high performance NAS storage (e.g. NetApp OnTAP) is available we would replace glusterFS with 1.2 TB NAS. The 500 GB of persistent storage could also live on NAS.
1.6.1.3 Object Storage	N/A
1.6.2.1 Geographical data storage requirements	All data that a particular customer uploads for an individual genome analysis (e.g. 0.8TB Next Generation Sequencing Data) needs to be in the same data centre as the compute capacity (e.g. 50-100 compute VMs) and connected via 10Gbit (optimal) or 1Gbit (feasible). However, different individual genome analyses can operate in different data centres after initial setup. The initial setup involves the installation of a persistent EMBL VM image (EMBL-provider transfer of 10 GB), a one-time population and transfer of a local Mysql DB (EMBL-provider transfer of 200 GB).

1.6.2.2 Large Volume Data Transfer	Large data transfer will occur when customers (such as EMBL or other labs) will upload Next Generation Sequencing data (e.g. 0.5-1.0 TB) for analysis. After an analysis customers will download the results (e.g. 10-100 GB). During a single analysis data will not be moved to other clouds.
1.7 Billing	
Please outline any preferred or required billing and purchasing models, for example 'burst' based resource billing versus 'subscription' based longer term billing.	For the initial setup phase a subscription model should be feasible; during the launch phase and production phase we would prefer a burst based resource billing.
If you have organisational requirements regarding reporting, invoicing, account use-age limits, agreed supplier lists, etc., please outline these in detail.	We need monthly reporting of the services used. Ideally we would have a live view with estimates for the future resource use based on previous / current usage.
1.8 Legal Documents	
1.8.1 Please use this section to outline any SLA requirements not already articulated in other sections of this agreement already.	None
1.8.2 Please specify any particular requirements or preferences with regards to this document here.	None
1.8.3 Outline here any specific needs with regards to privacy policy here.	It will be required that any data which will be sent to and from the Helix Nebula cloud infrastructure as well as all data that results from processing on Helix Nebula will be owned and can only be used by EMBL or the respective customer of our flagship's service.

1.8.4 If you have any aspects of usage or requirements that may be appropriate against such a policy please outline them here.	It will be required that any data which will be sent to and from the Helix Nebula cloud infrastructure as well as all data that results from processing on Helix Nebula will be owned and can only be used by EMBL or the respective customer of our flagship's service.
1.8.5 Any requirements regarding the treatment of your intellectual property please outline it here.	It will be required that any data which will be sent to and from the Helix Nebula cloud infrastructure as well as all data that results from processing on Helix Nebula will be owned and can only be used by EMBL or the respective customer of our flagship's service.
1.8.6 Any requirements regarding vendor neutrality please outline it here.	Vendor neutrality should be supported on functional / technical (e.g. APIs) and contractual levels
Other	N/A
Other	N/A
Other	We have rather limited experiences with the first three frameworks. However, our main requirements regarding a cloud management system would include • That a future framework is backed by a reasonably sized development and support organization. OpenNebula seems to lack both. • The capability to integrate the Amazon EC2 API for portability reasons • Capabilities to support the major commercial and open source cloud systems
Other	N/A

4.2. ESA – Super Site Exploitation Platform

Contact person (name, affiliation, email):	Lengert Wolfgang (ESA, Wolfgang.Lengert@esa.int); Jordi Farres (ESA, Jordi.Farres@esa.int); Giancarlo.Rivolta (Logica, Giancarlo.Rivolta@esa.int); Salvatore Pinto (Logica, Salvatore.Pinto@esa.int)
Scientific Objective:	The ESA flagship is working along GEO's disaster theme of "reducing the loss of life and property from natural disasters" The objective is: Improve disaster risk management and reduction by providing timely information relevant to the full cycle of disaster management (mitigation

	preparedness, warning, response and recovery). Adopt a multi-hazard end-to-end approach to ensure that relevant Earth observations and information effectively reach decision-makes and public. The Super Site Exploitation Platform (SSEP) shall become the R&D environment to fulfil these tasks.
Expected Impact and Benefits:	SSEP will allow easy access to all relevant data and tools (space & in-situ) for the specific geolocations which are at risk of earthquakes and Volcanoes. The impact / benefit would be the support of the 2 priority actions mentioned below and using HN to create an ecosystem, providing sustainability.
Existing or potential partnership:	The current partners are CNES, DLR, CNR and the potential partners are the committee on Earth Observation Satellite (CEOS: http://www.ceos.org), geophysical research organizations, IT and Earth Observation (EO) SMEs making, policy makers, rescue teams, EO value adding and commercial companies, ...)
Proposer Motivation:	The SuperSites Exploitation Platform shall improve the understanding of Geophysical processes creating earthquakes and volcano eruptions. SSEP supporting the GEO hazard supersites have visibility to policy makers since It supports risk assessment and risk mitigation. It is likely that combination of large scale satellite and in-situ data being easily accessible on a Cloud infrastructure with large scale computing capacity will attract business to build on the research findings. ESA expects that the availability of the "Science Cloud" will catalyse the exploitation of the Earth Observation data while significantly enlarge the science community leading to a drastic increase of publications and new ideas on the usage of data and information derived on SSEP. Beside the advantage for science it is expected that also business cases and policy interest might arise from the availability of a unique central pool of (i) large scale Space & in-situ data, (ii) focal point of the global geo-hazard science community, and (iii) large scale computing resources. A secondary objective is that the usage of a mature commercial service will reduce the IT cost for Earth Observation while providing a reliable data dissemination and processing services with SLAs and transparent and detailed reporting on the usage (e.g. going beyong Amazon accounting and billing) SSEP is only 1 of many possible "Exploitation Platforms" being supported by HN.
Proposer Long-term Objectives:	HN to create a federating European industry infrastructure with data providers and scientists in a federated open approach. This will allow to establish the concept of "Exploitation Platforms" where ESA stakeholders (Space and in-situ data providers, scientist, facilities, national/EC funding) meet to create an open science ecosystem.
1.1.1 Please define the	Internet

The Helix Nebula project is co-funded by the European Community Seventh Framework Programme (FP7/2007-2013) under Grant Agreement no 312301

connectivity method expected to be utilised. Please select all methods that are compatible with a successful deployment.	
1.1.4.1 Please check the formats of legacy server and/or drive images that will be provided for the proof of concept environment.	VMDK
1.1.4.2 Please select whether server images being provided are private or publicly available images.	Private
1.1.4.3 Please select the correct server image type based on the technology you use today.	VMWare, Amazon
1.1.4.4 Please select the access methods for importing and extracting server images for the flagship to the Helix Nebula cloud framework that you would wish to be made available.	HTTP, FTP, SCP
1.1 Resource Sizing	**Please outline different server node types required for the proposed flagship proof of concept deployment. Be especially sure to include server configuration types that represent the minimum and maximum requirements even if their aggregate number is small. Additionally, providing a % guide of the overall importance of each server type is also useful. Fill in as many tables as you deem appropriate with regards to covering the majority of your expected flagship deployment.**
1.1.5.1 Example Server	4

The Helix Nebula project is co-funded by the European Community Seventh Framework Programme (FP7/2007-2013) under Grant Agreement no 312301

Type A [CPU/Core]	
1.1.5.1 Example Server Type A [RAM]	10GB
1.1.5.1 Example Server Type A [HDD]	100GB
1.1.5.1 Example Server Type A [SSD]	N/A
1.1.5.1 Example Server Type A [NETWORK]	1GB
CPU	
RAM	
HDD	
SDD	
Network	
CPU	
RAM	
SDD	
HDD	
Network	
1.1.6 Virtualisation	**If you currently use virtualisation technologies and may rely on specific hypervisors please outline your usage currently. Select all applicable options for the flagship legacy deployments.**
[KVM]	
[XEN]	
[VMware]	Full

[Hyper-V]	
[ESX]	Full
[Other]	
1.2 Relevant Software Systems	
Please outline any relevant software systems in use for the flagship today and what systems might be required going forward as part of the initial flagship proof of concept.	Hadoop FS and NFS are required as shared file system. Contextualization mechanism in the API is also required. Possibility to dynamically plug/unplug custom disks is nice to have.
1.2.1 Please outline the current situation, possible restrictions and any requirements on the portability of applications used by the flagship.	EC2 and OCCI API suppost would been nice to have for portability
1.2.2 Please outline the current practices, considerations and requirements on the test and go live strategy used by the flagship	N/A
1.3 Operational	**Please outline the capacity requirements and nature with regards to the flagship**
1.3.1.1 Minimum; most likely; maximum number of servers in a request	Server type A: Min 10, Likey , 50, Maximum 100
1.3.1.2 Minimum; most likely; maximum	Min 30TB, Likely 200TB, Max 300TB

amount of storage in a request	
1.3.1.3 Minimum; most likely; maximum external network bandwith in a request	Min: 1GB, Likey 3GB, Max 10GB
1.3.1.4 Minimum; most likely; maximum lead time to provide the requested cloud service	Min 5 seconds; Likey 10 seconds; Max 1 minute per VM (and/or VM modification operation)
1.3.1.5 Minimum; most likely; maximum service provision to be supported	Min 5 seconds; Likey 10 seconds; Max 1 minute per VM (and/or VM modification operation)
1.3.1.6 Minimum; most likely; maximum service provision you are willing to commit to	minutes: 1 day: 1 month
1.3.1.7 Minimum; most likely; maximum period of extension of service provisoning	minutes: 1 day: 1 month
1.3.1.8 Minimum; most likely; maximum period for service provision	minutes: 1 day: 1 month
1.3.1.9 Minimum; most likely; maximum period for termination of service provision	minutes
1.3.1.10 Maximum allowed notification period for termination service provision	minutes
1.3.2.1 Specify the types of users	Customer users (use flagship resource); Technical users (infrastructure management access but no commercial visibility) who are authorized to order new services and who facilitate the provisioning of resources;

The Helix Nebula project is co-funded by the European Community Seventh Framework Programme (FP7/2007-2013) under Grant Agreement no 312301

	administrators (full IaaS access)
1.3.2.2 Which party should own the information about the users	ESA
1.3.2.3 Which party should maintain the information about the user	ESA
1.3.2.4 How should information about the users be exchanged (e.g. API, Web Console, Custom software)	No user data should be required
1.3.3.1 How should the service catalogues from the multiple suppliers be presented to the users?	Service catalogue are limited to ESA, it is not shared to the users. ESA acts as service reseller (cloud services are hidden from end users)
1.3.3.2 How do you want to access to the service catalogues	API, Web Service
1.3.4 Provisioning	**Please outline in the sub-sections below the resource provisioning models desired or required as part of the flagship deployment. Key areas to include are if automatic provisioning via API is required, aspects of any API functionality that may be required, expecting scaling requirements and models, required resource lead times, etc.**
1.3.4.1 Please outline the resource usage profile for deploying resources. Please indicate the variance of resource requirements over time, the length resources are expected to run for and the preferred method of usage.	Case 1: Data dissemination: 2VM and 20TB of storage always on Case 2: Processing campaigns: from 10 to 50VM on for a period from 15days to 3months Case 3: Sandbox service: 2VMs on during working hours (ad hoc machines for people to use)

The Helix Nebula project is co-funded by the European Community Seventh Framework Programme (FP7/2007-2013) under Grant Agreement no 312301

1.3.4.2 Please outline acceptable lead times for new long term capacity requirements as well as short term ad hoc variable requirements as appropriate	N/A
1.3.4.3 Please outline any requirements regarding availability and quality of service aspects of any provisioning tools you intend to use (such as API, web console etc.)	We expect to use the service to accommodate peaks in the processing. The scaling frequency is quite low, varying the number of VMs in matter of days (in relation to medium term processing requirements)
1.3.4.5 Please outline any requirements regarding the delivery of content during the provisioning like the method of adding the content to the provided resource and any content delivery tools you intend to use (such as API, web console etc.)	Output of the processing will be shared with the users via HTTP(s)
1.3.4.6 Please outline any requirements regarding the deployment of software during the provisioning like the method of adding / including the software to the image and any software deployment tools you intend to use (such as API, web console etc.)	Software deployment is done by the system attaching external disks or transfering the software via HTTP/FTP/SCP/GridFTP. Core OS drive with additional data set/application specific drives
1.3.5 External Management /	N/A

The Helix Nebula project is co-funded by the European Community Seventh Framework Programme (FP7/2007-2013) under Grant Agreement no 312301

Compatibility	
1.3.6 Monitoring	**Please outline any tools and requirements for monitoring of your system. If you have specific tools used already please outline them. Please also specify the aspects of computing being monitored, metrics, etc.**
1.3.6.1 System Monitoring	Nagios,OpenNebula from ESA;Internal monitoring from the Cloud provider
1.3.6.2 Performance Monitoring	Nagios,OpenNebula from ESA;Internal monitoring from the Cloud provider
1.3.6.3 Network Monitoring	Nagios,OpenNebula from ESA;Internal monitoring from the Cloud provider
1.3.6.4 Website Monitoring	Nagios from ESA
1.3.6.5 Security Monitoring	Internal monitoring from the Cloud provider
1.3.7 Service Level Reporting	**Please outline any requirements for service level reporting of the services.**
1.3.7.1 How should the service levels from the multiple suppliers be presented to the users?	SLA should be presented only to ESA
1.3.7.2 Who should have access to the service level reports?	ESA (technical users and administrators)
1.3.7.3 How do you want to access to the service level reports? (e.g. API, Web Console, Custom software)	Web Service, API
1.3.7.4 Other requirements for service level reporting	None
1.3.8.1 How should the contract information of	No preference

the services from the multiple suppliers be presented?	
1.3.8.2 Who should have access to the contract information of the service?	No preference
1.3.8.3 How do you want to access to the contract information (e.g. API, Web Console, Custom software)	No preference
1.3.8.4 Other requirements for service level reporting	No preference
1.3.9.1 Specify the tooling (mention supplier, software name and version) used to collect assets / software used by the flagships?	No preference
1.3.9.2 Specify the tooling (mention supplier, software name and version) used to administrate assets / software used by the flagships?	No preference
1.3.9.3 Who should collect assets / software used by the flagships running on, but not part of the cloud services?	No preference
1.3.9.4 How do you want to access to the asset / software license information?	No preference

The Helix Nebula project is co-funded by the European Community Seventh Framework Programme (FP7/2007-2013) under Grant Agreement no 312301

1.3.9.5 Other requirements for service level reporting	N/A
1.3.10 Preferred Cloud-Infrastructure	**Please outline any expected infrastructure management models for the various management aspects outlined below.**
1.3.10.1 Provisioning	API: OCCI, Amazon EC2 preferred; Web Console. Need contextualisation mechanism for VMs.
1.3.10.2 Security	API, Web Console potentially to manage firewalls and similar.
1.3.10.3 Networking	API: OCCI, Amazon EC2 preferred; Web Console
1.3.10.4 Billing	API, Web Console
1.3.11 Please outline the operating systems you intend to utilise as part of the flagship deployment including specific versions.	Centos 5/6, Scientific Linux 5/6
1.3.11.1 Other OS not listed above	None
Please outline your requirements with regards to software licensing. If you are expecting to host licensed commercial software as part of your flagship deployment please fill in this section.	None
Do you anticipate cloud infrastructure providers to have access to running cloud virtual machines? If so please specify acceptable access levels	No

The Helix Nebula project is co-funded by the European Community Seventh Framework Programme (FP7/2007-2013) under Grant Agreement no 312301

Please outline the remote access methods you intend to utilise to manage flagship cloud infrastructure	SSH
1.3.15.1 Availability	99%
1.3.15.2 Performance	None
1.3.15.3 Reliability/Durability	None
1.3.15.4 Accessibility	None
1.3.15.5 Disaster recovery	None
1.3.16 Technical Support Models	Helpdesk, Email, FAQ, Client Exposed Ticketing system. Coverage: normal working hours
1.3.17 Technical Support Models	None
1.3.18 Enterprise Application Operation	None
1.3.19 Data and resource management	Data that will be uplaoded by ESA or other customers via HTTP, FTP, SCP protocols. Data management will be done by ESA. Access to the data shall be restricted to ESA only (cloud provider cannot access it). Disks management (attach disk, shared disks between VMs, etc...) shall be done via API or Web Console
1.4 Security	**Please use this section to define any security requirements not mentioned elsewhere in the document; using the following subheadings as a guide (not all headings are required).**
1.4.1 Authentication, Authorisation and Accountability (AAA)	Helix-Nebula Single-sign-on or ESA EO SSO single sign on support is preferred (Shibboleth)
1.4.1.1 Notes	None
1.4.2 Remote management interface	Username/password (preferred), X509 certificate

authentication requirements	
1.4.3.1 LAN security (for internal transfers) – e.g. IPSec	No encryption needed if networks are isolated (see req 1.4.3.2)
1.4.3.2 Network segregation	Minimum one isolated network for all the VM. It is preferred to have the possibility to specify multiple isolated networks to assign dynamically to the VMs
1.4.3.3 Network access control (client health check)	No
1.4.4.1 Secure de-provisioning/ deletion requirements	Wipe data after de-provisioning is required
1.4.4.2 Hardware decommissioning requirements (degauss etc…)	Degauss required after resource de-provisioning
1.4.4.3 Specific data export/portability requirements (formats, time limits)	No
1.4.5.1 Minimum SSH key length policy for remote access	No
1.4.5.2 Key management	Any
1.4.5.3 At-rest encryption (e.g. encryption gateway)	No
1.4.5.4 Crypto hardware/acceleration	No
1.4.5.5 Entropy/randomness	Linux kernel basic one is enough

sources.	
1.4.6.1 Incident response services and service levels	SLA: Blocking: 1 hour (Work around), 3 day solving High: 1 day WA, 7 days solving Medium: 3 day WA, 14 days solving Low: On agreement
1.4.6.2 Incident/vulnerability severity classification used, if any	Blocking (No connection to the machines, hardware issues) High (ex VMs isolated from the interne) Medium (ex. Slow access to the disk) Low (ex. provisioning not working)
1.4.6.3 Incident reporting (to/by demand side)	Any
1.4.6.4 Vulnerability reporting and management (to/by demand side)	Any
1.4.6.5 Testing requirements (e.g. external pen-testing)	An external penetration test on-demand would been nice to have feature
1.4.6.6 Third party security services used, interfaces required.	None
1.4.7.1 Certifications required	None required, ESA CA will be used
1.4.7.2 Right to audit	No
1.4.7.3 Any other procedural security policy requirements that would have to be complied with (e.g. around personnel clearance, subcontracting, jurisdiction).	None
1.4.8.1	Hosting in member countries only.

Location/jurisdiction-limitations	
1.4.8.2 Third parties/subcontractor	No
1.4.8.3 Breach notification	Immediate notification to the ESA contacts
1.4.8.4 Maximum, minimum data retention	None
1.4.8.4 Access and rectification	None
1.4.8.5 Purpose limitation	None
1.5 Networking	**Please outline in more detail the expected networking requirements for the flagship. Include internal and external connectivity, availability requirements, private and public networking needs and an overall expected deployment topology.**
1.5.0.1 Estimate Necessary Capacity (peak bandwidth, 95th percentile bandwidth)	10 GBit inside the flagship infrastructure
1.5.0.2 Topology	One single network for each VM is the minimum. Possibility to create multiple isolated networks is nice to have. In that case, the networks will be about 5
1.5.0.3 Number of Nodes	10-100
1.5.0.4 Multiple Interfaces	One interface for private cluster network; one for public access
1.5.0.5 Fail Over Plan	None
1.5.1 Private Networking	Jumbo Frame
1.5.2 Public Networking	1Bit bandwidth Host Firewall is nice to have

The Helix Nebula project is co-funded by the European Community Seventh Framework Programme (FP7/2007-2013) under Grant Agreement no 312301

1.5.3 DNS	DNS for internet addresses or possibility to access external public DNS servers. Reverse DNS required. Hostname resolution nice to have.
1.6 Storage	**Please outline overall storage requirements for the flagship with respect to total capacity, availability and performance metrics**
1.6.0.1 Capacity	None
1.6.0.2 Availability	99.99% SLA
1.6.0.3 Performance	None
1.6.0.4 Usage Profile	Shared file system for internal network HTTP file download, FTP file upload for external (internet) network
1.6.1.1 Block Device Storage	Storage will be enhanced at blocks of 2.5TB
1.6.1.2 Volume Storage	N/A
1.6.1.3 Object Storage	N/A
1.6.2.1 Geographical data storage requirements	It is required for data to be in located in member country sites.
1.6.2.2 Large Volume Data Transfer	Large data transfer will occur only at start-up, to fill the system with the ESA database of EO data (around 20TB). After that time, data transfer will be limited to smaller amounts. Incremental updates <1TB month.
1.7 Billing	
Please outline any preferred or required billing and purchasing models, for example 'burst' based resource billing versus 'subscription' based longer term billing.	None
If you have organisational requirements regarding reporting, invoicing,	None

account use-age limits, agreed supplier lists, etc., please outline these in detail.	
1.8 Legal Documents	
1.8.1 Please use this section to outline any SLA requirements not already articulated in other sections of this agreement already.	None
1.8.2 Please specify any particular requirements or preferences with regards to this document here.	None
1.8.3 Outline here any specific needs with regards to privacy policy here.	It will be required that any data which will be sent to and from the Helix Nebula cloud infrastructure as well as all data that results from processing on Helix Nebula will be owned and can only be used by ESA or the respective customer of our flagship's service.
1.8.4 If you have any aspects of usage or requirements that may be appropriate against such a policy please outline them here.	It will be required that any data which will be sent to and from the Helix Nebula cloud infrastructure as well as all data that results from processing on Helix Nebula will be owned and can only be used by ESA or the respective customer of our flagship's service.
1.8.5 Any requirements regarding the treatment of your intellectual property please outline it here.	It will be required that any data which will be sent to and from the Helix Nebula cloud infrastructure as well as all data that results from processing on Helix Nebula will be owned and can only be used by ESA or the respective customer of our flagship's service.
1.8.6 Any requirements regarding vendor neutrality please outline it here.	None
Other	Need to define and implement non acceptable end user behaviour with automatic flagging/account revocation. Consider using subaccounts for

	technical users.
Other	-
Other	-
Other	-

4.3. CERN – ATLAS High Energy Physics

Contact person (name, affiliation, email):	Ian Bird (CERN, Ian.Bird@cern.ch)
Scientific Objective:	The LHC experiments at CERN are running a large scale distributed computing system on the WLCG grid infrastructure to perform processing and analysis of the particle collision data. The distributed computing environment consists of several pieces: a distributed data management component, a workflow manager, and the associated tools as well as the processing and analysis codes themselves. The objectives of this flagship use case are: ● To evaluate the available cloud technologies in relation to the use-cases of data management, processing, and analysis ● To design a model for transparently integrating cloud computing resources with the WLCG software and services ● To implement the cloud computing model into DDM, Panda, and related tools and services. This flagship is part of an encompassing ATLAS project to research cloud computing and its applicability to LHC computing. Specific additional objectives for more general use of cloud services by the LHC community are: ● Determine the costs of commercial cloud resources resulting from network transfers of data into and out of the cloud resource, short and long-term data storage in the cloud, and CPU resources for running the various experiment use cases. ● Develop an understanding of appropriate SLA's and how they might be defined for broader use. ● Understand the policy and legal constraints in moving scientific data across academic networks into commercial resources and back again.
Expected Impact and Benefits:	-
Existing or potential	-

The Helix Nebula project is co-funded by the European Community Seventh Framework Programme (FP7/2007-2013) under Grant Agreement no 312301

partnership:	
Proposer Motivation:	Most important factors to understand are: ● Cost of data transfer in and out and storage; overall cost ● Performance and reliability – in comparison to the WLCG as a baseline. ● Use of standard interfaces and interoperability between providers ● Can CERN transparently offload work to a cloud resource? ● Is there a requirement for a mechanism to make allocations between user groups in a cloud resource?
Proposer Long-term Objectives:	In the short term the benefits of running the LHC software in a commercial cloud system will be the possibility to dynamically acquire additional resources when needed. In the longer term use of commercial resources could become a real alternative to very large data centres owned and managed by the scientific community. The LHC scientific community of ~5000 physicists will directly benefit from this work. However not all of them will require access to the cloud services. In the organised data processing use cases the production managers only will need access (~10 people?), although many more may require accounts if the analysis use cases are successful. However, the work in all cases will be managed through the common Panda workflow service. The users may be based at any ATLAS or CMS institution around the world. Analysis use case workloads would originate from many different people, although they would still all run within the same framework. It is likely that a commercial cloud will not care that the workload comes from a different user – they will only see the framework and its owner. Probably the real point is that the direct impact to physicists is that they use these resources – so many more people get a direct benefit when compared to the data processing tasks. Also analysis may be a case where (depending on costs) data could be cached at a cloud site for re-use by different analyses. Eventual expansion of the use of cloud services by the entire LHC community could impact some 10000 physicists. The long-term objective would be to use commercial services as a significant fraction of the overall computing resources available to the experiments. At what point (if at all) does it become feasible economically and practically to rely on commercial or 3rd party providers?
1.1.1 Please define the connectivity method expected to be utilised. Please select all methods that are compatible with a successful deployment.	Internet, Secure Remote User Access

The Helix Nebula project is co-funded by the European Community Seventh Framework Programme (FP7/2007-2013) under Grant Agreement no 312301

1.1.4.1 Please check the formats of legacy server and/or drive images that will be provided for the proof of concept environment.	QCOW2, RAW
1.1.4.2 Please select whether server images being provided are private or publicly available images.	Public
1.1.4.3 Please select the correct server image type based on the technology you use today.	KVM
1.1.4.4 Please select the access methods for importing and extracting server images for the flagship to the Helix Nebula cloud framework that you would wish to be made available.	API via Https
1.1 Resource Sizing	**Please outline different server node types required for the proposed flagship proof of concept deployment. Be especially sure to include server configuration types that represent the minimum and maximum requirements even if their aggregate number is small. Additionally, providing a % guide of the overall importance of each server type is also useful. Fill in as many tables as you deem appropriate with regards to covering the majority of your expected flagship deployment.**
1.1.5.1 Example Server Type A [CPU/Core]	4
1.1.5.1 Example Server Type A [RAM]	8GB
1.1.5.1 Example Server Type A [HDD]	80GB

1.1.5.1 Example Server Type A [SSD]	No specific requirement
1.1.5.1 Example Server Type A [NETWORK]	1Gbps
CPU	4 cores * 2-3 CPU GHZ/core
RAM	8 RAM GB/per VM (typically we require 2GB RAM / core)
RAM	8 RAM GB/per VM (typically we require 2GB RAM / core)
HDD	80 HDD GB/per VM (typically we require 20GB HDD / core)
SDD	0 SDD GB/per VM
Network	1 Gbps connectivity to CERN per 100 VM
CPU	4 cores * 2-3 CPU GHZ/core
SDD	0 SDD GB/per VM
HDD	80 HDD GB/per VM (typically we require 20GB HDD / core)
Network	1 Gbps connectivity to CERN per 100 VM
1.1.6 Virtualisation	**If you currently use virtualisation technologies and may rely on specific hypervisors please outline your usage currently. Select all applicable options for the flagship legacy deployments.**
[KVM]	Full
[XEN]	Full
[VMware]	Full
[Hyper-V]	Full
[ESX]	No specific requirement
[Other]	No specific requirement
1.2 Relevant Software Systems	
Please outline any	The ATLAS flagship consisted in deploying a batch cluster in the cloud and

relevant software systems in use for the flagship today and what systems might be required going forward as part of the initial flagship proof of concept.	does not require any particular software system. The CernVM image already contains all the needed software. However it is of interest to use a common, popular interface or translator to guarantee the future integration with provisioning systems.
1.2.1 Please outline the current situation, possible restrictions and any requirements on the portability of applications used by the flagship.	As long as we can run the Cern VM image, the application is portable.
1.2.2 Please outline the current practices, considerations and requirements on the test and go live strategy used by the flagship	Start PanDA batch cluster on moderate size and scale up.
1.3 Operational	**Please outline the capacity requirements and nature with regards to the flagship**
1.3.1.1 Minimum; most likely; maximum number of servers in a request	50;250;2500 servers with 4 cores each
1.3.1.2 Minimum; most likely; maximum amount of storage in a request	20GB per core. 4TB;20TB;200TB
1.3.1.3 Minimum; most likely; maximum external network bandwith in a request	1Gbps;1Gbps;10Gbps
1.3.1.4 Minimum; most likely; maximum lead time to provide the	3 day; 10 days; 30 days

requested cloud service	
1.3.1.5 Minimum; most likely; maximum service provision to be supported	50;250;2500
1.3.1.6 Minimum; most likely; maximum service provision you are willing to commit to	1 day; 1 week; 1 month
1.3.1.7 Minimum; most likely; maximum period of extension of service provisoning	1 day; 1 week; 1 month
1.3.1.8 Minimum; most likely; maximum period for service provision	1 day; 1 week; 1 month
1.3.1.9 Minimum; most likely; maximum period for termination of service provision	2 hours; 1 day; 1 week
1.3.1.10 Maximum allowed notification period for termination service provision	1 working day
1.3.2.1 Specify the types of users	A few Grid/Cloud experts
1.3.2.2 Which party should own the information about the users	No user data should be required
1.3.2.3 Which party should maintain the information about the user	No user data should be required
1.3.2.4 How should	No user data should be required

The Helix Nebula project is co-funded by the European Community Seventh Framework Programme (FP7/2007-2013) under Grant Agreement no 312301

information about the users be exchanged (e.g. API, Web Console, Custom software)	
1.3.3.1 How should the service catalogues from the multiple suppliers be presented to the users?	We don't require a service catalogue
1.3.4 Provisioning	**Please outline in the sub-sections below the resource provisioning models desired or required as part of the flagship deployment. Key areas to include are if automatic provisioning via API is required, aspects of any API functionality that may be required, expecting scaling requirements and models, required resource lead times, etc.**
1.3.4.1 Please outline the resource usage profile for deploying resources. Please indicate the variance of resource requirements over time, the length resources are expected to run for and the preferred method of usage.	50-2500 4-core servers as described in 1.3.1
1.3.4.2 Please outline acceptable lead times for new long term capacity requirements as well as short term ad hoc variable requirements as appropriate	15 minutes; 1 hour; 24 hours.
1.3.4.3 Please outline any requirements regarding availability and quality of service aspects of any provisioning tools you intend to use (such as	We will scale depending on the computing requirement for the LHC

API, web console etc.)	
1.3.4.5 Please outline any requirements regarding the delivery of content during the provisioning like the method of adding the content to the provided resource and any content delivery tools you intend to use (such as API, web console etc.)	Job outputs are written to the data centre at CERN
1.3.4.6 Please outline any requirements regarding the deployment of software during the provisioning like the method of adding / including the software to the image and any software deployment tools you intend to use (such as API, web console etc.)	We have our own software deployment infrastructure built into CernVM.
1.3.5 External Management / Compatibility	None
1.3.6 Monitoring	**Please outline any tools and requirements for monitoring of your system. If you have specific tools used already please outline them. Please also specify the aspects of computing being monitored, metrics, etc.**
1.3.6.1 System Monitoring	Ganglia, HammerCloud, PanDA monitor, GlideInWMS
1.3.6.2 Performance Monitoring	HammerCloud, PanDA monitor
1.3.6.3 Network Monitoring	Ganglia
1.3.6.4 Website	No preference

The Helix Nebula project is co-funded by the European Community Seventh Framework Programme (FP7/2007-2013) under Grant Agreement no 312301

Monitoring	
1.3.6.5 Security Monitoring	No preference
1.3.7 Service Level Reporting	**Please outline any requirements for service level reporting of the services.**
1.3.7.1 How should the service levels from the multiple suppliers be presented to the users?	Ideally single integrated
1.3.7.2 Who should have access to the service level reports?	Open, API accessible.
1.3.7.3 How do you want to access to the service level reports? (e.g. API, Web Console, Custom software)	API, web console.
1.3.7.4 Other requirements for service level reporting	None
1.3.8.1 How should the contract information of the services from the multiple suppliers be presented?	Each supplier presents its own contract information for the delivered services
1.3.8.2 Who should have access to the contract information of the service?	The owner of the service
1.3.8.3 How do you want to access to the contract information (e.g. API, Web Console, Custom software)	Web Console
1.3.8.4 Other	None

requirements for service level reporting	
1.3.9.1 Specify the tooling (mention supplier, software name and version) used to collect assets / software used by the flagships?	FOSS
1.3.9.2 Specify the tooling (mention supplier, software name and version) used to administrate assets / software used by the flagships?	FOSS
1.3.9.3 Who should collect assets / software used by the flagships running on, but not part of the cloud services?	FOSS
1.3.9.4 How do you want to access to the asset / software license information?	API, web console.
1.3.9.5 Other requirements for service level reporting	None
1.3.10 Preferred Cloud-Infrastructure	**Please outline any expected infrastructure management models for the various management aspects outlined below.**
1.3.10.1 Provisioning	API
1.3.10.2 Security	API
1.3.10.3 Networking	API
1.3.10.4 Billing	API, Web Console
1.3.11 Please outline	Scientific Linux

the operating systems you intend to utilise as part of the flagship deployment including specific versions.	
1.3.11.1 Other OS not listed above	CernVM (based on Scientific Linux)
Please outline your requirements with regards to software licensing. If you are expecting to host licensed commercial software as part of your flagship deployment please fill in this section.	No commercial software is used.
Do you anticipate cloud infrastructure providers to have access to running cloud virtual machines? If so please specify acceptable access levels	In principle infrastructure providers should not need to access the VMs.
Please outline the remote access methods you intend to utilise to manage flagship cloud infrastructure	VNC, SSH
1.3.15.1 Availability	availability = time_running / scheduled_up_time
1.3.15.2 Performance	
1.3.15.3 Reliability/Durability	reliability = time_site_is_available / {total_time – time_site_is_scheduled_down}
1.3.15.4 Accessibility	n/a
1.3.15.5 Disaster recovery	We assume copy of data also held at CERN

The Helix Nebula project is co-funded by the European Community Seventh Framework Programme (FP7/2007-2013) under Grant Agreement no 312301

1.3.16 Technical Support Models	Helpdesk, Email, Client Exposed Ticketing system
1.3.17 Technical Support Models	-
1.3.18 Enterprise Application Operation	n/a
1.3.19 Data and resource management	-
1.4 Security	**Please use this section to define any security requirements not mentioned elsewhere in the document; using the following subheadings as a guide (not all headings are required)**
1.4.1 Authentication, Authorisation and Accountability (AAA)	Identity enrollment requirements – how identity is initially verified, Single sign-on requirements for management interface, RBAC management requirements – e.g. for remote management interface. Accountability requirements – security-related logging, signed time stamping, WORM functionality.
1.4.1.1 Notes	
1.4.2 Remote management interface authentication requirements	Username/password, Soft certificate (x.509)
1.4.3.1 LAN security (for internal transfers) – e.g. IPSec	Not required.
1.4.3.2 Network segregation	Ideally, public IP addresses (no NAT).
1.4.3.2 Network access control (client health check)	Not required.
1.4.4.1 Secure de-provisioning/ deletion requirements	Not required.
1.4.4.2 Hardware	Not required.

decommissioning requirements (degauss etc…)	
1.4.4.3 Specific data export/portability requirements (formats, time limits)	Not required
1.4.5.1 Minimum SSH key length policy for remote access	2048/4096
1.4.5.2 Key management	Via API.
1.4.5.3 At-rest encryption (e.g. encryption gateway)	Not required
1.4.5.4 Crypto hardware/acceleration	Not required
1.4.5.5 Entropy/randomness sources.	Not required
1.4.6.1 Incident response services and service levels	Undefined
1.4.6.2 Incident/vulnerability severity classification used, if any	Undefined
1.4.6.3 Incident reporting (to/by demand side)	Undefined
1.4.6.4 Vulnerability reporting and management (to/by demand side)	Undefined

1.4.6.5 Testing requirements (e.g. external pen-testing)	Undefined
1.4.6.6 Third party security services used, interfaces required.	Undefined
1.4.7.1 Certifications required	Bank guarantee
1.4.7.2 Right to audit	Undefined
1.4.7.3 Any other procedural security policy requirements that would have to be complied with (e.g. around personnel clearance, subcontracting, jurisdiction).	https://procurement.web.cern.ch/sites/procurement.web.cern.ch/files/key-reference/FC_5312-ii_GeneralConditionsCERNContracts_en__v2008-11-27.pdf
1.4.8.1 Location/jurisdiction-limitations	Undefined
1.4.8.2 Third parties/subcontractor	Undefined
1.4.8.3 Breach notification	Undefined
1.4.8.4 Maximum, minimum data retention	Undefined
1.4.8.4 Access and rectification	Undefined
1.4.8.5 Purpose limitation	Undefined
1.5 Networking	**Please outline in more detail the expected networking requirements for the flagship. Include internal and external connectivity, availability**

	requirements, private and public networking needs and an overall expected deployment topology.
1.5.0.1 Estimate Necessary Capacity (peak bandwidth, 95th percentile bandwidth)	1Gbit/s per 100 cores peak. 95th unknown.
1.5.0.2 Topology	Connection to GEANT preferred.
1.5.0.3 Number of Nodes	-
1.5.0.4 Multiple Interfaces	-
1.5.0.5 Fail Over Plan	Not defined
1.5.1 Private Networking	-
1.5.2 Public Networking	Reverse DNS Management
1.5.3 DNS	Full DNS and reverse DNS resolution.
1.6 Storage	**Please outline overall storage requirements for the flagship with respect to total capacity, availability and performance metrics**
1.6.0.1 Capacity	20GB per core system partition on the VM
1.6.0.2 Availability	High.
1.6.0.3 Performance	Spinning disks. SSD not required. (Our tasks are CPU bound).
1.6.0.4 Usage Profile	Unknown.
1.6.1.1 Block Device Storage	Shared network disk not required. Only local system disk is required.
1.6.1.2 Volume Storage	Not required.
1.6.1.3 Object Storage	Not currently required. But we can discuss extending the flagship app to use cloud storage.
1.6.2.1 Geographical data storage	Preferably in Europe

requirements	
1.6.2.2 Large Volume Data Transfer	n/a
1.7 Billing	
Please outline any preferred or required billing and purchasing models, for example 'burst' based resource billing versus 'subscription' based longer term billing.	The contract price shall be net, firm and inclusive of all costs relating to the performance of the contractor's obligations under the contract and take into account CERN's exoneration from VAT and import duties.
If you have organisational requirements regarding reporting, invoicing, account use-age limits, agreed supplier lists, etc., please outline these in detail.	None
1.8 Legal Documents	
1.8.1 Please use this section to outline any SLA requirements not already articulated in other sections of this agreement already.	None
1.8.2 Please specify any particular requirements or preferences with regards to this document here.	None
1.8.3 Outline here any specific needs with regards to privacy policy here.	None

1.8.4 If you have any aspects of usage or requirements that may be appropriate against such a policy please outline them here.	Target: 97% reliablitiy (ATLAS WLG Tier-1 sites)
1.8.5 Any requirements regarding the treatment of your intellectual property please outline it here.	None
1.8.6 Any requirements regarding vendor neutrality please outline it here.	None
Other	Need a framework style procurement agreement with acceptable terms and conditions. For our scientific data processing tasks that would be sufficient but for our enterprise/admin workloads we would need higher SLAs and also recognition of the diplomatic status of CERN as an inter-governmental organisation
Other	-
Other	-
Other	-

4.4. The Port d'Informació Científica (PIC)

Contact person (name, affiliation, email):	Yolanda Vives (PIC, yvives@pic.es); Manuel Delfino (delfino@pic.es); Jordi Delgado (jordidem@pic.es)
Scientific Objective:	The scientific objectives for this flagship are a vast improvement in speed and quality in the process of finding surrogate biomarkers based on medical images of the brain (Magnetic Resonance Imaging (MRI) or Positron Emission Tomography (PET)) for the early detection and the study of pharmaceutical effects in neurodegenerative diseases like Alzheimer, Parkinson, depression, etc. For this purpose, a large amount of medical images of patients and healthy

	control subjects are stored and processed with specialized software tools. Afterwards, statistical comparisons between both groups are usually using different metrics obtained after the processing of the images, like the cortical thickness, volume and shape of the different subcortical structures, volume of the different tissues etc. Over the last 5 years we have developed PICNIC, a web-based platform running over a cluster that allows neuroimage researchers to easily store and organize their medical image data and to rapidly process entire datasets, decreasing the turnaround time from several months to a few days. This has in turn opened the door to repeating the processing sweeping the algorithm parameter space or running different algorithms, allowing a complete treatment of systematic uncertainties for the first time in this type of research. The technological objective of this flagship is the porting and deployment of PICNIC to the cloud, in order to start the process of technology transfer towards a commercial environment.
Expected Impact and Benefits:	By implementing the flagship on a commercial cloud system what impact will the result have on the scientific field? What benefit will it bring to the scientific community that the proposer organisation(s) directly address? Provide details about the scientific community that will benefit from this flagship and expected impact on scientific results. Include an estimate the number and distribution of users that will need access to the system.
Existing or potential partnership:	SME: The non-profit private sector entity INNDACYT has been setup in order to provide the technology and methodology transfer between the public research support environment that PIC provides and the commercial sector. INNDACYT gives a high importance to implementing a cloud back-end for PICNIC along the mainstream of cloud use for science in Europe. For this reason, rather than promoting an ad-hoc implementation with a particular brand of cloud provider, INNDACYT strongly supports that the development be done in collaboration with Helix-Nebula. Algorithm providers: PIC has stable relationships with algorithm development groups in Europe and the United States, in order to install and maintain their applications on PICNIC and to do joint activities, such as the testing and certification of GPU acceleration of the algorithms. These relationships would be naturally extended to the collaboration on the Flagship Application with Helix-Nebula. Data providers and future beneficiaries: PIC and INNDACYT have stable relationships for research support with the following research centers in the Barcelona area : Institut d'Investigació Biomèdica Sant Pau (IIB-Sant Pau), Institut Hospital del Mar d'Investigacions (IMIM) and Institut de Diagnòstic per la Imatge de l'Hospital Vall d'Hebrón (IDI). These researchers provide a significant group of qualified, realistic test users immediately. In addition, the research institutes which host these researchers can evaluate the estimated costs and benefits of the cloud-based PICNIC. Finally, as was mentioned earlier, these researchers are in collaboration with research groups around Europe and the world, and

	can easily help to showcase the results of the Flagship Application.
Proposer Motivation:	The main aspects that we consider as the most important during the pilot phase in the porting of PICNIC to the cloud are the following. Secure access: taking into account the confidentiality of the medical data used in this flagship, the secure access is one of the most important aspects to investigate. Although PIC has historically specified to all researchers that all uploaded data should be anonimized by the researchers, the practical experience is that an additional automated anonimization step must be included in the upload process. Therefore, the Flagship Application must implement this for data stored in the cloud (temporal or permanent) . Moreover, we will investigate different ways of encrypting the data in order to ensure the confidentiality of these data and give confidence to the researchers that the cloud implementation is safe to use. Scalability of resources: we will investigate how the resources scale as a function of the necessities of the medical researchers. Financial models: we are also interested in the costs of PICNICC, to study which of the following two models would be the most appropriate for our application and to study the possibility to deploy both of them to offer the opportunity to the user to choose according to the needs: 1) Storage and computing at cloud and 2) only computing at cloud. The first proposal has more advantages for researchers with a regular access to the system. For example the organization of the data is ensured by the web portal and they have no need to have storage servers at home (which would imply hardware, software and system administrator manpower costs). If the researchers process images punctually and they do not have a large amount of images, the second proposal would be certainly cheaper. However, the transfer of the input and output would take a long time. We are interested in evaluating the cost of both models and define approximately at which amount of data and frequency of access the first or the second model are more convenient. A mixture of both models could be also proposed, where the raw data is stored at home and only the processed data is kept in the cloud.
Proposer Long-term Objectives:	Short term objectives: significantly boost the uptake, flexibility, functionality and reliability for research purposes. Long term objectives: to transfer the technology to a company that can exploit this idea as a service for the medical researchers.
1.1.1 Please define the connectivity method expected to be utilised. Please select all methods that are compatible with a successful deployment.	Internet, Secure Remote User Access

1.1.4.1 Please check the formats of legacy server and/or drive images that will be provided for the proof of concept environment.	QCOW2 , DMG, VMDK, RAW
1.1.4.2 Please select whether server images being provided are private or publicly available images.	Private
1.1.4.3 Please select the correct server image type based on the technology you use today.	Xen, KVM
1.1.4.4 Please select the access methods for importing and extracting server images for the flagship to the Helix Nebula cloud framework that you would wish to be made available.	FTP, Https
1.1 Resource Sizing	**Please outline different server node types required for the proposed flagship proof of concept deployment. Be especially sure to include server configuration types that represent the minimum and maximum requirements even if their aggregate number is small. Additionally, providing a % guide of the overall importance of each server type is also useful. Fill in as many tables as you deem appropriate with regards to covering the majority of your expected flagship deployment.**
1.1.5.1 Example Server Type A [CPU/Core]	-
1.1.5.1 Example Server Type A [RAM]	-
1.1.5.1 Example Server Type A [HDD]	-

1.1.5.1 Example Server Type A [SSD]	-
1.1.5.1 Example Server Type A [NETWORK]	-
CPU	1 core * 3 GHz
RAM	2GB RAM (2GB RAM/core)
HDD	2GB (2GB HDD/core)
SDD	Not required
Network	100Mbps
CPU	4*3GHz
RAM	-
SDD	Not required
HDD	8GB HDD / 8 GB RAM
Network	100Mbps
1.1.6 Virtualisation	**If you currently use virtualisation technologies and may rely on specific hypervisors please outline your usage currently. Select all applicable options for the flagship legacy deployments.**
[KVM]	Full
[XEN]	Full
[VMware]	
[Hyper-V]	
[ESX]	
[Other]	
1.2 Relevant Software Systems	

The Helix Nebula project is co-funded by the European Community Seventh Framework Programme (FP7/2007-2013) under Grant Agreement no 312301

Please outline any relevant software systems in use for the flagship today and what systems might be required going forward as part of the initial flagship proof of concept.	The PICNICC flagship consists in deploying a batch cluster in the cloud and does not require any particular software system. The PICNICC-VM image already contains all the needed software.
1.2.1 Please outline the current situation, possible restrictions and any requirements on the portability of applications used by the flagship.	As long as we can run the PICNICC-VM image, the application is portable.
1.2.2 Please outline the current practices, considerations and requirements on the test and go live strategy used by the flagship	Start PICNICC batch cluster on moderate size and scale up.
1.3 Operational	**Please outline the capacity requirements and nature with regards to the flagship**
1.3.1.1 Minimum; most likely; maximum number of servers in a request	50;100;300 servers of the minimum size (1 core each) and 10;25;100 servers of the maximum size (4 cores each)
1.3.1.2 Minimum; most likely; maximum amount of storage in a request	100GB;200GB;800GB
1.3.1.3 Minimum; most likely; maximum external network bandwith in a request	100Mbps; 1Gbps; 10 Gbps
1.3.1.4 Minimum; most likely; maximum lead	1 hour; 2 hours; 24 hours. Maximum lead time is acceptable to achieve a lower price.

time to provide the requested cloud service	
1.3.1.5 Minimum; most likely; maximum service provision to be supported	50;100;300
1.3.1.6 Minimum; most likely; maximum service provision you are willing to commit to	1 day; 1 week; 1 month
1.3.1.7 Minimum; most likely; maximum period of extension of service provisoning	1 day; 1 week; 1 month
1.3.1.8 Minimum; most likely; maximum period for service provision	1 day; 1 week; 1 month
1.3.1.9 Minimum; most likely; maximum period for termination of service provision	2 hours; 1 day; 1 week
1.3.1.10 Maximum allowed notification period for termination service provision	1 working day
1.3.2.1 Specify the types of users	A few Grid/Cloud experts
1.3.2.2 Which party should own the information about the users	PIC
1.3.2.3 Which party should maintain the information about the user	PIC

1.3.2.4 How should information about the users be exchanged (e.g. API, Web Console, Custom software)	Web console, API
1.3.3.1 How should the service catalogues from the multiple suppliers be presented to the users?	PIC selects the service catalogue that will be shown to the users. Cloud services are hidden to the end users.
1.3.3.2 How do you want to access to the service catalogues (e.g. API, Web Console, Custom software)	API, Web Console
1.3.4 Provisioning	**Please outline in the sub-sections below the resource provisioning models desired or required as part of the flagship deployment. Key areas to include are if automatic provisioning via API is required, aspects of any API functionality that may be required, expecting scaling requirements and models, required resource lead times, etc.**
1.3.4.1 Please outline the resource usage profile for deploying resources. Please indicate the variance of resource requirements over time, the length resources are expected to run for and the preferred method of usage.	PICNIC workflow ---------- User manages the data and job submissions through the PICNIC web portal. 1. User selects images to process a. The images are uploaded to a master cloud server. The transfer will be performed using FTP, SSH or other protocol. b. Raw disk space (N_images * 30 MB). c. Transfer estimated time: (several GB, hours) 2. Creation of the cluster a. N_nodes on demand. b. Processing software access. c. Transferring a single image and a processing script for each node. 3. Run the processing workflow. (For each node) a. Verification of the environment. b. Loading Input data image. c. Processing. d. Generating output data. e. Storage the output data to the master cloud server. 4. Download the Output data. a. Data volume estimated (N * 300MB)

	b. Transfer estimated time: (several GB, hours) 5. Deprovision the generated cluster (minutes).
1.3.4.2 Please outline acceptable lead times for new long term capacity requirements as well as short term ad hoc variable requirements as appropriate	1 hour; 2 hours; 24 hours.
1.3.4.3 Please outline any requirements regarding availability and quality of service aspects of any provisioning tools you intend to use (such as API, web console etc.)	Scaling frequency is low and it will be due to multiple instances of the cloud cluster and one for each simultaneous dataset to be processed.
1.3.4.5 Please outline any requirements regarding the delivery of content during the provisioning like the method of adding the content to the provided resource and any content delivery tools you intend to use (such as API, web console etc.)	Output of the processing will be shared with the users via a web console
1.3.4.6 Please outline any requirements regarding the deployment of software during the provisioning like the method of adding / including the software to the image and any software deployment tools you intend to use (such as API, web console etc.)	The PICNICC-VM already contains all the needed software.

The Helix Nebula project is co-funded by the European Community Seventh Framework Programme (FP7/2007-2013) under Grant Agreement no 312301

1.3.5 External Management / Compatibility	
1.3.6 Monitoring	**Please outline any tools and requirements for monitoring of your system. If you have specific tools used already please outline them. Please also specify the aspects of computing being monitored, metrics, etc.**
1.3.6.1 System Monitoring	Nagios
1.3.6.2 Performance Monitoring	Ganglia
1.3.6.3 Network Monitoring	Ganglia
1.3.6.4 Website Monitoring	Nagios
1.3.6.5 Security Monitoring	selinux, iptables, fail2ban
1.3.7 Service Level Reporting	**Please outline any requirements for service level reporting of the services.**
1.3.7.1 How should the service levels from the multiple suppliers be presented to the users?	PIC assembles the SLA from the different suppliers.
1.3.7.2 Who should have access to the service level reports?	PIC (technical users and administrators of the platform)
1.3.7.3 How do you want to access to the service level reports? (e.g. API, Web Console, Custom software)	Web console
1.3.7.4 Other requirements for service level reporting	None

1.3.8.1 How should the contract information of the services from the multiple suppliers be presented?	Each supplier presents its own contract information for the delivered services.
1.3.8.2 Who should have access to the contract information of the service?	Specific persons/roles within PIC
1.3.8.3 How do you want to access to the contract information (e.g. API, Web Console, Custom software)	Web Console
1.3.8.4 Other requirements for service level reporting	n/a
1.3.9.1 Specify the tooling (mention supplier, software name and version) used to collect assets / software used by the flagships?	n/a
1.3.9.2 Specify the tooling (mention supplier, software name and version) used to administrate assets / software used by the flagships?	n/a
1.3.9.3 Who should collect assets / software used by the flagships running on, but not part of the cloud services?	n/a
1.3.9.4 How do you want to access to the asset / software license	n/a

The Helix Nebula project is co-funded by the European Community Seventh Framework Programme (FP7/2007-2013) under Grant Agreement no 312301

information?	
1.3.9.5 Other requirements for service level reporting	None
1.3.10 Preferred Cloud-Infrastructure	**Please outline any expected infrastructure management models for the various management aspects outlined below.**
1.3.10.1 Provisioning	Web console, API
1.3.10.2 Security	Web console, API
1.3.10.3 Networking	Web console, API
1.3.10.4 Billing	Web console, API
1.3.11 Please outline the operating systems you intend to utilise as part of the flagship deployment including specific versions.	Scientific Linux
1.3.11.1 Other OS not listed above	None
Please outline your requirements with regards to software licensing. If you are expecting to host licensed commercial software as part of your flagship deployment please fill in this section.	The principal softwares used are for non-profit purposes, under the GPLv2 licenses (FreeSurfer and FSL cases). Other softwares (like SPM) needs matlab. Initially, we will not deploy this software in the cloud. In a second stage of the project, this software could be included. We would like to use matlab in a charge per use basis.
Do you anticipate cloud infrastructure providers to have access to running cloud virtual machines? If so please specify acceptable access levels	In principle infrastructure providers should not need to access the VMs.

The Helix Nebula project is co-funded by the European Community Seventh Framework Programme (FP7/2007-2013) under Grant Agreement no 312301

Please outline the remote access methods you intend to utilise to manage flagship cloud infrastructure	NX, SSH
1.3.15.1 Availability	MTBF > 2 months; RTO< 12 hours
1.3.15.2 Performance	
1.3.15.3 Reliability/Durability	MTBF > 2 months; RTO< 12 hours
1.3.15.4 Accessibility	MTBF > 2 months; RTO< 12 hours
1.3.15.5 Disaster recovery	Not needed. PIC will hold a copy of the data.
1.3.16 Technical Support Models	Helpdesk, Online Chat, Email, Ticketing
1.3.17 Technical Support Models	Not integrated.
1.3.18 Enterprise Application Operation	n/a
1.3.19 Data and resource management	PIC and other users will upload the data via a web portal. Data management will be performed by PIC and the access to the data shall be restricted to PIC.
1.4 Security	**Please use this section to define any security requirements not mentioned elsewhere in the document; using the following subheadings as a guide (not all headings are required).**
1.4.1 Authentication, Authorisation and Accountability (AAA)	The access to our portal, that will be initially in our installations, is performed via IP tables and username and password.
1.4.1.1 Notes	None
1.4.2 Remote management interface authentication requirements	Username/password, x.509 certificate

1.4.3.1 LAN security (for internal transfers) – e.g. IPSec	IPsec
1.4.3.2 Network segregation	n/a
1.4.3.3 Network access control (client health check)	n/a
1.4.4.1 Secure de-provisioning/ deletion requirements	Wipe data after deprovisioning is required.
1.4.4.2 Hardware decommissioning requirements (degauss etc...)	Not required.
1.4.4.3 Specific data export/portability requirements (formats, time limits)	n/a
1.4.5.1 Minimum SSH key length policy for remote access	1024
1.4.5.2 Key management	Via API
1.4.5.3 At-rest encryption (e.g. encryption gateway)	Not required.
1.4.5.4 Crypto hardware/acceleration	Not required.
1.4.5.5 Entropy/randomness sources.	Not required.
1.4.6.1 Incident response services and	n/a

service levels	
1.4.6.2 Incident/vulnerability severity classification used, if any	n/a
1.4.6.3 Incident reporting (to/by demand side)	n/a
1.4.6.4 Vulnerability reporting and management (to/by demand side)	n/a
1.4.6.5 Testing requirements (e.g. external pen-testing)	n/a
1.4.6.6 Third party security services used, interfaces required.	n/a
1.4.7.1 Certifications required	Not required.
1.4.7.2 Right to audit	-
1.4.7.3 Any other procedural security policy requirements that would have to be complied with (e.g. around personnel clearance, subcontracting, jurisdiction).	None
1.4.8.1 Location/jurisdiction-limitations	Inside the EU only
1.4.8.2 Third parties/subcontractor	n/a

1.4.8.3 Breach notification	n/a
1.4.8.4 Maximum, minimum data retention	n/a
1.4.8.4 Access and rectification	n/a
1.4.8.5 Purpose limitation	n/a
1.5 Networking	**Please outline in more detail the expected networking requirements for the flagship. Include internal and external connectivity, availability requirements, private and public networking needs and an overall expected deployment topology.**
1.5.0.1 Estimate Necessary Capacity (peak bandwidth, 95th percentile bandwidth)	1 Gbit/s
1.5.0.2 Topology	Connection to GEANT preferred
1.5.0.3 Number of Nodes	-
1.5.0.4 Multiple Interfaces	-
1.5.0.5 Fail Over Plan	none
1.5.1 Private Networking	Not required
1.5.2 Public Networking	Reverse DNS management
1.5.3 DNS	Full DNS and reverse DNS resolution
1.6 Storage	**Please outline overall storage requirements for the flagship with respect to total capacity, availability and performance metrics**
1.6.0.1 Capacity	1 TB
1.6.0.2 Availability	Standard

The Helix Nebula project is co-funded by the European Community Seventh Framework Programme (FP7/2007-2013) under Grant Agreement no 312301

1.6.0.3 Performance	Spinning disks. SSD not required (tasks are CPU bound)
1.6.0.4 Usage Profile	Each core would read one file independently.
1.6.1.1 Block Device Storage	n/a
1.6.1.2 Volume Storage	We will evaluate volume and object storage services.
1.6.1.3 Object Storage	We will evaluate volume and object storage services.
1.6.2.1 Geographical data storage requirements	Data should be stored in the EU.
1.6.2.2 Large Volume Data Transfer	n/a
1.7 Billing	
Please outline any preferred or required billing and purchasing models, for example 'burst' based resource billing versus 'subscription' based longer term billing.	Subscription based (pay per use)
If you have organisational requirements regarding reporting, invoicing, account use-age limits, agreed supplier lists, etc., please outline these in detail.	Monthly invoice (including provider tax identification)
1.8 Legal Documents	
1.8.1 Please use this section to outline any SLA requirements not already articulated in other sections of this	None

agreement already.	
1.8.2 Please specify any particular requirements or preferences with regards to this document here.	None
1.8.3 Outline here any specific needs with regards to privacy policy here.	It will be required that any data which will be sent to and from the Helix Nebula cloud infrastructure as well as all data that results from processing on Helix Nebula will be owned and can only be used by PIC or the respective customer of our flagship's service.
1.8.4 If you have any aspects of usage or requirements that may be appropriate against such a policy please outline them here.	It will be required that any data which will be sent to and from the Helix Nebula cloud infrastructure as well as all data that results from processing on Helix Nebula will be owned and can only be used by PIC or the respective customer of our flagship's service.
1.8.5 Any requirements regarding the treatment of your intellectual property please outline it here.	It will be required that any data which will be sent to and from the Helix Nebula cloud infrastructure as well as all data that results from processing on Helix Nebula will be owned and can only be used by PIC or the respective customer of our flagship's service.
1.8.6 Any requirements regarding vendor neutrality please outline it here.	None
Other	-
Other	-
Other	-
Other	-

The Helix Nebula project is co-funded by the European Community Seventh Framework Programme (FP7/2007-2013) under Grant Agreement no 312301

5. Conclusion

We have gathered an extensive set of revealing information about the technical and non-technical requirements relating to each of the current flagships. The requirements have been assessed and fine-tuned since the first Proof of Concept phase and have been updated to improve the implementation of the second year of flagship deployments accordingly. We have identified the need for more non-technical requirements to be added as Helix Nebula evolves as a research and business ecosystem within a fully operational federated cloud. Service architecture and delivery models have been developed further and the requirements framework has been expanded accordingly to meet the current needs of the consortium. The requirements framework has succeeded in providing information about the expectations of each flagship and the evolving capabilities of each supplier.

The Helix Nebula project is co-funded by the European Community Seventh Framework Programme (FP7/2007-2013) under Grant Agreement no 312301